Mapping Landscape Values with Social Media

Madeline Brown[1], Timothy Murtha[2], Luwei Wang[2], Yan Wang[2]

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

Abstract: This paper investigates public landscape values and perceptions in the Eastern United States using crowdsourced data from Twitter. Tweets contain information about how people use, value, and think about their environments. We collected two months of tweets (December 2018 and January 2019) using a keyword list focused on conservation and environment terms. These data are analyzed across four contexts based on location (Global, US national capital region, and Orlando, Florida) or keywords (National Parks). We find that Twitter data enables time-sensitive landscape perceptions over a large spatial extent that may be of interest to landscape planners. In addition, we compare landscape assessments across multiple contexts and examine landscape values in culturally relevant places, such as parks, scenic vistas, and waterways. Finally, we discuss future methodological and theoretical directions for landscape architects and planners interested in incorporating social media into research and design processes.

Keywords: Landscape assessment, landscape values, GIS, social media, Twitter

1 Introduction

Assessing how the public perceives and uses landscapes is a critical component of large landscape conservation design and planning. Knowing which places are culturally valued can inform planning and funding decision-making. However, assessing public perceptions over large spatial extents can be time consuming and resource intensive. Increasingly, researchers are turning to social media as a data source for understanding human behavior and beliefs. For landscape architects and planners, social media and crowdsourced data has been used to understand landscape preferences, disaster responses, and visitation rates (LEVIN et al. 2015, LEVIN et al. 2017, LANGEMEYER et al. 2018, FISHER et al. 2018). Advantages of social media are that large volumes of data can be collected rapidly with little cost other than data storage and processing. Our previous work identified the applicability of Twitter data for landscape architects and land managers working with cultural and natural resources (BROWN et al. 2019, BROWN & MURTHA 2019). Here we expand on that work with a larger dataset, clearer methods for improving data usability, and specific case studies of landscape assessments across multiple contexts.

We assessed tweets related to landscape and environmental values across three spatial extents: 1) global, 2) the national capital region of the United States, and 3) the city of Orlando, Florida, USA. In addition, a third subset based on keywords related to national parks was also examined. The two case study regions were selected as regions with notable cultural and natural resources. The National Capital Region (NCR) of the United States is home to many nationally important monuments, memorials, and scenic areas. The city of Orlando is most famous for its suite of amusement parks and attractions, notably Disneyworld and Universal Studios, but is also home to numerous urban parks and waterways. Comparing the spatial distribution and content of tweets across these two contexts offers a window into how landscape values are expressed in social media across regional contexts.

Journal of Digital Landscape Architecture, 5-2020, pp. 542-548. © Wichmann Verlag, VDE VERLAG GMBH · Berlin · Offenbach. ISBN 978-3-87907-690-1, ISSN 2367-4253, e-ISSN 2511-624X, doi:10.14627/537690055.

2 Methods

This paper draws on Twitter data gathered between December 2018 through January 2019. Data collection was conducted with the Intrinsic Landscape Assessment System (ILAS) 275-word keyword query (BROWN et al. 2019). This keyword query contains 275 words related to conservation and environmental topics. These terms were compiled primarily based on three glossaries and word lists from the United States Environmental Protection Agency (EPA), the Worldwide Fund for Nature (WWF) and the United Nation's Environmental Program (UNEP). The keyword query returned detailed information about each tweet including the text, date and time, username, user location, tweet location, and hashtags. Data scraping was conducted with Python. The majority of data analysis and visualization was completed with R. Here, we highlight a few core R packages. Data cleaning was conducted with the R packages tidyverse and stringr, data visualization and analysis was conducted using ggplot2, leaflet and other packages (ARNOLD et al. 2019, GRAUL 2016, WICKHAM 2016, 2019; WICKHAM et al. 2019, R CORE TEAM 2019, RAM et al. 2018). Text analysis was conducted with the R packages tidytext and cleanNLP (SILGE & ROBINSON 2016, ARNOLD 2017).

3 Results

During the two-month data collection period, 789,602 of the total tweets scraped were usable for our analysis. These tweets were selected based on their attributes of being in English and having coordinates or an identifiable place name that could be geocoded. Our results are organized into two sections discussing 1) text analysis and 2) spatial analysis.

