ILAS: Intrinsic Landscape Assessment System for Landscape Design and Planning in the National Capital Region

Madeline Brown[1], Timothy Murtha[1], Yan Wang[1], Luwei Wang[1]

[1]University of Florida, Florida/USA · madelinebrown@ufl.edu

Abstract: This study assesses the usefulness of social media for identifying public perceptions of intrinsic landscape values for landscape design and planning. We examine the ways individuals and institutions value and publicly discuss landscape perceptions and issues with an aim to inform landscape design and planning. We also assess the spatial distribution, content and sentiments of tweets. Our aim is to develop a methodology for landscape planners and architects to employ social media data to assess landscape values and perceptions in a variety of spatial and cultural contexts. These data could be used to develop design, planning, and conservation priorities, but our initial study offers clear directions for how to better capture and analyze these data for landscape design and planning.

Keywords: Landscape assessment, social media, conservation planning, crowdsourcing data, big data

1 Introduction

Social media is an integral component of contemporary social interactions and mass communications. However, the potential for social media to inform landscape planning and design has not been clearly identified. Recent studies suggest potential opportunities for social media to inform landscape planning through enhancing understanding public spatial and environmental preferences, effective public outreach and communication strategies, as well as the spread of ideas and behaviors through online channels (LEVIN et al. 2015, HAUSMANN et al. 2017, DI MININ et al. 2015, PARSONS et al. 2014, SHIFFMAN 2012, GUERRERO et al. 2016). This study explores a new avenue, namely how Twitter data may be mobilized to understand public perceptions and values around landscape and conservation issues. We aim to identify variation between Twitter topics and sentiments discussed across both spatial and social dimensions, including rural to urban and institutional to individual. This study primarily focuses on the National Capital Region of the United States of America (as defined by the National Park Service), however the methodology developed is intended to be applied in diverse contexts. By creating a novel Intrinsic Landscape Assessment System (ILAS) based on Twitter data, we aim to advance methodology for landscape planners and architects to leverage big data and social media to inform theory and practice in the field.

Landscape conservation planning increasingly involves input from diverse stakeholders and across broad spatial scales. However, traditional field studies and surveys can sometimes be both time consuming and costly, which may limit the level of effective community engagement among some landscape planning initiatives. Social media, including Twitter data, has shown promise as an avenue for gauging public use of protected and unprotected spaces (e. g. LEVIN et al. 2015, HAUSMANN et al. 2017), improving science and conservation communication (DI MININ et al. 2015, PARSONS et al. 2014, SHIFFMAN 2012, DEMETRIOUS 2017, LOVEJOY et al. 2012, RYBALKO & SELTZER 2010), and assessing cultural ecosystem services (FIGUEROA-ALFARO & TANG 2017, OTEROS-ROZAS et al. 2018). Social media also has

Journal of Digital Landscape Architecture, 4-2019, pp. 84-94. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-663-5, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537663009.

potential to inform landscape planning initiatives. In this study, we deploy social media data to assess how public and institutional actors think about and prioritize core landscape values, aesthetics and conservation. For this first effort we are tightly focused on ideas centered on landscape scale conservation design and planning, but are certain that this methodology could be adapted at multiple scales (potentially even the site scale).

2 Methods

This study involves four main phases: 1) data preprocessing, 2) semantic analysis, 3) geographical analysis, and 4) social network analysis. Here we present an initial investigation into the first three phases. To collect data that may be relevant to landscape planners, we created a 275-word query called ILTerms (Intrinsic Landscape Terms), which defines the keywords and terms related to the conservation (Table 1). The ILTerms list was generated based terminological resources from several large conservation and environmental organizations, including the U.S. EPA, the United Nations Environmental Program, and the World-wide Fund for Nature. The ILTerms lexicon was used to crowdsource Tweets that include the pre-defined keywords either in their text or hashtags. The term list was limited to 275 to fit with the Twitter query restrictions.

