Review

# Social media data for conservation science: A methodological overview

Tuuli Toivonen[a,d,*], Vuokko Heikinheimo[a,d,*], Christoph Fink[a,d], Anna Hausmann[a,d], Tuomo Hiippala[a,b,d], Olle Järv[a,d], Henrikki Tenkanen[a,d], Enrico Di Minin[a,c,d]

[a] Department of Geosciences and Geography, University of Helsinki, Finland
[b] Department of Languages, University of Helsinki, Finland
[c] School of Life Sciences, University of KwaZulu-Natal, Durban 4041, South Africa
[d] Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Finland

## ARTICLE INFO

## ABSTRACT

Improved understanding of human-nature interactions is crucial to conservation science and practice, but collecting relevant data remains challenging. Recently, social media have become an increasingly important source of information on human-nature interactions. However, the use of advanced methods for analysing social media is still limited, and social media data are not used to their full potential. In this article, we present available sources of social media data and approaches to mining and analysing these data for conservation science. Specifically, we (i) describe what kind of relevant information can be retrieved from social media platforms, (ii) provide a detailed overview of advanced methods for spatio-temporal, content and network analyses, (iii) exemplify the potential of these approaches for real-world conservation challenges, and (iv) discuss the limitations of social media data analysis in conservation science. Combined with other data sources and carefully considering the biases and ethical issues, social media data can provide a complementary and cost-efficient information source for addressing the grand challenges of biodiversity conservation in the Anthropocene epoch.

## 1. Introduction

Human activities are the main drivers of the ongoing rapid worldwide loss of biodiversity (Maxwell et al., 2016). Understanding human-nature interactions is crucial for finding successful conservation solutions that help address the biodiversity crisis and support the wellbeing of the people (Bennett et al., 2017; Venter et al., 2016). Collecting data on human-nature interactions such as protected area visitation or resource extraction is, however, time-consuming and requires more resources than are usually available (Waldron et al., 2013). There is a need for new, efficient ways of collecting relevant information on people, nature and their interactions.

The ongoing information age is characterized by an increasing volume of data generated by user activities in virtual networks (Castells, 2010). Big Data, i.e., the massive quantities of digital information available provide new research avenues in various fields of science (boyd and Crawford, 2012; Kitchin, 2014; Ruths and Pfeffer, 2014). Digital conservation (Arts et al., 2015; Di Minin et al., 2015; Ladle et al., 2016) is a sub-field of conservation science that uses novel data sources such as social media data and other large data sets to understand and potentially mitigate the biodiversity crisis. User-generated big data may

offer cost-efficient ways for biodiversity monitoring (Hampton et al., 2013), but more importantly, they allow studying human-nature interactions on an unprecedented scale (Ruths and Pfeffer, 2014).

Among other sources of big data, social media provide a rich source for studying people's activities in nature and understanding conservation debates or discussions online (Di Minin et al., 2015). Social media refers to "web-based services that allow individuals, communities and organizations to collaborate, connect, interact, and build a community by enabling them to create, co-create, modify, share, and engage with user-generated content that is easily accessible" (McCay-Peet and Quan-Haase, 2017). In this article, we focus particularly on data from social networking sites, microblogs, and media sharing services, which support textual and visual content and geotagging. These sites include social media platforms such as Facebook, Twitter, Instagram, Flickr, and Weibo, where users share content either in private networks or publicly online.

Information gathered from social media provide new approaches to studying visitation patterns in conservation areas (Levin et al., 2015; Tenkanen et al., 2017; Wood et al., 2013), preferences and activities of protected area visitors, and mapping cultural ecosystem services (Gliozzo et al., 2016; Richards and Friess, 2015; van Zanten et al.,

2016). Social media data have also been proposed as a source of information to monitor public reactions to conservation-related events (Lunstrum, 2017; Macdonald et al., 2016) and to understand the global trends in wildlife trade (Di Minin et al., 2018, 2019). Spontaneously generated social media data differ from data collected in citizen science campaigns, where people actively contribute data for research in a structured manner (See et al., 2016).

Computer science and related fields are currently making rapid advances in methods for social media analytics. These methods are generally adopted in conservation science with a delay (Di Minin et al., 2015; Sherren et al., 2017). There are several studies using, for example, geotagged social media data in spatial analysis of human-nature interactions (Hausmann et al., 2018; van Zanten et al., 2016), but analyses combining different elements of social media data and advanced analysis methods remain scarce. Most of the previous studies that use social media data rely on manual content analysis, especially for visual content, and relatively simple spatial analyses (see literature review in Section 2). Di Minin et al. (2018, 2019) have recently proposed and demonstrated the use of automated social media content analysis for tracking illegal wildlife trade, whereas Becken et al. (2017) presented a case study analysing human sentiment towards the environment from social media. To leverage the full potential of social media data, conservation science would benefit from adopting methods for automated content analysis and combining these methods with advanced spatio-temporal analytics.

In this review paper, we present an overview of current analytical approaches to extract information from social media data for conservation science. Specifically, we (i) describe what kind of relevant information can be retrieved from social media platforms, (ii) provide a detailed overview of advanced methods for spatio-temporal, content and network analysis, (iii) exemplify the potential of these approaches for real-world conservation challenges and (iv) discuss the potential and limitations of social media data analysis in conservation science. The article aims to point out relevant tools and literature for those interested in using social media data in their research.

## 2. Previous research using social media data

We carried out a systematic literature review to understand how social media data have been used in conservation science. This review updates the previous review by Di Minin et al. (2015) and provides an overview of current research and existing gaps in using social media data in conservation science. For details on how the literature review was carried out, see Supplement S1. Results of the literature review are summarized in Table 1. The search and selection of the literature resulted in 35 published journal articles, which we classified into three thematic categories, namely, "people in nature", "biodiversity

monitoring" and "online discussions".

Most of the reviewed papers fell in the category "people in nature" (n = 19) and focussed on spatial, spatio-temporal and content analysis of human-nature interactions. Most of these studies used social media locations and timestamps as a proxy for human presence (Allan et al., 2015; Fisher et al., 2018; Ghermandi, 2016; Hausmann et al., 2017; Levin et al., 2017, 2015; Martinez-Harms et al., 2018; Sonter et al., 2016; Spalding et al., 2017; Tenkanen et al., 2017; Tieskens et al., 2017; Walden-Schreiner et al., 2018) and social media content for understanding human experiences and activities in nature (Giovos et al., 2018; Gliozzo et al., 2016; Hausmann et al., 2018; Heikinheimo et al., 2017; Richards and Friess, 2015; van Zanten et al., 2016).

"Biodiversity monitoring" articles (n = 7) focussed on obtaining information about real-world species observations. These studies use different combinations of social media geotags, as well as text, image, and video content to retrieve information about species observations, often focussing on a single species and using manual data access methods (Campbell and Engelbrecht, 2018; Di Camillo et al., 2018; Dylewski et al., 2017; Havlin et al., 2018; Rocha et al., 2017; Schuette et al., 2018).

"Online discussions" articles (n = 9) focussed on understanding the discussions around and diffusion of nature-based content on the internet. Based on text analysis related to the topic or species of interest, the studies analysed media coverage of conservation-related events and news (Hawkins and Silver, 2017; Macdonald et al., 2016; Papworth et al., 2015), online reactions and discussions in relation to events and management actions (Ebeling-Schuld and Darimont, 2017; Greer et al., 2017; Lunstrum, 2017; Wu et al., 2018), and the analysis of species-specific information in online media (Jarić et al., 2016; Willemen et al., 2015).

## 3. What data are available?

### 3.1. Characteristics of different social media platforms

Social media allow people to share and exchange content in online networks (Kaplan and Haenlein, 2010). Various social media platforms enable users to share posts containing text, images and video, and users can like, share and comment on each other's posts, forming a network of users and content. Personal profiles and posts can be either private or public, depending on the platform and the user's settings (Lange, 2007). Many social media platforms allow users to geotag their posts, which makes social media data analogous to other types of geographic information (Sui and Goodchild, 2011). While 'social media' is a broad concept (Kaplan and Haenlein, 2010), here we focus specifically on social networking sites and content communities such as Facebook, Twitter, Instagram, Flickr, and Weibo, which are likely to contain

**Table 1**
Summary of publications (n = 35) using social media data in conservation science. Category and subcategory indicate which topics were investigated and additional columns show which elements of social media data were used in the publications. The original publications and their classifications are provided in the Supplementary material.

| Category and Subcategory | Publications | Spatial | Temporal | Textual content | Image content | Video content |
|---|---|---|---|---|---|---|
| **People in nature** | **19** | **18** | **8** | **2** | **4** | **1** |
| - Activities /experiences | 7 | 6 | 1 | 2 | 4 | 1 |
| - Human presence | 12 | 12 | 7 | - | - | - |
| **Biodiversity monitoring** | **7** | **4** | **1** | **4** | **3** | **3** |
| - Species observations | 7 | 4 | 1 | 4 | 3 | 3 |
| **Online discussions** | **9** | **-** | **3** | **9** | **-** | **-** |
| - Species on the internet | 2 | - | - | 2 | - | - |
| - Online reactions to events | 4 | - | 1 | 4 | - | - |
| - Media coverage of topics | 3 | - | 2 | 3 | - | - |
| **Column Total** | **35** | **22** | **12** | **15** | **7** | **4** |

**Table 2**

Description of social media platforms most used in conservation science. MAU = Monthly Active Users in millions in 2018. API = Application Programming Interface. Listed publications are related to conservation science (based on the literature reviewed in Section 2). Table content modified and updated from Di Minin et al. (2015), and platform statistics and API details have been retrieved from platform websites in October 2018. Categories of different social networking sites were adapted from McCay-Peet and Quan-Haase (2017). The table has been sorted based on the number of reviewed publications per platform.

| Platform | Type of social media | Description | MAU | API link | Notes about data access | Publications |
|---|---|---|---|---|---|---|
| Flickr | Media sharing | Image sharing platform. Popular among nature photographers. | 90 | Flickr API: https://www.flickr.com/services/api/ | API is accessible for non-commercial use. Users may have a Creative Commons license attached to their photos. | N = 11 (Fisher et al., 2018; Ghermandi, 2016; Gliozzo et al., 2016; Hausmann et al., 2018; Martínez-Harms et al., 2018; Richards and Friess, 2015; Sonter et al., 2016; Spalding et al., 2017; Tenkanen et al., 2017; Walden-Schreiner et al., 2018; Willemen et al., 2015) |
| Facebook | Social networking | General purpose social networking site. Content (text, image, video) is shared on personal profiles, groups and organizational pages. | 2 230 | Facebook Graph API: https://developers.facebook.com/docs/graph-api/ | Specific permissions are required for reading data using the Graph API. Access to Facebook Pages API, which provided access to public posts within a limited time, was closed in April 2018 (Freelon, 2018). | N = 7 (Campbell and Engelbrecht, 2018; Di Camillo et al., 2018; Jarić et al., 2016; Lunstrum 2017; Macdonald et al., 2016; Papworth et al., 2015; Rocha et al., 2017) |
| Twitter | Microblogging | Micro-blog service for sharing short messages (max 270 characters). | 336 | Different APIs available, mainly Twitter Search API and Twitter Streaming API: https://developer.twitter.com/en/docs | The standard Search API supports sampling posts within past 7 days. The Search API is rate limited, and may return an incomplete search result. The timeline endpoint of the Twitter API allows the collection of most recent tweets from a single user. Twitter also provides commercial options (Premium and Enterprise search) with improved access to data. Twitter Streaming API returns a sample of real-time Tweets based on a search criteria. Twitter firehose API provides access to full twitter data stream (enterprise only). | N = 6 (Hawkins and Silver, 2017; Jarić et al., 2016; Macdonald et al., 2016; Papworth et al., 2015; Schuette et al., 2018; Tenkanen et al., 2017) |
| Instagram | Media sharing | Photo- (and video-) sharing platform. | 1 000 | Instagram Platform API: https://www.instagram.com/developer/ | Major changes in the Instagram Platform API were implemented in spring 2016, and autumn 2018. The service will be deprecated completely by 2020 and replaced by Instagram Graph API (related to Facebook's Graph API). | N = 5 (Hausmann et al., 2017,2018; Heikinheimo et al., 2017; Tenkanen et al., 2017; van Zanten et al., 2016) |
| Panoramio | Media sharing | Photo-sharing platform linked to Google Earth and Google Maps. | | Panoramio API was deprecated in 2016. | The service was discontinued in January 2018. | N = 4 (Ghermandi, 2016; Gliozzo et al., 2016; Tieskens et al., 2017; Van Berkel et al., 2018) |
| YouTube | Media sharing | Video-sharing platform. | 1 900 | YouTube Data API: https://developers.google.com/youtube/v3/ | There is a daily quota limit for using the YouTube Data API. | N = 4 (Di Camillo et al., 2018; Dylewski et al., 2017; Giovos et al., 2018; Macdonald et al., 2016) |
| TripAdvisor | Travel destination reviews | Platform for rating and discussing about travel destinations | 456 | TripAdvisor content API: https://developer-tripadvisor.com/content-api/ | TripAdvisor Content API is targeted only for travel websites and apps (only commercial use). | N = 1 (Cong et al., 2014) |
| Weibo | Microblogging /Social networking site | Micro-blog | 431 | http://open.weibo.com/wiki/API%E6%96%87%E6%A1%A3/en | Official documentation has not been updated. Developer discussion forums are useful for additional information. Knowledge on Chinese is a plus. | N/A |
| Vkontakte | Social networking site | Generic social networking site. Content is shared on personal profiles, groups and organizational pages. | 97 | https://vk.com/dev/openapi | API is designed to integrate the functionality of VKontakte with other applications. It is possible to, e.g., retrieve event information. | N/A |

relevant information for studying human-nature interactions (Di Minin et al., 2015).