3.1 Text Analysis

Data processing and bot removal processes are an important part of working with social media data. Data quality is highly variable, and despite using a keyword filter, not all tweets will be relevant for landscape planners. Initially, tweet text is standardized by making all text lowercase, removing emojis and extra spaces, and standardizing punctuation characters. Text analysis begins by analyzing the individual words within tweets. When tweets are tokenized, there are 680,487 unique words in the global two-month dataset. Not all of these words are meaningful for analysis and are considered *stopwords*. These include words such as "a", "the", "his", "puts" and other words that have been identified by common lexicons. We removed 1,149 stopwords from the SMART, snowball and onix lexicons contained in the stop_words dataset from the tidytext R package (SILGE & ROBINSON 2016). For top words analysis we also removed numbers and ampersands.

Table 1: Dataset comparisons

	Global	Orlando	NCR	NPS
Total tweets	789,602	2,587	16,127	2,457
Unique words	680,487	10,312	37,169	11,826
Unique words (no stopwords)	679,809	9,766	36,528	11,291

We analyzed the top words across four contexts: globally, in the US national capital region (NCR), in Orlando, and within tweets referring to national parks (NPS). Data subsets for Orlando and the National Capital Region (NCR) were created by selecting only tweets that fell within these particular geographic regions. This final dataset (NPS) was created by selecting all the tweets mentioning the words "national park", "nationalpark" or "nps" in the text. Table 1 displays the total number of tweets and unique words across each of the contexts. The Orlando and NPS datasets are of similar size in terms of number of tweets though the NPS data contains more unique words. The NCR subset contains the most tweets of the subsets, but is still notably smaller than the total global dataset.

To compare variation and patterns in themes across the four contexts we analyzed the top 50 words from each data subset (Figure 1). In total, there are 121 unique words from the combined top 50 words from each context. In each subset, "nature" occurs within the top three most frequently used words, while the terms further down each frequency list vary.

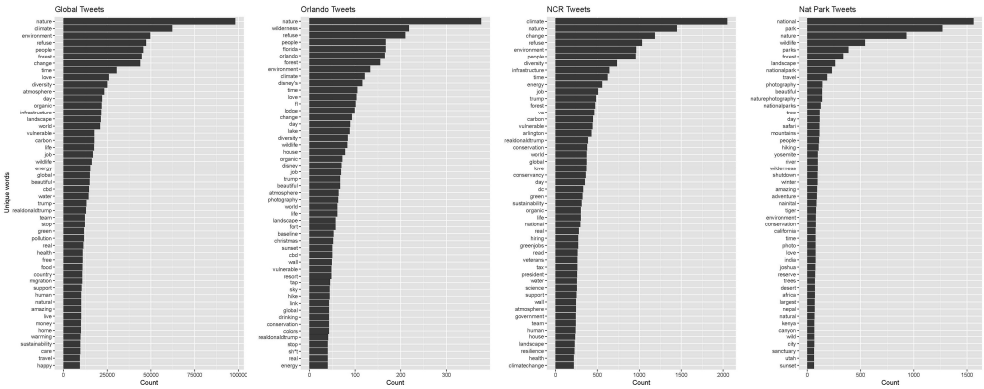

Fig. 1: Top 50 words in Tweets across four contexts (stopwords removed)

In addition to compiling the top words, we also compared the similarities and differences across wordlists using classical multidimensional scaling to reduce differences to two dimensions (with R Stats Package [R CORE TEAM 2019]). Figure 2 displays the results of this analysis, with labels colored by the number of lists in which they appear and jittered to avoid overlap. This approach reveals four clusters of words only found in one context, which aligns with the original four context sample. However, the NCR and Global wordlists were more similar to one another than to the other lists, while NPS and Orlando wordlists were more dissimilar to the other data subsets. This approach enables themes within landscape tweets to be filtered based on whether they are common topics across contexts or specific to particular locations. Comparing across Orlando and the National Capital Region does reveal some differences in common words that that are attributable to differences in local issues or geography. For example, the Orlando wordlist contains place names such as "Florida" and "Orlando" but also references to tourist attractions such as "Disney" and "resort." By contrast, tweets related to national parks contain far more words about natural landscape features such as mountains, rivers, and canyons, as well as specific names of national parks in the United States. In addition, "realdonaldtrump" and "job" are both common across the global, NCR, and Orlando distributions but not to the NPS tweets, indicating these are not major discussion

points among people tweeting about national parks. An area for future analysis will be to assess spatial correlations in sentiment and topics within and outside of public parks.