Table 1: Selection of terms from the 275-word ILTerms query

Selection of terms from the 275-word ILTerms query	
● Landscape ● Air quality ● Conservation ● CITES ● Wildlife	● Overfishing ● Solar energy ● Critical habitat ● Biodiversity ● World Heritage Site

Twitter users share short 140-character messages (this length is in the process of expanding). The texts can include words, URLs, @mentions, hashtags, emoticons, abbreviations, etc. To analyze this data, we clean the texts by removing URL links and user mentions (@), which are not relevant to the core meaning of the text in a large number of tweets. We also remove the special characters that are unnecessary for further analysis. We tokenize the tweets to unigrams based on regular expression patterns. Each tweet is segmented into their constituent words and converted to lowercase. We also remove the stop words that have no significance and words with less than two characters, conduct parts of speech tagging and lemmatization. The Data Preprocessing pipeline was created by a co-author in her previous study and will provide the basis to normalize the ILTerms Tweets.

In this study, we analyzed Twitter data from a two-week period between November 16-30, 2018 (except August 19). The Streaming Data were collected through a Twitter Streaming API in the Urban Complexity and Resilience Lab and was stored in pickle files every hour. In total, the ILTerms keyword query yielded 22,853,399 tweets, 19,327,380 of which were in English. Of the English language tweets, 217,556 contained “Place” attributes. Using the place names, we geocoded the tweets using the Data Science Toolkit’s geodatabase and the ggmaps R package, ultimately yielding 215,084 georeferenced tweets. Place name specificity varied from specific addresses (e. g. National Museum of Natural History), to town names

(e. g. Washington D.C.), to broader spatial locations. The text from these tweets was then cleaned by removing urls, punctuation, stop words, and contractions; primarily with stringr and the tidyverse package suite in R. Maps of tweet distributions were made with ArcMap.

To assess the common themes and semantic patterns in the Twitter data, we calculated the most frequent single words, bigrams (paired words), and trigrams (three sequential words) within the dataset. In order to select the variables that might be most relevant for landscape planners and conservationists, we examined the top 100 words from each list and selected up to 10 potential terms for further analysis.

3 Results

3.1 Text Analysis

Table 2 highlights a selection of some of the most frequent words found in the ~215,000 georeferenced tweets. Several of the common words – such as “climate”, “forest”, and “environment” – were part of the original ILTerms lexicon, while others – such as “people”, “love”, and “time” – were emergent common terms. In addition, we found that some of the tweets included spam, or meaningless tweets, such as job posts and product advertisements. This messy data suggests that the next iteration of the ILTerms Lexicon may benefit from using advanced natural language processing to further extract the semantic meanings of tweet text.

Table 2: Top words and paired words in English language geotagged tweets

RANK	WORD	N	BIGRAM	N
1	Nature	16,815	Climate change	9,010
2	Climate	16,223	Global warming	2,060
3	Forest	11,210	Air quality	1,784
4	Change	11,064	Mother nature	1,295
5	Environment	9,170	Nature conservancy	907
6	People	7,286	Cancer protecting	658
7	Refuse	7,272	Brother nature	630
8	Time	4,884	Hiring careerarc	529
9	Diversity	4,475	Fast paced	522
10	Love	4,447	Forest fires	466

We found that many of the keyword query terms that we identified as referring to specific landscape or conservation issues instead seemed to refer to alternative meanings in the context of Twitter content. For example, the word “environment” was not always used to refer to the natural or built environment, but rather, often referred to the political environment, social environments, or work environments. In addition, words such as “refuse” which was intended to refer to trash or pollution, is more frequently used in tweets as a verb than a noun.

Despite these limitations however, we did find that our query was able to identify current environmental issues – such as the wildfires in California or Trump's tweeting about forest management – as well as visual and recreational landscape values – such as beautiful sunsets, landscape photographs, and hikes. Some example Tweets from our ILTerms query illustrate the range of topics covered in these public microblogs:

> "taking it all in #nofilter #nofiltersneeded #sunset #sky #clouds #beautiful #nature #lake #lakeview @ south lake ta[hoe]"
> "just posted a photo @ marco island nature preserve & bird sanctuary"
>
> "one of our great @wvstateparks to enjoy □□□"
>
> "an admirable cause. As #landscape #architects we agree. trees so important to our lives."