Facebook is by far the most popular social networking site in the world, measured by the number of monthly active users (Table 2). Only in Russian-speaking countries, platforms VKontakte and Odnoklassiniki, and in China, QZone, are more popular than Facebook (http://vincos.it/world-map-of-social-networks/). Twitter and Instagram are widely popular in different parts of the world, especially in Europe, the Americas, and India. While the photosharing site Flickr is not as popular measured in number of users, it is often used for sharing nature-related content (Di Minin et al., 2015), and it has been widely used in research for studying nature recreation and cultural ecosystem services (see references in Table 2).

Information about the platform user base is a key element when accounting for platform-specific biases and population biases in social media analysis (Ruths and Pfeffer, 2014). Social media usage varies among different population groups, but detailed statistics are difficult to obtain globally. In the U.S., as an example, social media are popular in all age groups, but young adults (18- to 29-year-olds) are more likely than other age groups to use social media (Kemp, 2018). Surveys conducted in Finnish and South-African national parks support the same finding. Specifically, younger visitors to national parks are more likely to share their nature experiences in social media compared to older visitors (Heikinheimo et al., 2017), while women post their experiences online more likely compared to men (Hausmann et al., 2018). The lack of detailed statistics makes it challenging to estimate representativeness of the data in different spatial contexts, for example, within a specific national park.

Social media platforms differ in the type of content people share on them (Kaplan and Haenlein, 2010; Thelwall, 2009). Media-sharing platforms such as Flickr and Instagram are rich in visual content and related text descriptions and comments, while microblogs such as Twitter consist primarily of short text content with embedded images and links to other online content. Photo-sharing platforms seem to contain more content about people's activities and on-site observations, whereas microblogs might be more focussed on discussions around specific topics. General purpose social networking sites such as Facebook may contain a mixture of different content types. Despite these platform differences, people might share the same content across multiple platforms. For example, many geotagged tweets originate from Instagram (Heikinheimo et al., 2018).

### 3.2. Data acquisition

There are different approaches to acquire social media data ranging from manual searches to programmatic access to data (Batrinca and Treleaven, 2015; Lomborg and Bechmann, 2014). These approaches vary in required skill level of the analyst and the volume of data acquired. For instance, manually browsing through social media groups is time-consuming but requires little computational skill, while obtaining data using automated tools typically requires technical knowledge about the interface, programming skills, and an appropriate computing infrastructure for continuous data storage. Below, we examine the main data acquisition approaches in more detail.

**Application Programming Interfaces (APIs)** provide a defined set of methods for interacting with social media platforms in a programmatic way. Social media platforms often provide an API through which third parties such as application developers and researchers can interact with the platform automatically, which makes APIs an efficient tool for researchers (Lomborg and Bechmann, 2014). There are two prevailing architectures for APIs: *streaming APIs* (Joseph et al., 2014) and *REST* (representational state transfer) *interfaces* (Masse, 2011). Streaming APIs continuously deliver newly posted messages, e.g., on a specific topic in a read-only format. REST interfaces are used by requesting specific data from the API, allowing more flexible queries, also back in time. Using an API often requires a so-called access token or application key, which the platforms use to track and limit API usage. The process of acquiring access to an API varies among platforms (see notes in Table 2). Paying for improved API access is an option in some platforms, for example, in Twitter. Companies may alter which data are accessible through their API (Lomborg and Bechmann, 2014), and restrict or eliminate API access at any given time (Freelon, 2018). See an example of changes in Instagram data quality in the Supplementary material (Supplement, S3).

Several **data analysis software** packages contain functionalities for viewing, retrieving and analysing data from social media APIs. There are both proprietary tools and free and open-source solutions, which serve the purpose of collecting content from different social media platforms. For example, the ecosystem valuation software InVEST calculates the average annual number of photo-user-days based on geolocated Flickr photographs as a proxy for nature recreation (Wood et al., 2013). TAGS (Hawksey, 2010) is an open-source extension to Google Spreadsheet to continuously collect Twitter data based on selected keywords, while COSMOS is a standalone software that can be used to collect and analyse Twitter data with an easy-to-use user interface (Burnap et al., 2015).

There are also other options for social media data acquisition. The **purchase of data** from an authorized data vendor has several advantages such as little to no manual work and programming effort, and the availability to access time series of historical data. Costs may limit the practical usefulness of this approach in conservation science. **Web scraping**, or **web crawling**, is an approach for downloading and extracting data from web pages using an automated script. In comparison to APIs, web crawlers can only access the public web, while APIs may provide access to content that requires authentication (Lomborg and Bechmann, 2014).

There are various limitations when using computational tools for querying data from social media platforms (Brooker et al., 2016). While social media as such are subject to limitations and bias (boyd and Crawford, 2012), additional gaps and quality issues may be introduced when acquiring the data computationally (Brooker et al., 2016). For example, an API might return only a subsample of the requested data due to rate limits and access levels, and the quality of this sample may be difficult to evaluate (boyd and Crawford, 2012; Brooker et al., 2016).

Regardless of the data acquisition method, the researcher is responsible for storing and analysing the data in an ethical and responsible way (Lomborg and Bechmann, 2014; Zook et al., 2017). Each social media platform has its own Terms of Service that define how data acquired from these platforms can be stored and used by third parties. Clearly, these conditions should be acknowledged when retrieving data for research purposes (Batrinca and Treleaven, 2015). For example, web scraping often violates the terms and conditions of service providers, and researchers should understand the potential consequences of this violation (Freelon, 2018).

### 3.3. Elements of social media data

The information content of a social media post can be broken down into several elements: user information (full name, username, number of followers, user-defined home location), content (text, image, sound, video), timestamp (time when content was shared), geotag (automatic or user-defined location for the post), and comments and likes by other users (Fig. 1).

**Text content** is a core element of a social media post. Textual content often consists of short messages and captions that may include hashtags, emojis and external links. Additional text content is present in the comments related to the post. It is important to acknowledge that users post text in different languages and that a single post can contain text written in multiple languages (Hiippala et al., 2018). The language used in social media is often a spoken language containing abbreviations, emojis, hashtags and sarcasm. Twitter posts (tweets) are primarily textual, but tweets may also include interactive links to other

**Fig. 1.** Elements of a typical social media post viewed on a mobile device. Adapted from Poorthuis et al. (2016).

external media (image, GIF and video, web-page). The length of a tweet is currently limited to 280 characters (140 characters prior to November 2017). In contrast, Instagram and Flickr are primarily platforms for visual content – each post is an image or video that may be supplemented by a textual caption. Additional text content is present in the comments related to a post.

**Image content** often contains photographs taken by the user but may also contain memes, infographics and other types of visualizations. The user may have applied a filter to the content, which changes the colours and brightness of the image. Based on the manual content of social media images shared in national parks, users often share content related to biodiversity, landscape, and various human activities (Hausmann et al., 2018; Heikinheimo et al., 2017). In a Finnish national

park, users shared more content related to people and human activities (Heikinheimo et al., 2017), while in Kruger National Park, South Africa, most of the analysed content was related to charismatic biodiversity (Hausmann et al., 2018). The type of shared photographs also varies between platforms. Flickr was used more extensively for sharing biodiversity-related content, while Instagram was relatively more used for sharing content about social aspects of the visit (Hausmann et al., 2018).

**Geographic information** is another relevant element of a social media post. Location information might reveal the exact location where the post was shared. Users might also geotag or mention places they want to discuss, without actually having been there (e.g., "I wish I could travel to #Yellowstonenationalpark!"). A geotagged social media post
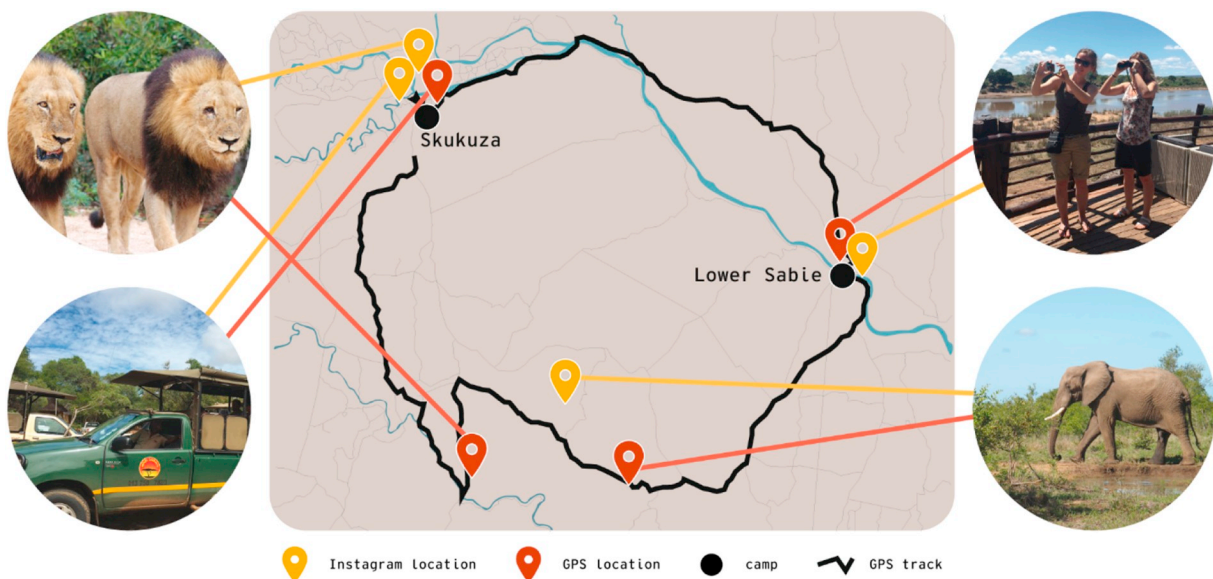


**Fig. 2.** Locations of social media images are not always accurate as shown in an example from Kruger National Park. The map shows Instagram (in yellow) and the spots where the photos were actually taken (in red). Images and GPS data by authors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
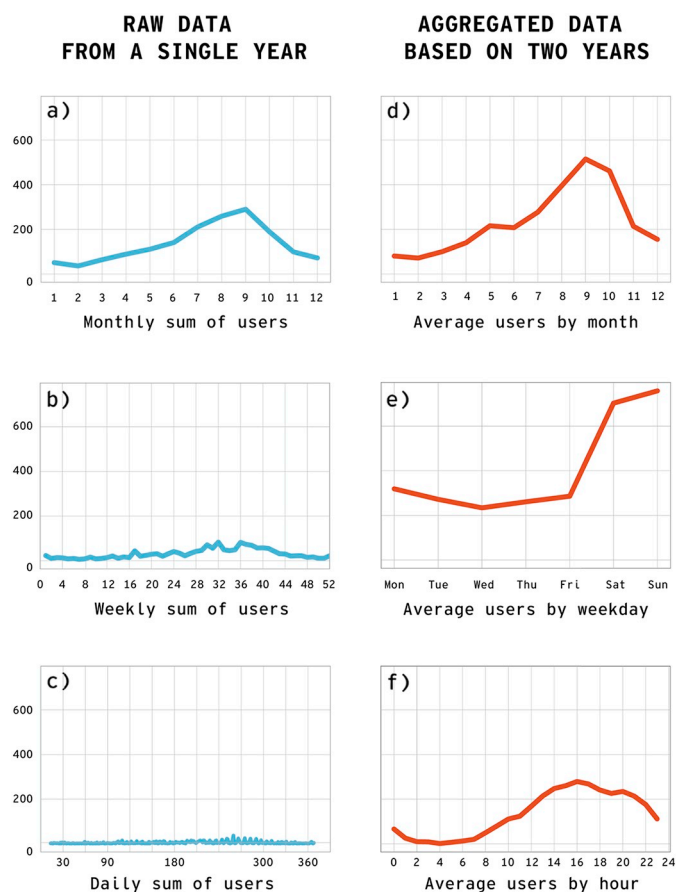
## RAW DATA
## FROM A SINGLE YEAR

## AGGREGATED DATA
## BASED ON TWO YEARS



**Fig. 3.** Timestamps in social media posts allow analysis of temporal patterns of people or public attention. The sporadic nature of data does not always allow a high level of detail, but aggregation may help. Blue lines demonstrate how increasing temporal granularity of inspection influences the number of observations per time unit: (a) shows the monthly visitation pattern, which is lost when the same data are inspected on a weekly (b) or a daily (c) basis. In the finest scale examination, many days are left with zero users. Aggregation of data may help to overcome this problem if temporal patterns are repeating (red lines). Here, data have been aggregated over two years for monthly (d), daily (e) and hourly (f) averages. The example is based on Instagram data for Nuuksio National Park in Finland for the years 2014 and 2015. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contains latitude and longitude coordinates in its metadata. However, the spatial accuracy of these coordinates depends on the user and the platform (Hochmair et al., 2018). For example, in Flickr, a post can have the exact coordinates automatically derived from the camera device, or the user can add spatial location to a post after using a webmap application. In Instagram, users geotag their content by selecting a place name from a pre-defined list of locations (such as, 'Kruger National Park' or 'The Dolomites') and hence, the geotag location approximates the actual location where the photo was taken (Fig. 2). Consequently, micro-scale (e.g., trail-level) spatial analysis is not always meaningful even with geotagged data (Wu et al., 2017). Only a small portion of all social media posts are geotagged, but the small portion is compensated by the vast volume of social media data (Poorthuis and Zook, 2017). In addition to coordinates, geographic information is also shared using place names and hashtags within the textual content of a post.