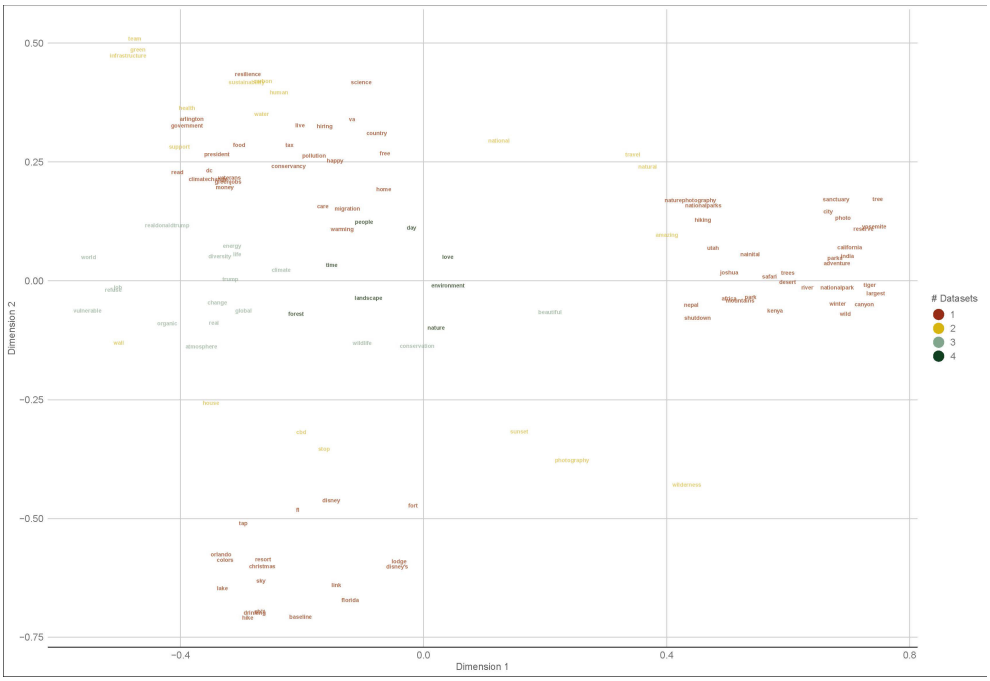

Fig. 2: MDS of top 50 words across four contexts

3.2 Mapping Landscape Tweets

Some tweets contain information about their location or direct coordinates and can be mapped. Tweets containing terms related to recreation and landscape are mapped in Figures 3 and 4. Figure 3 illustrates the ability to map distinct keywords across the landscape to highlight patterns in tweet topics across different locations.

Tweets contain diverse types of information and themes. Some tweets are highly personal, others are broadly political, and still others aim to advertise. This diversity may make it difficult to isolate tweets of interest to landscape planners. However, if particular keywords or hashtags are queried, it is possible to extract a more targeted subset of tweets. Even with the 275-word query, we found that many of the tweets might be relevant for land managers seeking to understand public perceptions and use of local landscapes. For example, Figure 4 shows the distribution of 2 months of tweets that refer to hiking, biking, and trails and which are located in the Orlando area. A targeted query like this may demonstrate which trails are being used for particular kinds of activities. Similar queries about photography, canoeing, birdwatching, or other recreational activities could be used to document the diversity of activities occurring on public lands.