Some of these tweets may be useful for identifying locally relevant environmental issues or valued public assets. For example, aggregating tweets that refer to activities or aesthetics may be a way to identify key geographic locations associated with these values (e. g. kayaking, photography, stargazing).

Table 3: Counts of words frequently paired before or after the following key terms: *environment, landscape, park, quality*. These data are only derived from English-language, geotagged tweets gathered using the ILTerms keyword query

Word 1	**Word 2**	**N**	**Word1**	**Word2**	**N**
Social	Environment	160	Healthcare	Landscape	255
Collaborative	Environment	144	Landscape	Landscapephotography	85
Hostile	Environment	130	Beautiful	Landscape	66
Diverse	Environment	77	Landscape	Photography	64
Learning	Environment	76	Nature	Landscape	63
Forest	Park	209	Air	Quality	1,784
National	Park	153	Quality	Index	70
Park	Forest	121	Quality	Health	33
Wildlife	Park	106	Unique	Quality	32
WY	Park	87	Quality	Drinks	31

3.2 National Capital Region of the United States

A major focus of this project is on the geographic distribution of ILTerms tweets. Figure 1 highlights the spatial distribution of the ILTerms tweets from the two-week period sampled. This figure illustrates the global distribution and representativeness of the study tweets. Ultimately, we aim to assess whether the spatio-temporal distribution of ILTerms can be used to identify areas with contemporary conservation problems or to improve landscape conservation planning.

Our pilot study aims to test the ILTerms assessment system in the National Capital Region of the United States. This is a region with rich cultural and natural resources, located on both public and private land. In addition, as a primarily urban and peri-urban region, the likelihood of tweets occurring on public lands may be greater than is possible in more remote regions and parks. Using the common tweet terms, we found that many of these terms were commonly discussed in the National Capital Region. Figure 2 displays the distribution of tweets with given keywords and bigrams across the National Capital Region. Some of the tweets come from public institutions, such as museums, while others are from the general public. There does seem to be some clustering of the tweets based on their thematic content, but at this stage, the relationship between land tenure, land use, and other spatial data layers and the ILTerms data remains an area for future investigation.

One potential limitation in our data collection is that November is the end of Autumn in the National Capital Region, meaning that fewer people may be outside enjoying public spaces than would occur in the summer. A quick analysis of tweets in warmer weather climates, such as Florida, suggested that more of the ILTerms tweets were actually discussing the physical landscape and experiences of landscapes.