**Timestamps** in social media data reflect the date and time when the user sent a post online, or when a photo was taken. Timestamps provide useful information on temporal patterns of observations and activities. For example, temporal changes in national park visitor patterns can be

analysed based on social media timestamps (Tenkanen et al., 2017) (see also Box 3). Combined with content analysis techniques, timestamps may reveal public reactions to events and the time-lags for these reactions, for example, regarding the conservation of a specific species (see Box 4). Accuracy of the timestamp and the appropriate temporal scale of the analysis depend highly on the user who generated the data. Users might not post content immediately when an event happens, but only much later when, e.g., with good internet connection. It is also common to post old memories on social media, to reshare old content with the hashtag "#throwbacktime" or "#tb", or to talk about future events, making the timestamp of posting less relevant. Additionally, the sporadic nature of social media posting may hinder very detailed temporal analyses. If the temporal patterns are regular (for example, repeating weekly patterns in national park visitation), the data can be aggregated over longer periods of time to extract meaningful temporal patterns (Fig. 3).

**The user profile** of a social media account holder contains varying degrees of information about the user. Username, profile picture and a short description ("bio") available in the user profile can reveal different background characteristics about the user (Table 3). Some platforms allow users to add information about their home location/place of residence with a varying accuracy (Facebook, Twitter, and Flickr), while other platforms lack this information (e.g., Instagram). User profile information combined with content and spatial locations of posts has been used to infer further information such as age, gender, nationality, household type, and origin of the user (Longley et al., 2015). While these pieces of information are seldom accurate, they may provide relevant information on who posts what and when in social media (Di Minin et al., 2015) and how the preferences, activities, opinions or spatial patterns differ among different groups of people.

**Likes and comments**, as well as information about friends and followers of a user, can reveal interaction and communication on social media. Likes may reveal, for example, appreciation and preferences towards specific species (Hausmann et al., 2018), and comments may tell about the reactions and sentiment towards different topics. Likes, comments and friendships may reveal social networks around specific posts, topics and users. Information about social networks can be used, for example, for studying public reactions to specific events such as "rhino poaching" or "#cecilthelion", or to detect influential individuals and actors on social media.

## 4. How to analyse the data?

### 4.1. Overview of social media analysis workflow

Extracting meaningful information from social media data is an iterative process (Fig. 4) that often follows a typical data mining workflow (Han et al., 2011), in which a researcher needs to consider questions ranging from the acquisition and storage of data in an efficient and safe manner, to cleaning, filtering, transforming and enriching the data into a format that can be used to conduct the actual analyses.

Social media data, like big data sources in general, are characterized by large volumes of data, internal variability, varying veracity and high velocity of accumulation (Kitchin, 2014). Applied research often benefits from making big data small (Poorthuis and Zook, 2017), i.e., extracting only relevant data for further analysis. Filtering is indeed needed as social media may contain data generated by bots, meaning automated data generation algorithms, and advertisers, and the data may be inaccurately georeferenced, purposefully sarcastic or contain merely a circulating meme or a funny cat video. Hence, cleaning and filtering the data are two of the most crucial steps in any social media data analysis workflow (Di Minin et al., 2018; Varol et al., 2017). Transforming the structure and format of the data (for example, spatial and temporal aggregation (Tenkanen et al., 2017)) and enriching the data with relevant metadata (for example, language identification

**Table 3**
Different approaches for deriving information on users. ●●● = useful data, ●● = potentially useful data, ● = sometimes useful data.

| Social media variables | Age | Gender | Nationality/ Origin | Language (mother tongue) | Profession/ Occupation | Social Status/ Income Type | Lifestyle/ Personality | Place of Residence | Societal Influence |
|---|---|---|---|---|---|---|---|---|---|
| User name | ●● | ●●● | ● | | | | | | |
| User profile image | ●●● | ●●● | | | | | | | |
| User "bio" description | ● | ● | ● | ● | ● | ● | ● | ● | |
| User-defined place of residence | | | | | | | | ●●● | |
| User spatio-temporal mobility | | | ●● | | ●● | | ●● | ●●● | |
| Image and textual content | | | ●● | ●●● | ●● | ●● | ●● | ● | |
| Social network information | | | | | | | | ● | ●●● |
| Further readings: | Longley et al. (2015), Sloan et al. (2015) | Longley et al. (2015), Schwartz et al. (2013) | Longley et al. (2015) | Hiippala et al. (2018) | Sloan et al. (2015) | Sloan et al. (2015), Preoţiuc-Pietro et al. (2015) | Schwartz et al. (2013) | Hawelka et al. (2014) | Bakshy et al. (2011) |



**Fig. 4.** Main steps in social media data analysis and a checklist of questions to consider in different steps. See Supplementary material (S7) for detailed processing steps related to the practical analysis examples presented in Boxes 1–4.

(Hiippala et al., 2018)) might also be necessary before the actual analysis. At every step, but particularly when drawing conclusions on the results, a critical assessment of the results is needed. The results should be compared with other available information, if possible. Those interpreting the results should also be reminded about the biases, not to assume that social media reflects all behaviours of people.

In the following sections, we focus on one step of this iterative data mining flow: the analysis. We describe contemporary methods for spatio-temporal analysis, content analysis and network analysis in the context of social media data mining.

### 4.2. Spatio-temporal analysis

**Spatio-temporal analyses** of social media data take advantage of the geographic information and timestamp elements of social media posts (see Section 3.3). Gathering spatially and temporally explicit data on human activities has become widely available through the use of smartphones. Smartphones record their locations both for the mobile network operator services, as well as through mobile applications such as social media (Kitchin, 2014; Sui and Goodchild, 2011). The data collected by mobile devices are used as proxies to understand the changing distributions of people, i.e., dynamic population, but also to study the movement patterns of people (Frank et al., 2014; González et al., 2008; Järv et al., 2014). At large, there are two main groups of spatio-temporal analysis approaches that are useful for social media data analytics, location-based and person-based analysis (see Box 1).

#### 4.2.1. Location-based approach

A **location-based (or place-based) approach** is a relatively straightforward way of analysing the spatial patterns of social media posts and users using standard spatial analysis approaches from plotting the points on a map to calculating density surfaces or aggregating data over spatial units. Location-based approaches have been used, for example, to assess the presence of people in protected areas or to identify hotspots of human-nature interactions (Levin et al., 2015; Tenkanen et al., 2017). Such information may be complementary to official visitor statistics (see Box 1). Statistical models with ancillary data allow us to explain the variation in social media posts/user densities with various environmental variables, to gain understanding of, for example, landscape values or visitation preferences of people (Hausmann et al., 2018; van Zanten et al., 2016; Yan et al., 2018). Longitudinal location-based analyses are useful when studying the temporal variation in protected area visitation or 'dynamic populations'. Such variation is difficult to track with conventional data collection methods (Tenkanen et al., 2017; Wood et al., 2013). Combinations of spatio-temporal and content analysis approaches for social media data (see Section 3.3) have been used to study spatial distribution of species observations (Willemen et al., 2015), landscape values (van Zanten et al., 2016), human activities in nature (Heikinheimo et al., 2017), and sentiments and semantics related to a place (Hu, 2018; Liu et al., 2015).

#### 4.2.2. Person-based approach

A **person-based approach**, which focuses on human mobility at the individual level, requires much more challenging and advanced methods compared to location-based analyses (Järv et al., 2014; Rashidi et al., 2017; Wang et al., 2018). Person-based spatio-temporal analysis can provide more in-depth knowledge about people's actual movements and the places they have visited, for example, trips between different national parks (Fig. 5) (Hawelka et al., 2014; Heikinheimo et al., 2017; Huang and Wong, 2016). Additionally, comparing the location information of the post and the location provided by user's profile data may reveal global scale mobility patterns (Hawelka et al., 2014; Levin et al., 2015). Advanced analyses of visited locations and movement trajectories over a prolonged time period may benefit from additional human mobility analysis methods. These methods include anchor point modelling (Ahas et al., 2010) to reveal origins of social

media users, space-time prisms (Neutens et al., 2008) to evaluate potentially reachable locations within a national park, and sequence alignment (Shoval and Isaacson, 2007) to reveal visitation frequencies between locations and national parks. Further, linking an individual's mobility pattern with the content of social media posts over a prolonged time period can provide additional information on visitors' background (see Table 3), their potential visitation preferences, and actual experiences. It could be possible, for example, to classify visitors as different tourist types such as eco-, adventure, culture and leisure tourists based on geotagged posts in previously visited locations.

Hence, person-based analysis is useful for profiling protected areas based on visitors and visitation features, and for monitoring needs for nature conservation based on mobility trajectories within and between protected areas. Not the least, this knowledge is essential for developing managing and marketing of protected areas (Sevin, 2013), especially where visitor monitoring systems do not exist.

### 4.3. Content analysis

**Social media content analysis** refers to a collection of qualitative and quantitative methods for systematically describing the content that users post on social media platforms. Quantitative analyses of social media content remain scarce in conservation science, and previous studies have relied mainly on time- and resource-consuming manual content analysis (Eid and Handal, 2018; Hausmann et al., 2018; Hinsley et al., 2016). However, research on artificial intelligence is making rapid progress in analysing and understanding visual and textual content, which has enabled computers to classify species in photographs, to identify patches of forests in satellite images and to evaluate the emotions expressed by social media users. Many of these advances emerge from an approach to machine learning known as **deep learning**, which uses artificial neural networks that can learn to perform tasks without human supervision, given the networks are provided with numerous examples of paired input data (e.g., an image) and the desired output (e.g., a label for an object in the image) (Goodfellow et al., 2017; LeCun et al., 2015). The following sections describe the advances brought about by deep learning in the fields of computer vision and natural language processing, which provide methods for analysing visual and textual content on social media.

#### 4.3.1. Computer vision methods for visual content analysis

**Computer vision** is a broad, interdisciplinary field that studies the automatic processing and understanding of images. Visual content available on social media platforms such as photographs posted on Flickr, Instagram and Twitter can be analysed using computer vision methods. Visual content can be analysed, for example, to monitor species and ecosystems, as well as the threats they face. Computer vision methods can be used to identify species automatically for monitoring purposes (Norouzzadeh et al., 2018), to detect which species or wildlife products are being illegally traded on social media (Di Minin et al., 2018) or to identify cultural ecosystem services (Lee et al., 2019). Deep learning has achieved state-of-the-art results for challenging computer vision tasks such as classifying the contents of photographic images (Rawat and Wang, 2017), finding objects and identifying their outlines (He et al., 2017) and generating descriptions of entire images or their parts (Johnson et al., 2016; Karpathy and Fei-Fei, 2015; Zellers et al., 2018) (see Boxes 2 and 3).

#### 4.3.2. Natural language processing methods for textual content analysis

**Natural language processing (NLP)** is a field that studies the computational processing of all natural languages such as English or Swahili. Natural language processing methods can be used to analyse textual content on social media platforms such as Instagram captions or Twitter posts. Applications of NLP methods in conservation science have been rare but hold much potential for extracting useful information from textual content (Becken et al., 2017). NLP methods allow, for

**Box 1**

Assessing National Park Visitor Movements – Case Study from Kruger National Park, South Africa.

**Background:** Nature-based tourism generates important revenues to support conservation and management in protected areas. Understanding the mobility patterns of protected area visitors is crucial for developing conservation strategies both for local management as well as global marketing.