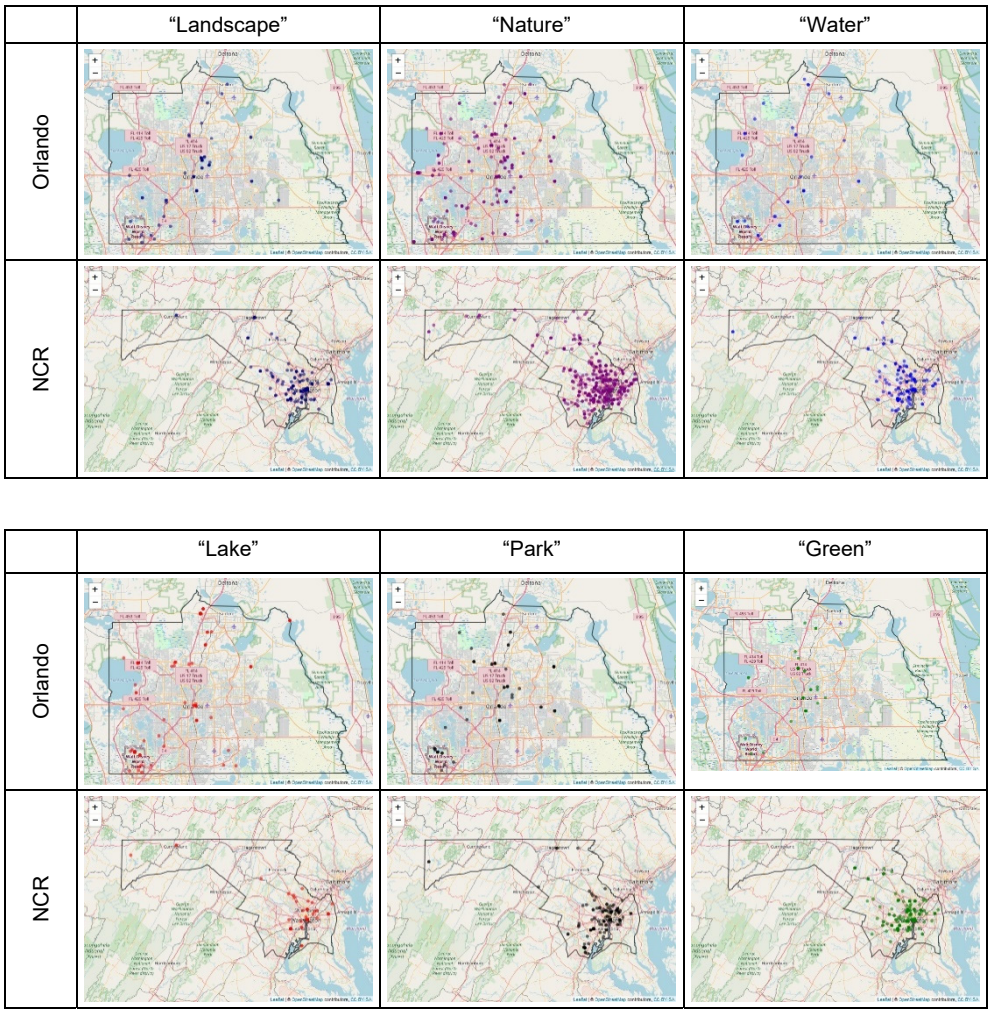

Fig. 3: Maps of recreation and landscape related tweets in Orlando and the National Capital Region. Tweets are subset based on the presence of each term in tweet text.

Events are also often documented in twitter data. For example, one tweet from Four Mile Run Park states: "wow! 717 lbs of litter collected today in the sunnyside stream by 2 dozen volunteers!! #cleanstreamextreme #alexparklove". Temporal analysis of tweets within a particular park could be a useful tool to document all the various activities (both formal and informal) that occur within park borders year-round.

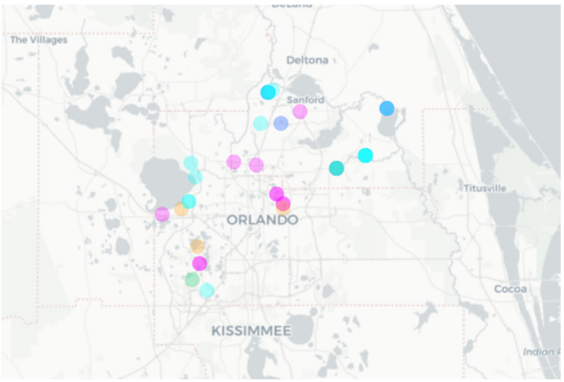

Fig. 4:
This map of Orlando contains tweets with the words "hike/hiking" (pink), "bike/biking" (orange) and "trail" (blue). Darker circles have multiple tweets while lighter circles have fewer. Notably, some of the tweets are clustered along the shoreline of a lake where there is a public trail. The ones near the top right of the lake refer to a trail while those on the lower edge of the lake refer to hiking and biking.