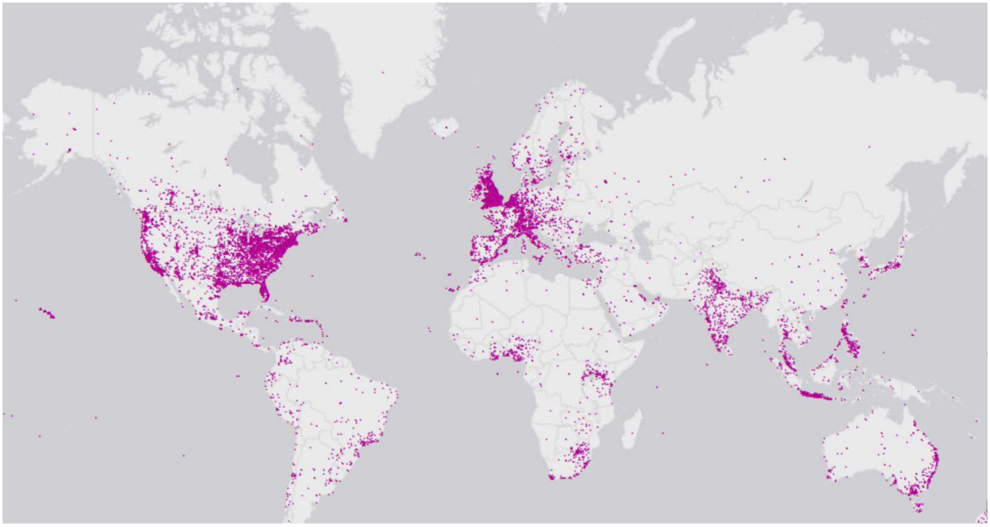

Fig. 1: English language tweets from ILTerms query from November 16-30, 2018

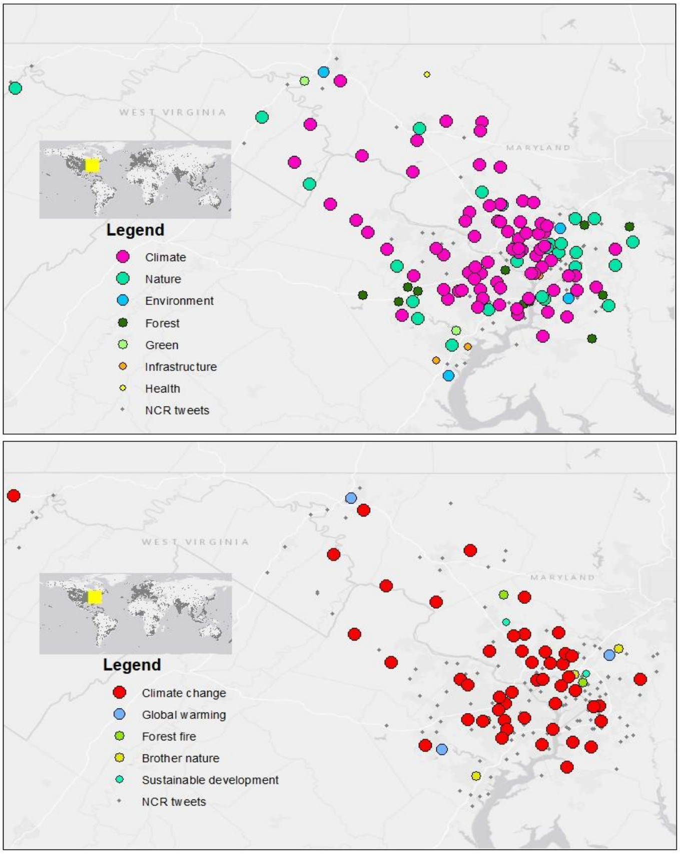

Fig. 2: Distribution of Tweets in the U.S. National Capital Region. Tweets are colored according to keywords in the text. Top map is single keywords and bottom map is bigrams.

3.3 Sentiment and Hashtag Analysis

Assessing landscape values and perception involves sentiment analyses of public experiences of landscapes. To assess the sentiment of ILTerms tweets, we used the NRC Word-Emotion Association Lexicon, which is built into the tidytext R package. Table 4 displays the sentiments included in the NRC Lexicon. Based on analyzing the sentiment of the sampled tweets, words such as "change," "bad," and "government" were associated with fear, while words such as "tree," "cancer," and "hate" were associated with anger. Such assessments of the emotional sentiment of tweets may be of limited utility to landscape planners. Instead, landscape architects may be more interested in aesthetic or sensory experiences of a place, as well as emotive responses to a place that are coupled with specific place-based descriptions. Simply assessing emotional content of tweets without the broader context of the entity that is evoking a particular emotional response may be of limited utility to landscape architects. This remains an area for future inquiry.

Table 4: Sentiments from the NRC Word-Emotion Association Lexicon (NRC 2019)

Sentiments				
Anger	Anticipation	Disgust	Fear	Joy
Negative	Positive	Sadness	Surprise	Trust

Combining multiple words, such as an adjective and a noun (e. g. 'beautiful' and 'landscape' or 'polluted' and 'water') in order to filter and analyze tweets may prove helpful for identifying landscape values. Our current sample size is rather small for the geotagged data, making linking sentiments to geographic locations difficult at this phase.