**Case example**: Kruger National Park (KNP), South Africa, is a popular wildlife-watching destination, attracting both national and international tourists. Tourists actively share their experiences on social media during their visit. Spatial and temporal patterns of geo-located social media posts may help reveal visitors' country or region of origin. Inside the park, geolocated posts may reveal hotspots for human-nature interactions, as well as areas that may be exposed to potential pressure on biodiversity such as crowdedness. In Fig. B1, we present examples of spatial analyses of park visitors based on social media data: Where are the hotspots of human-nature interaction in the park, what kind of trajectories do tourists have in the park, and where do the visitors come from?

**Method:** To identify the most popular roads within Kruger National Park (KNP), we retrieved geolocated Instagram posts (n = 81 379) inside the park boundaries for the year 2014. These data were matched with the closest road segments (using spatial join) making it possible to detect the roads with the highest social media activity. We analysed trajectories of visitors by drawing a line between subsequent social media post locations for each user account. These locations could be clustered to identify the common travel itineraries within the park. To obtain information on the countries of origin of the visitors in KNP, we traced back all locations globally between 2014 and 2015 (n = 136 141 000) for each user having posted within KNP in 2014 (n = 1408). These were visualized using a Sankey diagram.

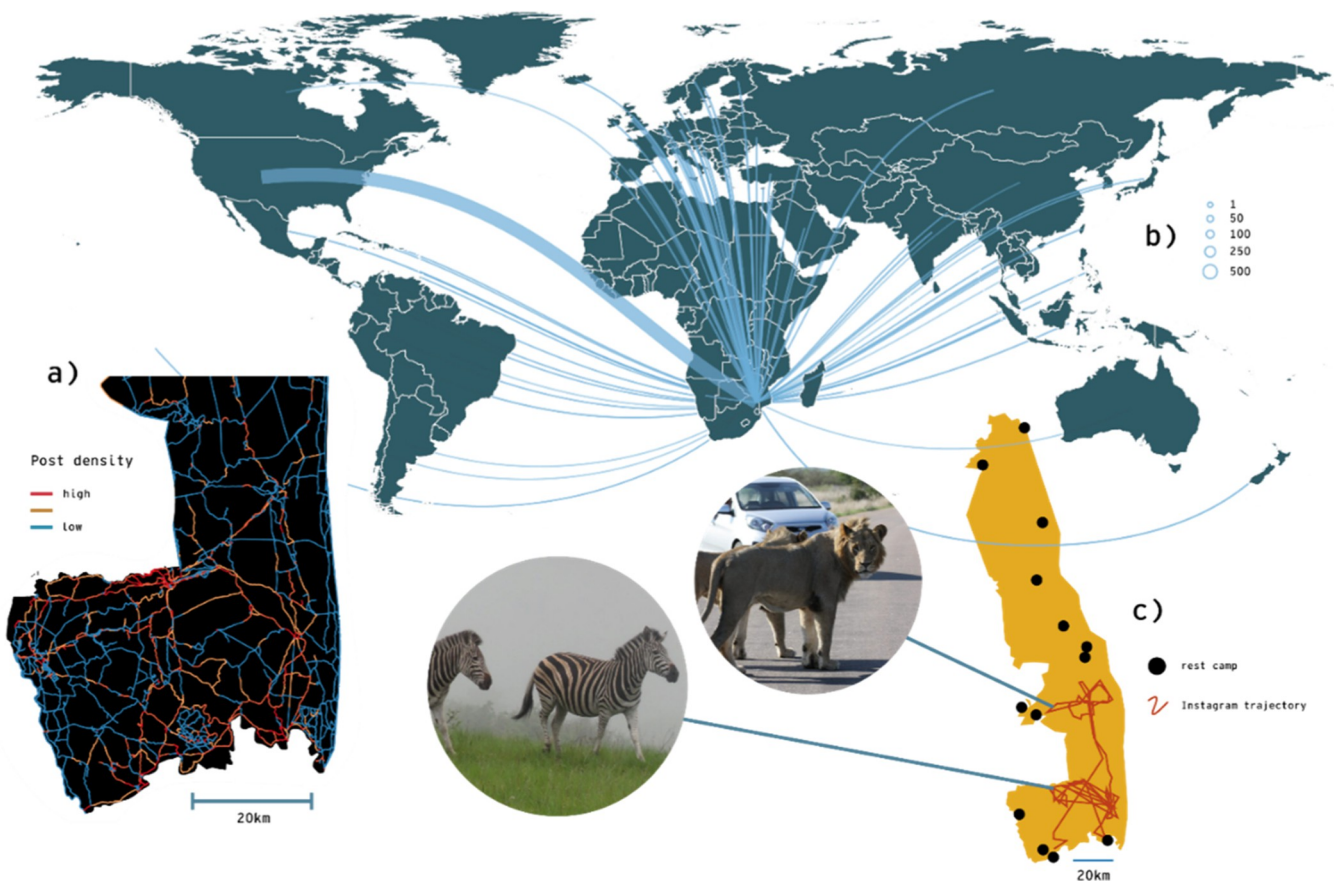**Related readings:** Hausmann et al., 2018 and Supplementary material.



**Fig. B1.** a) Location-based visualization of Instagram posts reveal hotspots of human-nature interaction on the roads of Kruger National Park. Person-based analysis of visitors in Kruger NP also indicate: c) visitor routes within the park, or b) the potential origin of social media users who visited the park. Read more details about location-based and person-based spatio-temporal analysis in Section 4.2.

instance, **automatic language identification**, **sentiment analysis** (e.g., negative posts related to crowdedness in a national park or the killing of 'Cecil' the lion) and **named entity recognition** for extracting the names of locations, organizations, individuals and species mentioned in social media posts. These NLP tasks have benefitted from recent advances in representing natural language using a technique known as word embeddings (Bojanowski et al., 2017; Mikolov et al., 2013; Peters et al., 2018). Word embeddings use neural networks to map words to vectors of real numbers, which can then be used as input data for downstream NLP tasks. However, applying NLP methods to social media content remains challenging due to the use of non-standard spellings, abbreviations, creative language and multiple languages
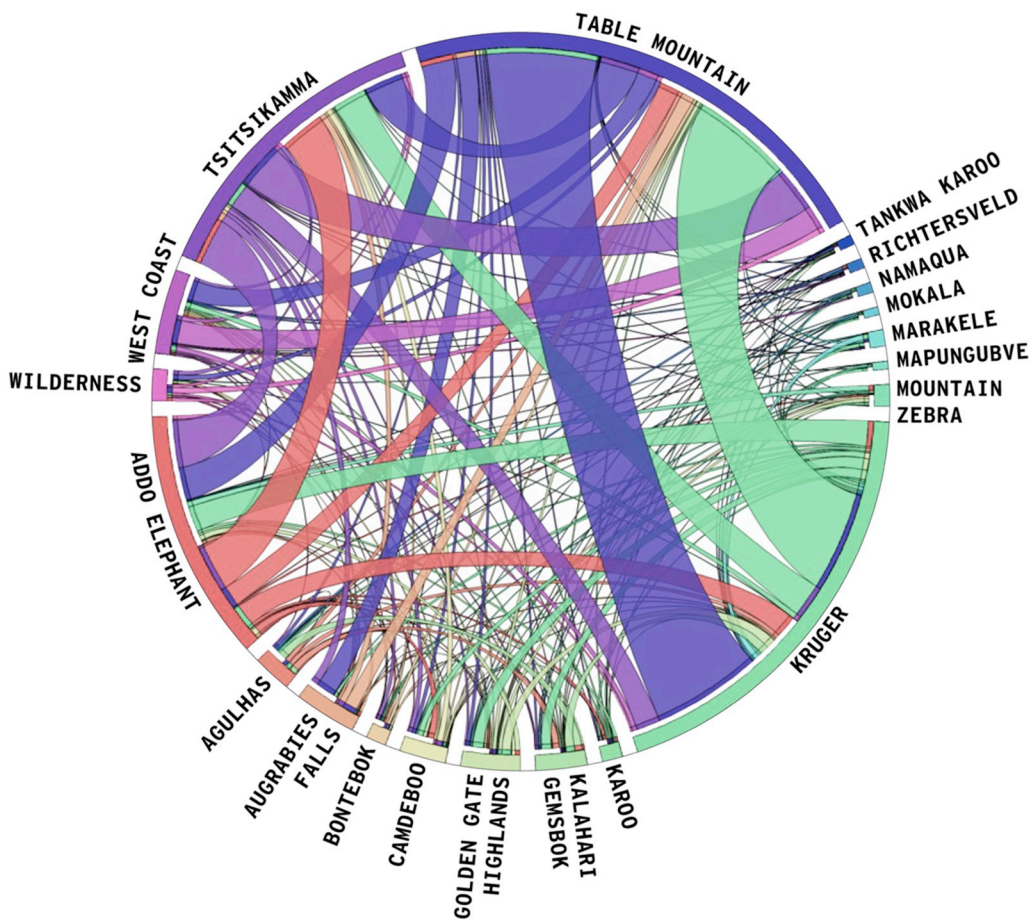
within a single post (Carter et al., 2013; Hiippala et al., 2018). NLP models are typically language- and domain-specific, which means that a model trained using English Wikipedia articles is likely to produce poor results when applied to data collected from social media platforms.

### 4.3.3. Combining visual and textual content analysis and current challenges

Social media users typically express themselves using combinations of visual and textual content such as a photograph and its caption. Accounting for this phenomenon, known as **multimodality**, is an emerging topic in automatic content analysis (Ramachandram and Taylor, 2018). These combinations can improve the performance of tasks such as sentiment analysis, as the combination of photograph and caption may prove more informative for determining the sentiment of a social media post than the photograph or caption alone (You et al., 2016) because the relevant information can be distributed across both visual and textual content. The same holds true for identifying species or human activities in social media posts. Captions may provide detailed information on the species or activity, whereas the computer vision model may recognize only the broader class to which they belong.

One drawback of using deep learning is the need for high volumes of training data for learning models that generalize well on unseen data. These training datasets are missing from conservation science and additional efforts will be needed to create them. This effort can be supported by deep learning, particularly for time-consuming tasks such as segmenting objects in images (Maninis et al., 2018), which reduces the resources needed for creating domain-specific, tailored datasets for conservation science. Shortcomings in data volume may be alleviated using a technique known as transfer learning, which involves learning representations on larger data sets and fine-tuning them to different downstream tasks such as classification (Weiss et al., 2016). Pre-trained

models and training data exist for various computer vision tasks (Deng et al., 2009; Krishna et al., 2017; Lin et al., 2014), whereas word embeddings and other language resources are now available for a large number of languages (Grave et al., 2018).

### 4.4. Social network analysis based on likes, comments and followers

Network analysis is a common methodology in computer science, and it has also been applied in ecology and social sciences (Cumming et al., 2010). **Social network analysis** is an approach for mapping and measuring the relationships between people, groups, and discussion topics as well as the dissemination of ideas (Wellman, 2001). Technically, a social network consists of a set of nodes and links that connect these nodes to each other (Marin and Wellman, 2014). In social media data, nodes can be individuals, posts or topics that are linked through online interaction such as likes, followerships, friendships, comments or shares (Fig. 6).

Forming a social network makes it possible to understand how individuals interact with each other; how strong the social ties are between individuals; how ideas, values and opinions spread in the network; and how different communities form around specific topics or opinions (Chamberlain et al., 2018; Crandall et al., 2010; Croitoru et al., 2015; Grabowicz et al., 2012; Reihanian et al., 2016; Takhteyev et al., 2012; Zhang et al., 2010). Social network analysis could provide new information for applications such as understanding the relationships among users and sellers of wildlife products (Di Minin et al., 2019), identifying influential actors in a conservation debate (Malan, 2011), or evaluating the influence of social networks when estimating the economic value of tourism and biodiversity (Spalding et al., 2017). Of the various analytical approaches developed for social network analysis, community detection and centrality analyses may have the

**Box 2**

Automated Content Detection from Social Media Images.

---

**Background:** Observations by nature enthusiasts contribute to species monitoring worldwide. In addition to citizen science platforms designed specifically for this purpose, there is an increased interest in using social media for collecting information about nature observations and human-nature interaction. Extracting relevant information from high volumes of social media data often requires automated content analysis, but note: it is advisable to check the social platform terms of use before applying these methods.

**Case example:** We show how two commonly used content analysis methods, dense captioning and instance segmentation, can be used to identify species or human-nature interactions in social media images. We demonstrate the differences between the methods by applying them to the same images. Dense captioning, which combines computer vision and natural language processing, detects visual regions of interest in the images and generates a linguistic description for each region. Instance segmentation locates objects in images and estimates their outline. This information can be used to count objects and to filter visual content for further processing such as more detailed species classification. The two approaches produce different types of content recognition, as shown with the examples.

**Method:** Dense captioning uses a pre-trained neural network with a VGG16 backbone, trained and implemented by the authors (Johnson et al., 2016). Instance segmentation uses a pre-trained deep neural network, namely, Mask R-CNN (He et al., 2017) with a ResNeXt X-101-32x8d-FPN backbone, as implemented in the Detectron library (Girshick et al., 2018) (Fig. B2).