4 Discussion and Future Directions

We suggest landscape architects, planners, and land managers can benefit from analyzing twitter data in several ways. Social media data enables the identification of the following entities: 1) Points of interest, 2) Problems or issues in infrastructure, 3) Ideas for infrastructure development, 4) Landscape functions or visitor activities, 5) Local events, 6) Species diversity. For example, searching for tweets about signage, trails, scenic vistas, or other infrastructure in public lands can be a tool for land managers to rapidly assess visitor perceptions of park features. In addition, tweets often mention wildlife sightings and points of interest, both of which could be useful for land managers hoping to document park assets. However, scraping global twitter data may be impractical for most land managers and planners. Instead, developing ways for local tweets to be rapidly assessed would be more useful for these user groups.

The importance of social media for understanding public perceptions and values is well recognized. Moreover, land managers and planners are increasingly turning to crowdsourcing and online tools for understanding visitor preferences and volumes in public lands. Our study expanded upon a two-week pilot study using the same keyword query focused on environment and conservation terms. We found that over two months, many of the tweets collected contain themes that might be helpful for land managers and planners. Further work in this area might benefit from considering topic modeling and analyzing seasonal variation in tweets discussing landscape features.

References

ARNOLD, T. (2017), A Tidy Data Model for Natural Language Processing using cleanNLP. The R Journal, 9 (2), 1-20. https://journal.r-project.org/archive/2017/RJ-2017-035/index.html.

ARNOLD, J., DAROCZI G., WERTH, B. et al. (2019), ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'. http://github.com/jrnold/ggthemes.

BROWN, M. & MURTHA, T. (2019), Anthropological Approaches for Cultural Resource Conservation Design and Planning. Environmental Practice 21 (4), 179-188. https://doi.org/10.1080/14660466.2019.1688083.

BROWN, M., MURTHA, T., WANG, Y. & WANG, L. (2019), ILAS: Intrinsic Landscape Assessment System for Landscape Design and Planning in the National Capital Region. Journal of Digital Landscape Architecture, 4-2019, 84-94.

FISHER, D. M., WOOD, S. A., WHITE, E. M. et al. (2018), Recreational Use in Dispersed Public Lands Measured Using Social Media Data and On-Site Counts. Journal of Environmental Management, 222 (September), 465-474. https://doi.org/10.1016/j.jenvman.2018.05.045.

GRAUL, C. (2016), leafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. R package version 0.4-0. http://cran.r-project.org/package=leafletR.

LANGEMEYER, J., CALCAGNI, F. & BARÓ, F. (2018), Mapping the Intangible: Using Geolocated Social Media Data to Examine Landscape Aesthetics. Land Use Policy, 77 (September), 542-552. https://doi.org/10.1016/j.landusepol.2018.05.049.

LEVIN, N., KARK, S. & CRANDALL, C. (2015), Where Have All the People Gone? Enhancing Global Conservation Using Night Lights and Social Media. Ecological Applications, 25 (8), 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. Applied Geography, 79 (February), 115-126. https://doi.org/10.1016/j.apgeog.2016.12.009.

MANNING, C. D., SURDEANU, M., BAUER, J. (2014), The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, 55-60.

R CORE TEAM (2019), R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org.

RAM, K., WICKHAM, H., RICHARDS, C. & BAGGETT, A. (2018), wesanderson: A Wes Anderson Palette Generator. https://github.com/karthik/wesanderson.

SILGE, J., ROBINSON, D. (2016), tidytext: Text Mining and Analysis Using Tidy Data Principles in R. JOSS, 1 (3). http://dx.doi.org/10.21105/joss.00037.

WICKHAM, H. (2016), ggplot2: Elegant Graphics for Data Analysis. Springer, New York. https://ggplot2.tidyverse.org.

WICKHAM, H. (2019), stringr: Simple, Consistent Wrappers for Common String Operations. https://github.com/tidyverse/stringr.

WICKHAM, H. et al. (2019), Welcome to the tidyverse. Journal of Open Source Software, 4 (43), 1686. https://doi.org/10.21105/joss.01686.