Hashtags represent shared units of meaning that Twitter users deploy in order to link their tweet to broader social ideas, places, memes, or groups. Some hashtags, such as #optoutside, are also used by businesses like REI as a way to promote brand values. Such tags may also expand into broader public usage as well. There are many diverse hashtags included in the ILTerms tweets, some of which reflect the keyword query directly (e. g. #landscape, #nature), while others may indicate how people are using and valuing landscapes (e. g. #shotoftheday, #getoutstayout, #sunset, #lakeview, or #hiking). Table 5 conveys the top 15 hashtags posted within the sampled tweets. Notably, four of the top ten hashtags are related to job advertisements. Closer inspection of these tweets reveals that many are advertising environmental jobs, explaining why they may have been captured by the keyword query. In future work, it will be important to develop more detailed processes for removing spam and irrelevant tweets from the dataset.

Hashtags can also be analyzed in aggregate, according to which hashtags are most frequently used within the same tweets. We conducted a network analysis of the top 100 hashtags, linked based on whether or not they are used as co-tags in individual tweets. Then we used a spinglass community detection algorithm to identify the clusters of hashtags that are most closely associated with one another. Seven communities of hashtags were identified through this community detection analysis. We then assessed the common pattern among the hashtags within each community to determine which topic these hashtags are generally associated with. Figure 3 conveys these primary themes found in the top 99 hashtags (one was removed due to its lack of ties and relevance) according to their community membership. Employing a network approach to outline hashtag topic clusters allows for rapid identification of salient themes within the ILTerms tweets. In particular, this enables efficient filtering of off-topic or spam tweets, such as those included in the hashtag clusters focused on job advertisement or google business reviews. At the same time, this approach identifies current events, such as Giving Tuesday. These hashtag communities may be highly responsive to temporal changes in hashtag trends. Other hashtag communities such as landscape photography or visual landscape features may be more consistent over time. Further analysis of the temporal and spatial hashtag variation, coupled with thematic community detection analysis may prove useful for identifying which topics are most relevant for landscape assessment across these dimensions. In addition, zeroing in on specific hashtag communities may reduce noise in the data and allow more efficient landscape value assessment.

Table 5: Top 15 hashtags from ILTerms data. Highlighted terms are ones that may be relevant to landscape conservation planning and design

Hashtag	**N**	**Hashtag**	**N**	**Hashtag**	**N**
Nature	3439	Hiring!	891	Travel	628
Job	2424	Careerarc	833	Diversity	521
Hiring	1045	Climatechange	752	Autumn	479
Landscape	938	Sustainability	655	Sunset	465
Photography	907	Green	628	Naturephotography	464

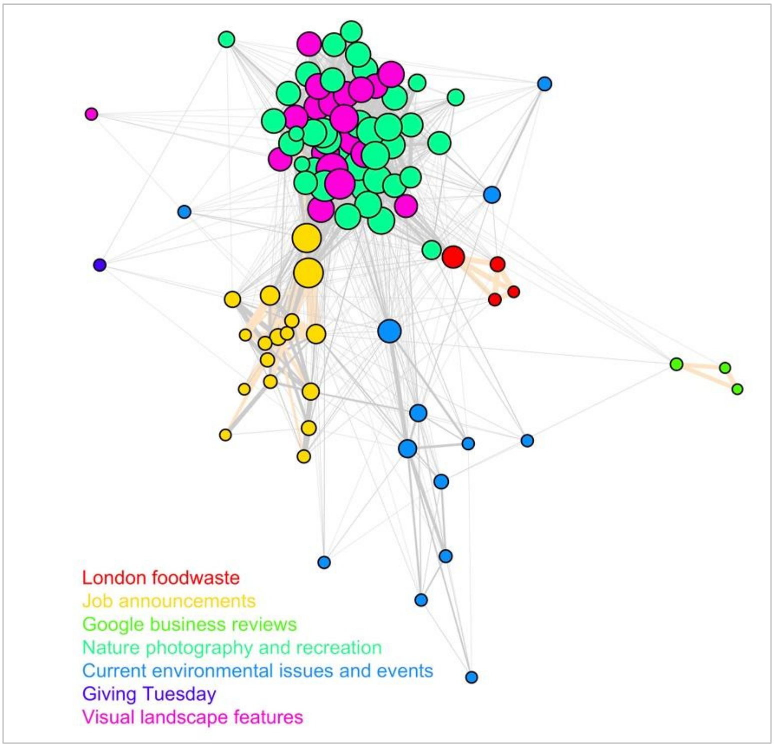

Fig. 3: Co-tagging network of top 100 hashtags. Each node represents a distinct hashtag found in the ILTerms tweets. Nodes are linked when they are both tagged in the same tweet. Edges are weighted, with wider beige edges representing the top 10 % of edges. Nodes are sized according to their degree centrality. Colors represent distinct communities detected through a spinglass community detection algorithm. Each community was then given a descriptive name based on the common themes among the included tags.