**Fig. B2.** A comparison of results for instance segmentation (left) and dense captioning (right) for two pairs of images. Instance segmentation shows promising results for detecting objects, even if they are only partially visible or overlap each other, which enables, for example, counting the number of objects such as individuals of a certain species. Dense captioning provides richer descriptions, which may be used for searching image contents but lacks accuracy for object detection. In both cases, the results are limited to the classes and descriptions present in the training data: the model cannot detect objects that are not present in the training data. For more examples, see the Supplementary material, S4.

---

highest potential for such conservation-related questions.

**Community detection approaches** can be categorized broadly into topology-based and topic-based community detection approaches (Ding, 2011). The topology-based community detection algorithms such as the Link clustering method (Ahn et al., 2010), Clique Percolation method (Palla et al., 2005), or BIGCLAM (Yang and Leskovec, 2013) use information about social connections such as friendship, followership or communication between users to cluster nodes into

distinguishable groups. Topic-based community detection aims at grouping and separating individuals into specific communities based on the topics that are discussed, or sentiments or opinions about a specific topic. Topic-oriented community detection (Zhao et al., 2012) combines natural language processing techniques with clustering and community detection algorithms that can reveal communities that are interested in specific topics such as a species or a protected area. Sentiment modelling can further reveal if the communities address a

**Box 3**

Temporal Variation of Visitor Activities in Protected Areas.

---

**Background:** Recreational activities in protected areas provide people access to cultural ecosystem services such as physical recreation and sense of place. However, the activities of visitors may cause direct or indirect disturbance to the environment. Inspecting the content of social media data in space and time provides increased understanding of activities and their relationship, e.g., to the phenological stages of nature.

**Case example:** Pallas-Yllästunturi in Lapland is the most visited national park in Finland. The park visitation is extensively monitored through visitor counters and systematic visitor surveys, repeated every 5 years. The surveys reveal that hiking and skiing are the most popular activities in the park, depending on the season. Social media data provide complementary information about the activities in the park, as it allows the analysis of the activities of people over time. This information is useful for understanding temporal trends in activities, the relationship of these trends to changes in nature and for detecting emerging activities (Heikinheimo et al., 2017). Here, we demonstrate how social media text and imagery capture the temporal fluctuation of activities in different seasons.

**Methods:** We used Instagram posts extracted within a 10-km buffer zone around the Pallas-Yllās National Park between January 2014 and April 2016 (N = 19,939) (Heikinheimo et al., 2017). We searched for all posts mentioning skiing, hiking and biking, and related environmental conditions (snow and autumn colours) in the most popular languages (English, Finnish) using a keyword search (Fig. B3).

**Further readings:** Heikinheimo et al., 2017 and Supplementary material.
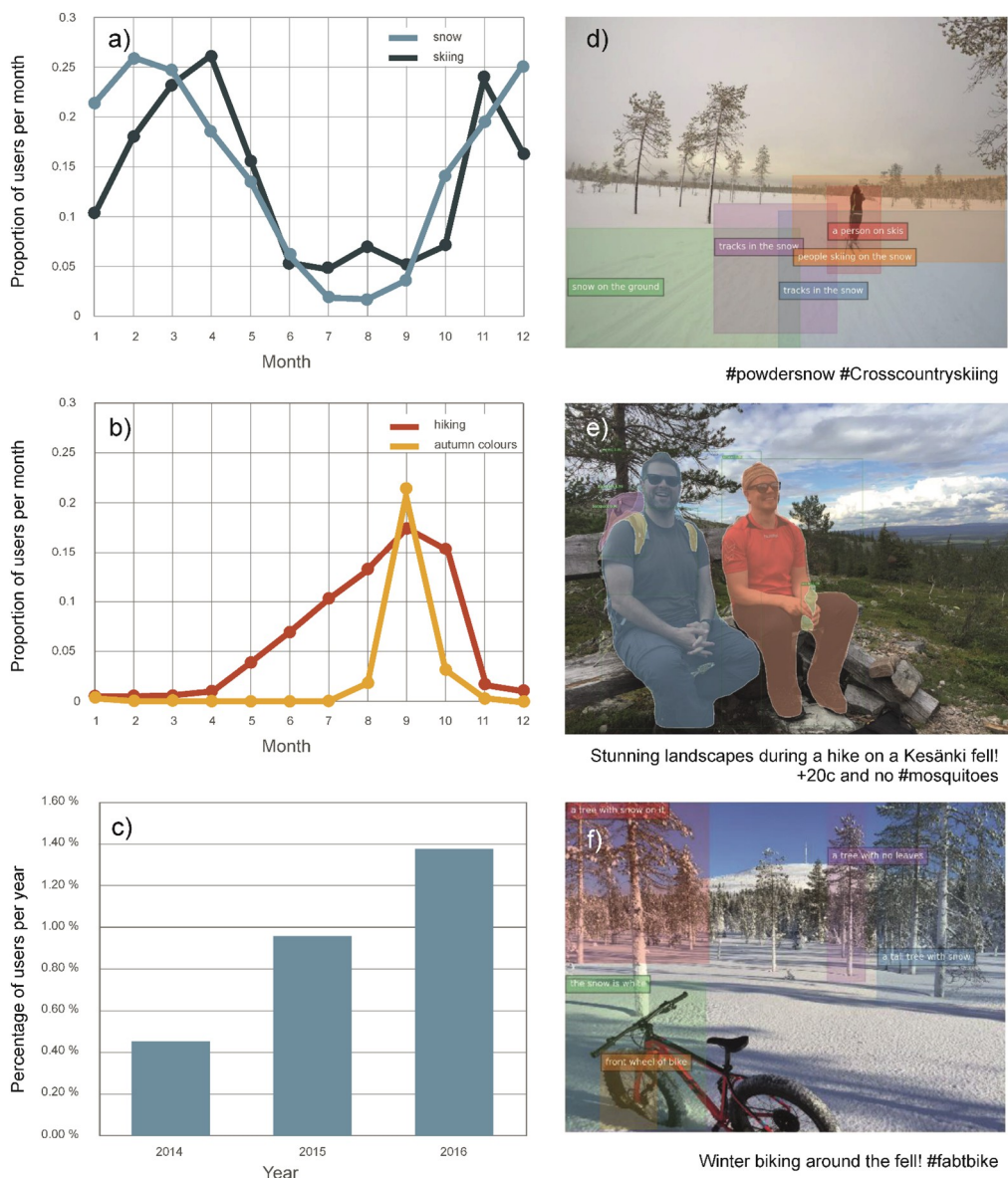


**Fig. B3.** Temporal patterns of Instagram posts mentioning popular activities (skiing, hiking and biking), and the most characteristic seasonal elements (snow and autumn colours) in Pallas-Yllästunturi national park in Finnish Lapland. For skiing (a) and hiking (b), the seasonal trends are clear and they match with the phenological changes in the landscape. Biking is an emerging activity that has been growing in popularity during the recent years, as visible also in social media posts (c). The graphs (a-c) are based on the text content only. Photos (d)-(f) provide examples of visual content from the national park labelled using dense captioning (d, f) and instance segmentation (e) (see Box 2 and the Supplementary material for more examples).
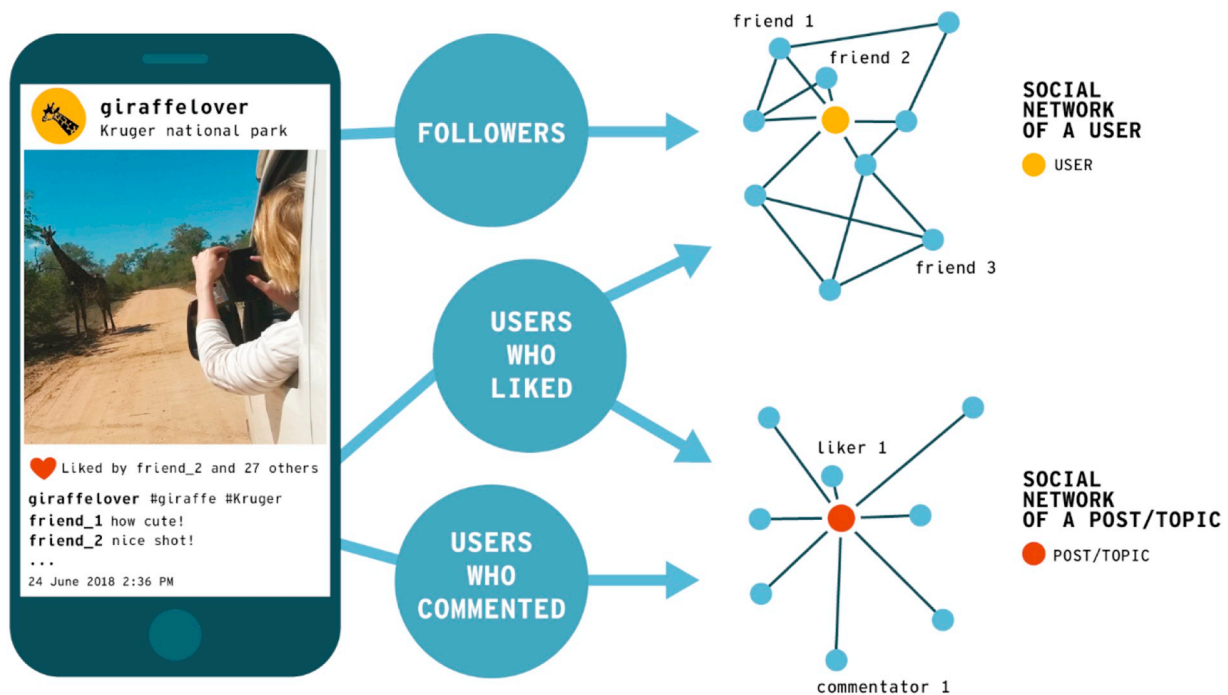
**Fig. 6.** Social media data make it possible to form a social network of a user based on combination of follower information, likers and commentators. It is also possible to form a social network around a post or topic by generating the network based on users who liked or commented on the post/topic.

specific topic positively or negatively (see Box 4). Various centrality measures such as betweenness centrality, degree centrality or eigenvector centrality (Landherr et al., 2010) make it possible to identify key nodes from the social network such as opinion leaders or key informants of a social group that can be useful for targeting conservation marketing.

## 5. Discussion and conclusions

### 5.1. Most possibilities are still unexplored

Digital innovation, particularly social media analysis, has brought new research avenues to conservation science (Arts et al., 2015; Di Minin et al., 2015). In this paper, we have presented concrete data acquisition and analysis approaches that might benefit conservation scientists interested in using social media data as an input for their research. We focussed on social media data retrieved from social networking sites and content communities (Kaplan and Haenlein, 2010), where users can share textual and visual content and geotag their posts. The methods presented for spatio-temporal analysis, content analysis and network analysis can also be applied to other types of data such as Wikipedia and other digital corpora (Ladle et al., 2016), camera trap imagery (Norouzzadeh et al., 2018) and various kinds of citizen-science data.

There is a rapidly growing body of literature presenting different use cases of social media data in conservation science (see Table 1). Most of the published research examples have used social media for studying human-nature interactions from a place-based perspective, for example, using social media as a proxy of visitation in natural areas (Sonter et al., 2016; Tenkanen et al., 2017). Indeed, most of the existing studies are based on mapping georeferenced social media data (Section 2, Table 1). However, there is potential to move "beyond the geotag" (Crampton et al., 2013) and to leverage the potential of more advanced social media analysis techniques also in conservation science. Applying more sophisticated analysis methods and examining several elements of social media data together, from spatially and temporally tagged

content to information on users and their networks, can provide new information about human-nature interactions, particularly from the perspective of people. Ultimately, social media provide information on people's behaviour (Ruths and Pfeffer, 2014), and human dimensions of environmental issues, which are essential to understand for successful conservation actions (Bennett et al., 2017).

### 5.2. Big but small social media data

Social media data are generally referred to as 'big data' (Crampton et al., 2013; Kitchin, 2014), and the volume of user-generated data worldwide is overwhelming. However, the amount of data needed in end analysis can be reduced significantly if one is interested only in a specific topic or a specific region (Di Minin et al., 2018; Poorthuis and Zook, 2017). The data are generally rich in terms of visual and textual content and variable across topics and languages (Hiippala et al., 2018). Despite the evident noisiness of these data, social media content has been found to contain relevant information for studying human-nature interactions, from relatively local to global scale (Hausmann et al., 2018; Heikinheimo et al., 2017; Levin et al., 2015). Overall, social media data seem most useful for studying relatively broad areas that are frequently visited by people (Tenkanen et al., 2017) or topics that are popular among social media users. On limited spatial and temporal scales, the content may become too sporadic for meaningful analysis (Fig. 4).

Data quality remains a challenge in any social media analysis workflow (boyd and Crawford, 2012; Longley et al., 2015). Data extracted from social media can be heavily biased towards specific user groups and geographic regions (Kemp, 2018). Social media users tend to share more positive and likeable content online, instead of negative experiences, causing a positivity bias in the content (Reinecke and Trepte, 2014). Photos and text are often meant to entertain rather than document and hence their value, e.g., biodiversity monitoring is varying. Despite the biases, social media content has been found to reflect surveyed preferences and activities of national park visitors in South Africa and Finland (Hausmann et al., 2018; Heikinheimo et al.,

**Box 4**

Assessing Public Sentiment for Conservation – Case Pangolin.

**Background:** Big data mining approaches can be used to follow and analyse debates around conservation (Di Minin et al., 2015; Ladle et al., 2016). Social media content analysis may reveal the volume of online discussion but also the sentiment users express in response to species or events and how these change over time. Understanding the sentiments of people towards conservation of species or protection of areas is important for developing conservation strategies and communication.

**Case example:** Pangolins are the most trafficked mammal species in the world. Awareness campaigns on social media have raised attention about the critical conservation status of the species. Social media content and sentiment analysis shows how much and with what sentiments social media users have posted about the species. With improved knowledge about public opinion, awareness campaigns can be made more effective, and public support for the conservation of the species can be increased.

**Method:** We collected Twitter posts (*tweets*) mentioning Pangolin in one of 20 relevant languages from February to May 2018, using the platform's API. We then used the sentiment classification framework of Webis (Hagen et al., 2015), which is optimized for short text length such as in tweets, to identify the sentiment (positive, neutral, negative) of all posts. Finally, we calculated a mean sentiment of all tweets posted in one day by assigning "positive" messages a value of 1.0, "negative" messages –1.0, and neutral 0.0. This results in a value in the interval [–1.0, 1.0] to describe the mean sentiment for each day in the dataset (Fig. B4).
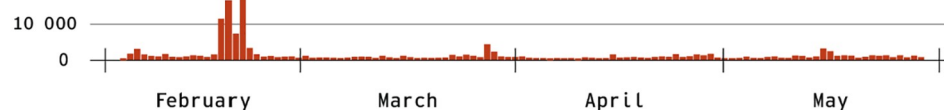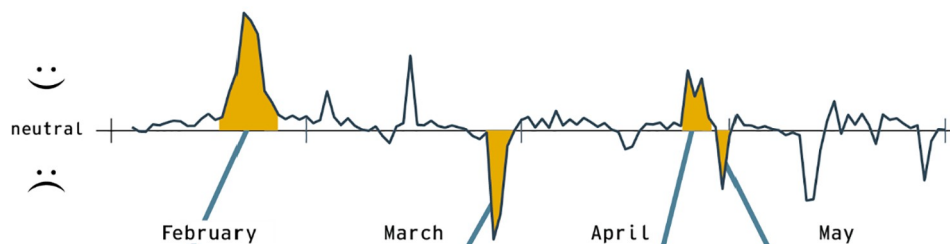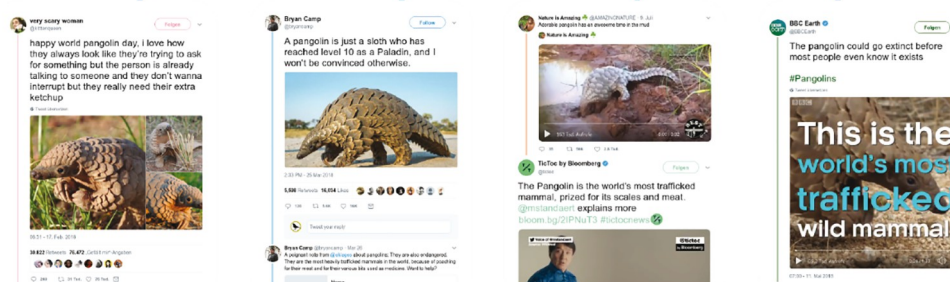


**Fig. B4.** The volume of Twitter messages relating to the keyword "pangolin" changes over time, with users reacting to awareness campaigns, conservation events or similar news items (a). Sentiment analysis of the content of the post allows us to track positive, neutral or negative attitudes. The variation of the sentiment over time (b) can be used to identify prevailing topics and link them to specific events (c). World Pangolin Day (17 February 2018) accounts for a clear spike in both volume and sentiment.

2017), and social media usage rates correlate with official visitor statistics in popular nature destinations in different parts of the world (Tenkanen et al., 2017; Wood et al., 2013).

In addition to inherent biases in social media data, computational data acquisition methods may introduce additional problems to the retrieved data (Brooker et al., 2016). Most platforms offer only limited datasets and metadata to researchers, and the sampling algorithms for platform APIs remain unknown (Joseph et al., 2014). Textual content in social media is rich in information but difficult to analyse due to its variability. For example, detecting sarcasm when analysing online sentiment is challenging, particularly as posts are made with tens of languages. Specific words in one language may have completely

different meanings in other languages or they may refer to different things even with one language. As an example, geotagged posts about tigers may be found from Africa simply because people post about drinking Tiger beer.

Social media can also serve as a platform in citizen science campaigns, especially regarding studies related to biodiversity monitoring (see Table 1). For example, researchers can encourage interested citizens to post animal sightings in a special interest group on social media (Campbell and Engelbrecht, 2018; Rocha et al., 2017), or using a specific hashtag. Using popular social media platforms in active data collection campaigns may engage wider audiences in the data collection efforts and could save time from developing separate applications.

If possible, it is important to compare the patterns observed from social media with other data sources gathered with more traditional means of data acquisition such as surveys and statistics [see examples in (Hausmann et al., 2018; Heikinheimo et al., 2017; Tenkanen et al., 2017)], to assess the quality of the data and to understand who and what the data represent. Social media data are created spontaneously by the users, and while this spontaneity generates a bias to the data, it is also a source of information revealing what and where people consider worth posting about.

### 5.3. Access to data and privacy issues

In conservation science, researchers have traditionally relied on their own data collection or data collected by conservation authorities, NGOs or active citizen scientists. These data are increasingly available online with clear licensing and well-described APIs (Di Minin and Toivonen, 2015). Social media data sources differ from this, as the platforms are first and foremost profit-making companies with business interests. Companies may, without prior notice, change the usage rights to the data, data structure, or access to APIs (Freelon, 2018; Lomborg and Bechmann, 2014), with consequences to data availability and quality (see Supplement S3). It may be challenging to keep up with ongoing changes in the APIs and use rights, as the changes are seldom openly documented. This limitation is problematic particularly for longitudinal research projects and monitoring, and it challenges the reproducibility of research (Lomborg and Bechmann, 2014). If planning for long-term data collection from social media, it is good to acknowledge the risk of discontinuity of the platform or data access.

One way to overcome the access challenges is an increased direct collaboration between researchers and the companies. At best, the broader research community of digital conservation could establish collaboration platforms with the social media companies. With established collaboration models and infrastructures in place, the administrative work related to contracts would be reduced and analyses could be run within the safe environments of social media platforms without breaching privacy fences. This kind of collaboration could be very helpful in, for example, investigating illegal wildlife trade on social media (Di Minin et al., 2019).

Moreover, new mechanisms could be established so that social media users are able to actively donate their data for research purposes via the platforms. Using automated content recognition techniques, Instagram has already prompted users not to share content that is abusive to animals. What if next time you post a picture of an animal or a plant in your favourite social media platform, you would receive a prompt: "*Looks like you are sharing a nature observation. Would you like to share this post for research use?*", allowing you to save a copy of your post to a research database. The newly implemented European Union directive GDPR supports this, as it obligates companies to store their data in a format that is portable, i.e., machine readable by other applications.

### 5.4. Future directions

In a workshop organized at the European Conference for

Conservation Biology on 12 June 2018, we asked expert participants (n = 23) to identify new research avenues for applying social media data in conservation science (see Supplement S2). Brainstorming resulted in novel ideas for biodiversity monitoring, visitor mapping, as well as understanding values and sentiments – similar topics emerged from our literature review (Section 2). The majority of suggestions concentrated on gathering new knowledge on people's values and appreciation related to biodiversity, either in certain locations or related to certain species, or more broadly to understand nature-centred societal discourses. Social media data were also seen as a platform for active communication by individuals and organizations; a channel for nudging people towards more biodiversity-aware lifestyles.

We foresee significant advances in assessing these topics by using automated multimodal (text, image, video and audio together) content analysis. Rapid advancements in image recognition, natural language processing and other applications of machine learning provide possibilities to use social media data more comprehensively (see Box 4). Mining opinions and sentiments and other techniques linked to conservation culturomics (Ladle et al., 2016) becomes more accurate through improved training of algorithms to interpret correctly, e.g., sarcastic expressions or content in multiple languages. Advanced content analysis methods used together with spatial and temporal analytics may allow, for example, monitoring of public conservation discourses (Becken et al., 2017), investigations of illegal wildlife trade (Di Minin et al., 2018, 2019), identifying emerging opportunities and threats for biodiversity conservation in protected areas, or analysing return on investment of biodiversity campaigns.

When interpreting results based on social media data analysis, researchers must carefully consider ethical issues, minimize potential harm and aim at fairness (Zook et al., 2017). Privacy issues and data anonymization should be considered even if using publicly shared content. The potential biases should be borne in mind not to accidentally suggest that the data reflect the opinion or behaviour of the entire population. Overall, social media data are best used with other data sources to gain a full and dynamic image on a conservation issue, for the benefit of people and biodiversity.

### Acknowledgments

### Appendix A. Supplementary material

Supplementary material (more results and further method descriptions) to this article can be found online at https://doi.org/10.1016/j.biocon.2019.01.023.

### References

Ahas, R., Silm, S., Järv, O., Saluveer, E., Tiru, M., 2010. Using mobile positioning data to model locations meaningful to users of mobile phones. J. Urban Technol. 17, 3–27. https://doi.org/10.1080/10630731003597306.

Ahn, Y.-Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. Nature 466, 761–764. https://doi.org/10.1038/nature09182.

Allan, J.D., Smith, S.D.P., McIntyre, P.B., Joseph, C.A., Dickinson, C.E., Marino, A.L., Biel, R.G., Olson, J.C., Doran, P.J., Rutherford, E.S., Adkins, J.E., Adeyemo, A.O., 2015. Using cultural ecosystem services to inform restoration priorities in the Laurentian Great Lakes. Front. Ecol. Environ. 13, 418–424. https://doi.org/10.1890/140328.

Arts, K., van der Wal, R., Adams, W.M., 2015. Digital technology and the conservation of nature. Ambio 44, 661–673. https://doi.org/10.1007/s13280-015-0705-1.

Bakshy, E., Hofman, J.M., Mason, W.A., Watts, D.J., 2011. Everyone's an influencer: quantifying influence on Twitter. In: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11. ACM, New York, NY, USA,

pp. 65–74. https://doi.org/10.1145/1935826.1935845.

Batrinca, B., Treleaven, P.C., 2015. Social media analytics: a survey of techniques, tools and platforms. AI & Soc. 30, 89–116. https://doi.org/10.1007/s00146-014-0549-4.

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

Bennett, N.J., Roth, R., Klain, S.C., Chan, K., Christie, P., Clark, D.A., Cullman, G., Curran, D., Durbin, T.J., Epstein, G., Greenberg, A., Nelson, M.P., Sandlos, J., Stedman, R., Teel, T.L., Thomas, R., Veríssimo, D., Wyborn, C., 2017. Conservation social science: understanding and integrating human dimensions to improve conservation. Biol. Conserv. 205, 93–108. https://doi.org/10.1016/J.BIOCON.2016.10.006.

Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. Enriching word vectors with subword information. Trans. Assoc. Comput. Linguist. 5, 135–146.

boyd, danah, Crawford, K., 2012. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf. Commun. Soc. 15, 662–679. https://doi.org/10.1080/1369118X.2012.678878.

Brooker, P., Barnett, J., Cribbin, T., Sharma, S., 2016. Have we even solved the first 'big data challenge?' Practical issues concerning data collection and visual representation for social media analytics. In: Snee, H., Hine, C., Morey, Y., Roberts, S., Watson, H. (Eds.), Digital Methods for Social Science: An Interdisciplinary Guide to Research Innovation. Palgrave Macmillan UK, London, pp. 34–50. https://doi.org/10.1057/9781137453662_3.

Burnap, P., Rana, O., Williams, M., Housley, W., Edwards, A., Morgan, J., Sloan, L., Conejero, J., 2015. COSMOS: towards an integrated and scalable service for analysing social media on demand. Int. J. Parallel Emergent Distrib. Syst. 30, 80–100. https://doi.org/10.1080/17445760.2014.902057.

Campbell, H., Engelbrecht, I., 2018. The Baboon Spider Atlas – using citizen science and the 'fear factor' to map baboon spider (Araneae: Theraphosidae) diversity and distributions in Southern Africa. Insect Conserv. Divers. 11, 143–151. https://doi.org/10.1111/icad.12278.