4 Discussion and future directions

4.1 Discussion

Social media is ubiquitous in many spheres of modern life, yet its potential to influence design and planning has not been clearly identified. This study investigates the usefulness of Twitter as a data source for investigating intrinsic and aesthetic landscape values and perceptions. Creating a novel lexicon of aesthetic landscape terms, we collected landscape-related tweets over an almost complete two-week period. We analyzed the distribution and content of these tweets to assess public perceptions of landscape values across social and spatial contexts. Initial assessment found that many tweets discussed specific environmental issues and landscape features. The tweets examined in this study cannot be adequately understood without an understanding of the broader political and social context in which they were created. During the time period sampled (November 16-30 2018), a National Climate Assessment (USGCRP 2018) was released, massive wildfires were occurring in California, and the President of the United States made a highly publicized and controversial tweet about Finland raking their forest floors (BBC 2018). Each of these events is likely to have impacted the Twittersphere, given the responsiveness of this medium to current events. Indeed, air quality, forest fires, and forest floors are all common bigrams appearing in our data sample, likely related to the latter two events.

LEVIN and colleagues (2015) combined spatial data of night sky lights and photo-sharing to identify locations of high human activity and recreation in both public and private lands. This approach allowed them to identify existing protected areas with high levels of social media activity as well as "photography hotspots" in unprotected non-urban areas (LEVIN et al. 2015). Similarly, GUERRERO and colleagues (2016) identified urban nature hotspots in Copenhagen using Instagram data. Using geotagged Twitter data may similarly allow landscape values and recreational hotspots to be identified. By specifically assessing tweets within national parks or other public lands, it will be possible to rapidly identify public sentiments and recreational use of these landscapes.

4.2 Outlook

Our pilot study examines both tweet content and spatial distribution. Initially, we plan to further assess the scale of content mentioned in landscape tweets, including whether the tweets primarily focus on issues of local, national, or global concern. The next step in our analysis will be to integrate georeferenced data with additional social and environmental spatial data layers to compare the distribution of conservation tweets to other variables such as: 1) public and private lands, 2) urban and rural areas, and 3) land-use types. We will also explore the spatial and temporal pattern of sentiment of the collected tweets as well as tweets related to certain topics or issues. Understanding how people value specific conservation actions and landscapes will be important for prioritizing landscape conservation and planning actions.

References

AGARWAL, S. D., LANCE BENNETT, W., JOHNSON, C. N. & WALKER, S. (2014), A Model of Crowd-Enabled Organization: Theory and Methods for Understanding the Role of Twitter in the Occupy Protests. International Journal of Communication, 8, 646-672.

BBC (2018), California wildfires: Finland bemused by Trump raking comment. November 19. https://www.bbc.com/news/world-europe-46256296.

BENNETT, W. L. (2012), The Personalization of Politics: Political Identity, Social Media, and Changing Patterns of Participation. Eds. D. V. SHAH, L. A. FRIEDLAND, C. WELLS, Y. MIE KIM & H. ROJAS. The ANNALS of the American Academy of Political and Social Science, 644 (1), 20-39. https://doi.org/10.1177/0002716212451428.

BLEI, D. M., NG, A. Y. & JORDAN, M. I. (2017), Latent dirichlet allocation. Journal of machine Learning research, 3, 993-1022.

DEMETRIOUS, K. (2017), Contemporary Publics, Twitter and the Story of PR: Exploring Corporate Interventions to Promote 'Clean Coal' in Australia. Contemporary Publics 1. 20.