Carter, S., Weerkamp, W., Tsagkias, M., 2013. Microblog language identification: overcoming the limitations of short, unedited and idiomatic text. Lang. Resour. Eval. 47, 195–215. https://doi.org/10.1007/s10579-012-9195-y.

Castells, M., 2010. The information age. Media Stud. A Read. 2, 152.

Chamberlain, B.P., Levy-Kramer, J., Humby, C., Deisenroth, M.P., 2018. Real-time community detection in full social networks on a laptop. PLoS One 13, e0188702. https://doi.org/10.1371/journal.pone.0188702.

Cong, L., Wu, B., Morrison, A.M., Shu, H., Wang, M., 2014. Analysis of wildlife tourism experiences with endangered species: an exploratory study of encounters with giant pandas in Chengdu, China. Tour. Manag. 40, 300–310. https://doi.org/10.1016/J.TOURMAN.2013.07.005.

Crampton, J.W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W., Zook, M., 2013. Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. Cartogr. Geogr. Inf. Sci. 40, 130–139. https://doi.org/10.1080/15230406.2013.777137.

Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J., 2010. Inferring social ties from geographic coincidences. Proc. Natl. Acad. Sci. U. S. A. 107, 22436–22441. https://doi.org/10.1073/pnas.1006155107.

Croitoru, A., Wayant, N., Crooks, A., Radzikowski, J., Stefanidis, A., 2015. Linking cyber and physical spaces through community detection and clustering in social media feeds. Comput. Environ. Urban. Syst. 53, 47–64. https://doi.org/10.1016/J.COMPENVURBSYS.2014.11.002.

Cumming, G.S., Bodin, Ö., Ernstson, H., Elmqvist, T., 2010. Network analysis in conservation biogeography: challenges and opportunities. Divers. Distrib. 16, 414–425. https://doi.org/10.1111/j.1472-4642.2010.00651.x.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'09), pp. 248–255.

Di Camillo, C.G., Ponti, M., Bavestrello, G., Krzelj, M., Cerrano, C., 2018. Building a baseline for habitat-forming corals by a multi-source approach, including Web Ecological Knowledge. Biodivers. Conserv. 27, 1257–1276. https://doi.org/10.1007/s10531-017-1492-8.

Di Minin, E., Toivonen, T., 2015. Global protected area expansion: creating more than paper parks. Bioscience 65, 637–638. https://doi.org/10.1093/biosci/biv064.

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

Di Minin, E., Fink, C., Tenkanen, H., Hiippala, T., 2018. Machine learning for tracking illegal wildlife trade on social media. Nat. Ecol. Evol. 2, 406–407. https://doi.org/10.1038/s41559-018-0466-x.

Di Minin, E., Fink, C., Hiippala, T., Tenkanen, H., 2019. A framework for investigating illegal wildlife trade on social media with machine learning. Conserv. Biol. 33, 210–213. https://doi.org/10.1111/cobi.13104.

Ding, Y., 2011. Community detection: topological vs. topical. J. Inf. Secur. 5, 498–514. https://doi.org/10.1016/J.JOI.2011.02.006.

Dylewski, L., Mikula, P., Tryjanowski, P., Morelli, F., Yosef, R., 2017. Social media and scientific research are complementary—YouTube and shrikes as a case study. Sci. Nat. 104, 48. https://doi.org/10.1007/s00114-017-1470-8.

Ebeling-Schuld, A.M., Darimont, C.T., 2017. Online hunting forums identify achievement as prominent among multiple satisfactions. Wildl. Soc. Bull. 41, 523–529. https://doi.org/10.1002/wsb.796.

Eid, E., Handal, R., 2018. Illegal hunting in Jordan: using social media to assess impacts on wildlife. Oryx 52, 730–735. https://doi.org/10.1017/S0030605316001629.

Fisher, D.M., Wood, S.A., White, E.M., Blahna, D.J., Lange, S., Weinberg, A., Tomco, M., Lia, E., 2018. Recreational use in dispersed public lands measured using social media data and on-site counts. J. Environ. Manag. 222, 465–474. https://doi.org/10.1016/J.JENVMAN.2018.05.045.

Frank, M.R., Williams, J., Mitchell, L., Bagrow, J., Dodds, P.S., Danforth, C.M., 2014. Constructing a taxonomy of fine-grained human movement and activity motifs through social media. arXiv (Prepr. arXiv1410.1393).

Freelon, D., 2018. Computational research in the post-API age. Polit. Commun. 0, 1–4. https://doi.org/10.1080/10584609.2018.1477506.

Ghermandi, A., 2016. Analysis of intensity and spatial patterns of public use in natural treatment systems using geotagged photos from social media. Water Res. 105, 297–304. https://doi.org/10.1016/j.watres.2016.09.009.

Giovos, I., Keramidas, I., Antoniou, C., Deidun, A., Font, T., Kleitou, P., Lloret, J., Matić-Skoko, S., Said, A., Tiralongo, F., Moutopoulos, D.K., 2018. Identifying recreational fisheries in the Mediterranean Sea through social media. Fish. Manag. Ecol. 25, 287–295. https://doi.org/10.1111/fme.12293.

Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K., 2018. Detectron.

Gliozzo, G., Pettorelli, N., Muki Haklay, M., 2016. Using crowdsourced imagery to detect cultural ecosystem services: a case study in South Wales, UK. Ecol. Soc. 21, art6. https://doi.org/10.5751/ES-08436-210306.

González, M., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782. https://doi.org/10.1038/nature06958.

Goodfellow, I., Bengio, Y., Courville, A., 2017. Deep Learning. MIT Press, Cambridge, MA and London.

Grabowicz, P.A., Ramasco, J.J., Moro, E., Pujol, J.M., Eguiluz, V.M., 2012. Social features of online networks: the strength of intermediary ties in online social media. PLoS One 7, e29358. https://doi.org/10.1371/journal.pone.0029358.

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T., 2018. Learning word vectors for 157 languages. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), pp. 3483–3487.

Greer, K., Day, K., McCutcheon, S., 2017. Efficacy and perception of trail use enforcement in an urban natural reserve in San Diego, California. J. Outdoor Recreat. Tour. 18, 56–64. https://doi.org/10.1016/J.JORT.2017.02.002.

Hagen, M., Potthast, M., Büchner, M., Stein, B., 2015. Webis: an ensemble for twitter sentiment detection. In: Proceedings of SemEval.

Hampton, S., Strasser, C., Tewksbury, J., Gram, W., Budden, A., Batcheller, A., Duke, C., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162. https://doi.org/10.1890/120103.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.

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. Sci. Rep. 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. Conserv. Lett. 11. https://doi.org/10.1111/conl.12343.

Havlin, P., Caravaggi, A., Montgomery, W.I., 2018. The distribution and trophic ecology of an introduced, insular population of red-necked wallabies (Notamacropus rufogriseus). Can. J. Zool. 96, 357–365. https://doi.org/10.1139/cjz-2017-0090.

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geolocated Twitter as proxy for global mobility patterns. Cartogr. Geogr. Inf. Sci. 41, 260–271. https://doi.org/10.1080/15230406.2014.890072.

Hawkins, R., Silver, J.J., 2017. From selfie to #sealfie: Nature 2.0 and the digital cultural politics of an internationally contested resource. Geoforum 79, 114–123. https://doi.org/10.1016/J.GEOFORUM.2016.06.019.

Hawksey, M., 2010. TAGS – using Google Spreadsheet to automatically monitor Twitter event hashtags and more [WWW document]. URL. https://mashe.hawksey.info/2010/06/using-google-spreadsheet-to-automatically-monitor-twitter/.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988. https://doi.org/10.1109/ICCV.2017.322.

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 Int. J. Geo-Information 6, 85. https://doi.org/10.3390/ijgi6030085.

Heikinheimo, V., Tenkanen, H., Hiippala, T., Toivonen, T., 2018. Digital imaginations of national parks in different social media: a data exploration. In: Proceedings of PLATIAL'18: Workshop on Platial Analysis. Heidelberg, Germany, September 20–21, pp. 45–52. https://doi.org/10.5281/zenodo.1472745. Zenodo.

Hiippala, T., Hausmann, A., Tenkanen, H., Toivonen, T., 2018. Exploring the linguistic landscape of geotagged social media content in urban environments. Digit. Scholarsh. Humanit. https://doi.org/10.1093/llc/fqy049.

Hinsley, A., Lee, T.E., Harrison, J.R., Roberts, D.L., 2016. Estimating the extent and structure of trade in horticultural orchids via social media. Conserv. Biol. 30, 1038–1047. https://doi.org/10.1111/cobi.12721.

Hochmair, H.H., Juhász, L., Cvetojevic, S., 2018. Data quality of points of interest in selected mapping and social media platforms. In: Progress in Location Based Services 2018 (Vol. Lecture Notes in Geoinformation and Cartography). Springer, Cham, pp. 293–313. https://doi.org/10.1007/978-3-319-71470-7_15.

Hu, Y., 2018. Geo-text data and data-driven geospatial semantics. Geogr. Compass, e12404. https://doi.org/10.1111/gec3.12404.

Huang, Q., Wong, D.W.S., 2016. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? Int. J. Geogr. Inf. Sci. 30, 1873–1898. https://doi.org/10.1080/13658816.2016.1145225.

Jarić, I., Courchamp, F., Gessner, J., Roberts, D.L., 2016. Data mining in conservation research using Latin and vernacular species names. PeerJ 4, e2202. https://doi.org/10.7717/peerj.2202.

Järv, O., Ahas, R., Witlox, F., 2014. Understanding monthly variability in human activity spaces: a twelve-month study using mobile phone call detail records. Transp. Res. Part C Emerg. Technol. 38, 122–135. https://doi.org/10.1016/j.trc.2013.11.003.

Johnson, J., Karpathy, A., Fei-Fei, L., 2016. DenseCap: fully convolutional localization networks for dense captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'16), pp. 4565–4574.

Joseph, K., Landwehr, P.M., Carley, K.M., 2014. Two 1%s Don't Make a Whole: Comparing Simultaneous Samples from Twitter's Streaming API. Springer, Cham, pp. 75–83. https://doi.org/10.1007/978-3-319-05579-4_10.

Kaplan, A.M., Haenlein, M., 2010. Users of the world, unite! The challenges and opportunities of Social Media. Bus. Horiz. 53, 59–68. https://doi.org/10.1016/j.bushor.2009.09.003.

Karpathy, A., Fei-Fei, L., 2015. Deep visual-semantic alignments for generating image descriptions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), pp. 3128–3137.

Kemp, S., 2018. Digital in 2018 - essential insights into internet, social media, mobile, and ecommerce use around the world [WWW document]. URL https://wearesocial.com/blog/2018/10/the-state-of-the-internet-in-q4-2018.

Kitchin, R., 2014. The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. Sage.

Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D.A., Bernstein, M.S., Fei-Fei, L., 2017. Visual genome: connecting language and vision using crowdsourced dense image annotations. Int. J. Comput. Vis. 123, 32–73. https://doi.org/10.1007/s11263-016-0981-7.

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

Landherr, A., Friedl, B., Heidemann, J., 2010. A critical review of centrality measures in social networks. Bus. Inf. Syst. Eng. 2, 371–385. https://doi.org/10.1007/s12599-010-0127-3.

Lange, P.G., 2007. Publicly private and privately public: social networking on YouTube. J. Comput. Commun. 13, 361–380.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Lee, H., Seo, B., Koellner, T., Lautenbach, S., 2019. Mapping cultural ecosystem services 2.0 – potential and shortcomings from unlabeled crowd sourced images. Ecol. Indic. 96, 505–515. https://doi.org/10.1016/J.ECOLIND.2018.08.035.

Levin, N., Kark, S., Crandall, D., 2015. Where have all the people gone? Enhancing global conservation using night lights and social media. Ecol. Appl. 25, 2153–2167. https://doi.org/10.1890/15-0113.1.

Levin, N., Lechner, A.M., Brown, G., 2017. An evaluation of crowdsourced information for assessing the visitation and perceived importance of protected areas. Appl. Geogr. 79, 115–126. https://doi.org/10.1016/j.apgeog.2016.12.009.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common objects in context. In: Proceedings of the 13th European Conference on Computer Vision (ECCV). Springer International Publishing, Cham, pp. 740–755.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social sensing: a new approach to understanding our socioeconomic environments. Ann. Assoc. Am. Geogr. 105, 512–530. https://doi.org/10.1080/00045608.2015.1018773.

Lomborg, S., Bechmann, A., 2014. Using APIs for data collection on social media. Inf. Soc. 30, 256–265. https://doi.org/10.1080/01972243.2014.915276.

Longley, P.A., Adnan, M., Lansley, G., 2015. The geotemporal demographics of Twitter usage. Environ. Plan. A 47, 465–484. https://doi.org/10.1068/a130122p.

Lunstrum, E., 2017. Feed them to the lions: conservation violence goes online. Geoforum 79, 134–143. https://doi.org/10.1016/J.GEOFORUM.2016.04.009.

Macdonald, D.W., Jacobsen, K.S., Burnham, D., Johnson, P.J., Loveridge, A.J., 2016. Cecil: a moment or a movement? Analysis of media coverage of the death of a lion, Panthera Leo. Animals 6. https://doi.org/10.3390/ani6050026.

Malan, L.-C., 2011. The new conservation debate: the view from practical ethics. Biol. Conserv. 144, 948–957. https://doi.org/10.1016/J.BIOCON.2010.04.001.

Maninis, K.-K., Caelles, S., Pont-Tuset, J., Van Gool, L., 2018. Deep extreme cut: from extreme points to object segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18), pp. 616625. https://doi.org/10.1016/0165-022X(85)90009-0.

Marin, A., Wellman, B., 2014. Social network analysis: an introduction. In: Schott, J., Carrington, P.J. (Eds.), The SAGE Handbook of Social Network Analysis. SAGE Publications Inc.. https://doi.org/10.4135/9781446294413.n2.

Martinez-Harms, M.J., Bryan, B.A., Wood, S.A., Fisher, D.M., Law, E., Rhodes, J.R., Dobbs, C., Biggs, D., Wilson, K.A., 2018. Inequality in access to cultural ecosystem services from protected areas in the Chilean biodiversity hotspot. Sci. Total Environ. 636, 1128–1138. https://doi.org/10.1016/J.SCITOTENV.2018.04.353.

Masse, M., 2011. REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces. O'Reilly Media.

Maxwell, S.L., Fuller, R.A., Brooks, T.M., Watson, J.E.M., 2016. Biodiversity: the ravages of guns, nets and bulldozers. Nature 536, 143–145. https://doi.org/10.1038/536143a.

McCay-Peet, L., Quan-Haase, A., 2017. What is social media and what questions can social media research help us answer. In: Sloan, L., Quan-Haase, A. (Eds.), The SAGE Handbook of Social Media Research Methods. SAGE, pp. 13–26.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In: Proceedings of Advances in Neural Information Processing Systems 26 (NIPS'13), pp. 3111–3119.

Neutens, T., Van de Weghe, N., Witlox, F., De Maeyer, P., 2008. A three-dimensional network-based space–time prism. J. Geogr. Syst. 10, 89–107. https://doi.org/10.1007/s10109-007-0057-x.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. https://doi.org/10.1073/pnas.1719367115.

Palla, G., Derényi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818. https://doi.org/10.1038/nature03607.

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

Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.S., 2018. Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 2227–2237. https://doi.org/10.18653/v1/N18-1202.

Poorthuis, A., Zook, M., 2017. Making big data small: strategies to expand urban and geographical research using social media. J. Urban Technol. 24, 115–135. https://doi.org/10.1080/10630732.2017.1335153.

Poorthuis, A., Zook, M., Taylor Shelton, M.G., Stephens, M., 2016. Using geotagged digital social data in geographical research. In: Key Methods in Geography. Sage, pp. 248.

Preoţiuc-Pietro, D., Volkova, S., Lampos, V., Bachrach, Y., Aletras, N., 2015. Studying user income through language, behaviour and affect in social media. PLoS One 10, e0138717. https://doi.org/10.1371/journal.pone.0138717.

Ramachandram, D., Taylor, G.W., 2018. Deep multimodal learning: a survey on recent advances and trends. IEEE Signal Process. Mag. 34, 96–108. https://doi.org/10.1109/MSP.2017.2738401.

Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S., Waller, S.T., 2017. Exploring the capacity of social media data for modelling travel behaviour: opportunities and challenges. Transp. Res. Part C Emerg. Technol. 75, 197–211. https://doi.org/10.1016/j.trc.2016.12.008.

Rawat, W., Wang, Z., 2017. Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29, 2352–2449. https://doi.org/10.1162/neco_a_00990.

Reihanian, A., Minaei-Bidgoli, B., Alizadeh, H., 2016. Topic-oriented community detection of rating-based social networks. J. King Saud Univ. Comput. Inf. Sci. 28, 303–310. https://doi.org/10.1016/J.JKSUCI.2015.07.001.

Reinecke, L., Trepte, S., 2014. Authenticity and well-being on social network sites: a two-wave longitudinal study on the effects of online authenticity and the positivity bias in SNS communication. Comput. Hum. Behav. 30, 95–102. https://doi.org/10.1016/J.CHB.2013.07.030.

Richards, D.R., Friess, D.A., 2015. A rapid indicator of cultural ecosystem service usage at a fine spatial scale: content analysis of social media photographs. Ecol. Indic. 53, 187–195. https://doi.org/10.1016/j.ecolind.2015.01.034.

Rocha, R.G., Vieira, B.P., Rodrigues, V., Fonseca, C., 2017. Public engagement offers insights on the Eurasian red squirrel distribution. Eur. J. Wildl. Res. https://doi.org/10.1007/s10344-017-1145-y.

Ruths, D., Pfeffer, J., 2014. Social media for large studies of behavior. Science (80- ) 346, 1063–1064. https://doi.org/10.1126/science.346.6213.1063.

Schuette, S., Folk, R.A., Cantley, J.T., Martine, C.T., 2018. The hidden Heuchera: how science Twitter uncovered a globally imperiled species in Pennsylvania, USA. PhytoKeys 96, 87–97. https://doi.org/10.3897/phytokeys.96.23667.

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P., Ungar, L.H., 2013. Personality, gender, and age in the language of social media: the open-vocabulary approach. PLoS One 8, e73791. https://doi.org/10.1371/journal.pone.0073791.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pődör, A., Olteanu-Raimond, A.-M., Rutzinger, M., 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int. J. Geo-Information 5, 55. https://doi.org/10.3390/ijgi5050055.

Sevin, E., 2013. Places going viral: Twitter usage patterns in destination marketing and place branding. J. Place Manag. Dev. 6, 227–239. https://doi.org/10.1108/JPMD-10-2012-0037.

Sherren, K., Smit, M., Holmlund, M., Parkins, J.R., Chen, Y., 2017. Conservation culturomics should include images and a wider range of scholars. Front. Ecol. Environ. 15, 289–290. https://doi.org/10.1002/fee.1507.

Shoval, N., Isaacson, M., 2007. Sequence alignment as a method for human activity analysis in space and time. Ann. Assoc. Am. Geogr. 97, 282–297. https://doi.org/10.1111/j.1467-8306.2007.00536.x.

Sloan, L., Morgan, J., Burnap, P., Williams, M., 2015. Who tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user metadata. PLoS One 10, e0115545. https://doi.org/10.1371/journal.pone.0115545.

Sonter, L.J., Watson, K.B., Wood, S.A., Ricketts, T.H., 2016. Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. PLoS One 11. https://doi.org/10.1371/journal.pone.0162372.

Spalding, M., Burke, L., Wood, S.A., Ashpole, J., Hutchison, J., zu Ermgassen, P., 2017. Mapping the global value and distribution of coral reef tourism. Mar. Policy 82, 104–113. https://doi.org/10.1016/J.MARPOL.2017.05.014.

Sui, D., Goodchild, M., 2011. The convergence of GIS and social media: challenges for GIScience. Int. J. Geogr. Inf. Sci. 25, 1737–1748. https://doi.org/10.1080/13658816.2011.604636.

Takhteyev, Y., Gruzd, A., Wellman, B., 2012. Geography of Twitter networks. Soc. Networks 34, 73–81. https://doi.org/10.1016/j.socnet.2011.05.006.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: assessing the usability of social

media data for visitor monitoring in protected areas. Sci. Rep. 7, 17615. https://doi.org/10.1038/s41598-017-18007-4.

Thelwall, M., 2009. Social network sites. Users and uses. In: Zelkowitz, M. (Ed.), Advances in Computers. Elsevier, pp. 19–73. https://doi.org/10.1016/S0065-2458(09)01002-X.

Tieskens, K.F., Schulp, C.J.E., Levers, C., Lieskovský, J., Kuemmerle, T., Plieninger, T., Verburg, P.H., 2017. Characterizing European cultural landscapes: accounting for structure, management intensity and value of agricultural and forest landscapes. Land Use Policy 62, 29–39. https://doi.org/10.1016/j.landusepol.2016.12.001.

Van Berkel, D.B., Tabrizian, P., Dorning, M.A., Smart, L., Newcomb, D., Mehaffey, M., Neale, A., Meentemeyer, R.K., 2018. Quantifying the visual-sensory landscape qualities that contribute to cultural ecosystem services using social media and LiDAR. Ecosyst. Serv. 31, 326–335. https://doi.org/10.1016/J.ECOSER.2018.03.022.

van Zanten, B.T., Van Berkel, D.B., Meentemeyer, R.K., Smith, J.W., Tieskens, K.F., Verburg, P.H., 2016. Continental-scale quantification of landscape values using social media data. Proc. Natl. Acad. Sci. 113, 12974–12979. https://doi.org/10.1073/pnas.1614158113.

Varol, O., Ferrara, E., Davis, C.A., Menczer, F., Flammini, A., 2017. Online Human-Bot Interactions: Detection, Estimation, and Characterization. (CoRR abs/1703.0).

Venter, O., Sanderson, E.W., Magrach, A., Allan, J.R., Beher, J., Jones, K.R., Possingham, H.P., Laurance, W.F., Wood, P., Fekete, B.M., Levy, M.A., Watson, J.E.M., 2016. Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. Nat. Commun. 7, 12558. https://doi.org/10.1038/ncomms12558.

Walden-Schreiner, C., Leung, Y.F., Tateosian, L., 2018. Digital footprints: incorporating crowdsourced geographic information for protected area management. Appl. Geogr. 90, 44–54. https://doi.org/10.1016/j.apgeog.2017.11.004.

Waldron, A., Mooers, A.O., Miller, D.C., Nibbelink, N., Redding, D., Kuhn, T.S., Roberts, J.T., Gittleman, J.L., 2013. Targeting global conservation funding to limit immediate biodiversity declines. Proc. Natl. Acad. Sci. U. S. A. 110, 12144–12148. https://doi.org/10.1073/pnas.1221370110.

Wang, Z., He, S., Leung, L., 2018. Applying mobile phone data to travel behaviour research: a literature review. Travel Behav. Soc. 11, 141–155.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3, 9. https://doi.org/10.1186/s40537-016-0043-6.

Wellman, B., 2001. Computer networks as social networks. Science 293, 2031–2034. https://doi.org/10.1126/science.1065547.

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. https://doi.org/10.1371/journal.pone.0129785.

Wood, S.A., Guerry, A.D., Silver, J.M., Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. Sci. Rep. 3, 2976. https://doi.org/10.1038/srep02976.

Wu, X., Lindsey, G., Fisher, D., Wood, S.A., 2017. Photos, tweets, and trails: are social media proxies for urban trail use? J. Transp. Land Use 10, 789–804. https://doi.org/10.5198/jtlu.2017.1130.

Wu, Y., Xie, L., Huang, S.-L., Li, P., Yuan, Z., Liu, W., 2018. Using social media to strengthen public awareness of wildlife conservation. Ocean Coast. Manag. 153, 76–83. https://doi.org/10.1016/j.ocecoaman.2017.12.010.

Yan, Y., Kuo, C.-L., Feng, C.-C., Huang, W., Fan, H., Zipf, A., 2018. Coupling maximum entropy modeling with geotagged social media data to determine the geographic distribution of tourists. Int. J. Geogr. Inf. Sci. 32, 1699–1736. https://doi.org/10.1080/13658816.2018.1458989.

Yang, J., Leskovec, J., 2013. Overlapping community detection at scale: a nonnegative matrix factorization approach. In: Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, WSDM '13. ACM, New York, NY, USA, pp. 587–596. https://doi.org/10.1145/2433396.2433471.

You, Q., Cao, L., Jin, H., Luo, J., 2016. Robust visual-textual sentiment analysis: when attention meets tree-structured recursive neural networks. In: Proceedings of the 24th ACM International Conference on Multimedia (MM'16), pp. 1008–1017.

Zellers, R., Yatskar, M., Thomson, S., Choi, Y., 2018. Neural motifs: scene graph parsing with global context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18), pp. 5831–5840.

Zhang, W., Johnson, T.J., Seltzer, T., Bichard, S.L., 2010. The revolution will be networked. Soc. Sci. Comput. Rev. 28, 75–92. https://doi.org/10.1177/0894439309335162.

Zhao, Z., Feng, S., Wang, Q., Huang, J.Z., Williams, G.J., Fan, J., 2012. Topic oriented community detection through social objects and link analysis in social networks. Knowl.-Based Syst. 26, 164–173. https://doi.org/10.1016/j.knosys.2011.07.017.

Zook, M., Barocas, S., boyd, danah, Crawford, K., Keller, E., Gangadharan, S.P., Goodman, A., Hollander, R., Koenig, B.A., Metcalf, J., Narayanan, A., Nelson, A., Pasquale, F., 2017. Ten simple rules for responsible big data research. PLoS Comput. Biol. 13, e1005399. https://doi.org/10.1371/journal.pcbi.1005399.