DI MININ, E., HENRIKKI, T. & T. TOIVONEN, T. (2015), Prospects and Challenges for Social Media Data in Conservation Science. Frontiers in Environmental Science 3 (September 9). https://doi.org/10.3389/fenvs.2015.00063.

FIGUEROA-ALFARO, R. & TANG, Z. (2017), Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on Panoramio and Flickr. Journal of Environmental Planning and Management, 60 (2), 266-281. doi: 10.1080/09640568.2016.1151772.

GUERRERO, P., MOLLER, M., OLAFSSON, A. & SNIZEK, B. (2016), Revealing Cultural Ecosystem Services through Instagram Images: The Potential of Social Media Volunteered Geographic Information for Urban Green Infrastructure Planning and Governance. Urban Planning, 1 (2), 1-17.

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: Social Media Data in Protected Areas. Conservation Letters, 11 (1), e12343.
https://doi.org/10.1111/conl.12343.

LEVIN, N., KARK, S. & CRANDALL, D. (2015), Where Have All the People Gone? Enhancing Global Conservation Using Night Lights and Social Media. Ecological Applications, 25 (8), 2153-2567. https://doi.org/10.1890/15-0113.1.

LOVEJOY, K., WATERS, R. D. & SAXTON, G. D. (2012), Engaging Stakeholders through Twitter: How Nonprofit Organizations Are Getting More out of 140 Characters or Less. Public Relations Review, 38 (2), 313-318. https://doi.org/10.1016/j.pubrev.2012.01.005.

MAURYA, A. et al. (2016), Semantic scan: detecting subtle, spatially localized events in text streams. arXiv preprint arXiv:1602.04393.

NIELSEN, F. Å. (2011), A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.

NRC Emotion Sentiment Lexicon (2019), http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.

OTEROS-ROZAS, E., MARTÍN-LÓPEZ, B., FAGERHOLM, N., BIELING, C. & PLIENINGER, T. (2018), Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. Ecological Indicators, 94 (2), 74-86. https://doi.org/10.1016/j.ecolind.2017.02.009.

PARSONS, E. C. M., SHIFFMAN, D. S., DARLING, E. S., SPILLMAN, N. & WRIGHT, A. J. (2014), How Twitter Literacy Can Benefit Conservation Scientists: Editorial. Conservation Biology, 28 (2), 299-301. https://doi.org/10.1111/cobi.12226.

RIBEIRO, F. N, ARAÚJO, M., GONÇALVES, P., BENEVENUTO, F. & GONÇALVES, M. A. (2015), SentiBench – a benchmark comparison of state-of-the-practice sentiment analysis methods. arXiv preprint arXiv:151201818.

RYBALKO, S. & SELTZER, T. (2010), Dialogic Communication in 140 Characters or Less: How Fortune 500 Companies Engage Stakeholders Using Twitter. Public Relations Review, 36 (4), 336-341. https://doi.org/10.1016/j.pubrev.2010.08.004.

SHIFFMAN, D. S. (2012), Twitter as a Tool for Conservation Education and Outreach: What Scientific Conferences Can Do to Promote Live-Tweeting. Journal of Environmental Studies & Sciences, 2 (3), 257-262. https://doi.org/10.1007/s13412-012-0080-1.

U.S. GLOBAL CHANGE RESEARCH PROGRAM (USGCRP) (2018), 4th National Climate Assessment (November 23). https://nca2018.globalchange.gov/.

WANG, Y. & TAYLOR, J. E. (2018), Detecting Urban Emergencies Technique (DUET): A Data-Driven Approach based on Latent Dirichlet Allocation (LDA) Topic Modeling. Journal of Computing in Civil Engineering (in Press).

WOOD, S. A., D. GUERRY, A., SILVER, J. M. & LACAYO, M. (2013), Using Social Media to Quantify Nature-Based Tourism and Recreation. Scientific Reports, 3 (1). https://doi.org/10.1038/srep02976.

XIE, W., ZHU, F., JIANG, J., LIM, E.-P. & WANG, K. (2016), Topicsketch: Real-time bursty topic detection from twitter. IEEE Transactions on Knowledge and Data Engineering, 28, 2216-2229.