



The Effect of COVID-19 on the Origins of Florida State Park Visitors and Online Reviewers

Innocensia Owuor¹, Hartwig H. Hochmair¹, and Gernot Paulus²

¹ Geomatics Sciences, University of Florida, Florida, USA

² Spatial Information Management, Carinthia University of Applied Sciences, Villach, Austria

Correspondence: Innocensia Owuor (innocensia.owuor@ufl.edu)

Abstract. The COVID-19 pandemic has caused substantial socioeconomic disruptions and affected travel behaviour at various geographic scales. Public parks were some of the few recreational spaces that people could visit while adhering to the COVID-19 mitigation measures such as social distancing, travel restrictions, or curfews. This study draws data from profiles of TripAdvisor and Yelp users who reviewed Florida state park as well as origins of state park visitors from SafeGraph patterns data to compare the spatial distribution of reviewer and visitor origins before (2019) and during (2020) the pandemic. Results show an increase in the proportion of reviewers and visitors from within the State of Florida during the pandemic as well as shorter distance between reviewer/visitor origin states and Florida. This reflects reduced mobility at the national scale due to COVID-19 related travel restrictions but also highlights the important role of outdoor recreational areas, such as parks, for local communities during such crises.

Keywords. social media, TripAdvisor, Yelp, parks.

1 Introduction

Social media data is prominently employed in various tourism studies as it provides insights such as visitor travel patterns (Hasnat and Hasan, 2018). Traditionally, visitor data has been gathered through surveys, on-site questionnaires, and interviews but these methods are time and labour intensive, and not spatially exhaustive (Arkema et al., 2021). Therefore, social media and online review data have gained interest for the analysis of visitation patterns of outdoor recreational areas and other points of interest (POI) over the years (Tenkanen et al., 2017).

TripAdvisor and Yelp are popular online review platforms for travel planning. Globally, TripAdvisor has over one billion reviews (TripAdvisor, 2022) and Yelp boasts almost a quarter billion reviews (Statista, 2021). TripAdvisor reviews have been positively correlated with official tourist visitation statistics (Ma and Kirilenko, 2021) and Yelp reviews have been used to make recommendations to tourists (Rossetti et al., 2016). Other novel data sources such as SafeGraph, which is based on aggregated data such as mobile phone precise locations and advertisements identifiers (Google android and Apple iOS IDs) from mobile application developers (SafeGraph, 2021), has been used in the analysis of visitation patterns to different attractions (Juhász and Hochmair, 2020). SafeGraph Places has over 11 million POIs in 200 countries, including foot traffic data to POIs in the US, Canada, and the UK (Berryman, 2021).

Information about tourist origins and destinations is important for marketing, sustainable management of tourist attractions, and for analyzing the impact of events on travel (Cirer-Costa, 2017). The user profile on social media applications often contains the user's stated origin (home) location which can be specified at different granularity levels such as city, state, or country. Studies that leverage data from social media platforms, such as Twitter, pointed out that these locations may not paint a true picture of the user's origins (home) location (Hecht et al., 2011). However, there are methods that can be used to validate these locations, such as the analysis of the Twitter user timeline and network (Zheng et al., 2018). Analysis of three separate social media communities (Twitter, Flickr, Swarm) showed that the localness assumption, i.e., that social media users can be considered locals from everywhere they post geotagged content, is true in only about 75% of cases (Johnson et al. 2016). More specifically, only about 53% of Flickr users and

23% of Wikipedia users contribute content that is 100 km or less from their specified home location (Hecht and Gergle, 2010). Analyzing SafeGraph visitation data for three Florida cities showed that the median distance between a POI and the origin of its visitors varies by season and POI type and that travel accommodation and amusement parks are associated with longest median distances from home (Juhász and Hochmair, 2020).

The COVID-19 pandemic widely impacted healthcare, economic and social aspects of our society through measures implemented to curb the virulence of the disease (Ebrahim et al., 2020). With the closure of many indoor recreational facilities such as gyms or indoor swimming pools, parks were one of the few spaces that the public could use while still following social distancing guidelines. Various studies identified increased visitation numbers to parks during the COVID-19 pandemic as people used them for physical exercise and mental health care (Volenc et al., 2021).

This study infers origins of Florida state park visitors from (a) the user profiles of reviewers of Florida state parks on TripAdvisor and Yelp, and (b) from SafeGraph visitation patterns to parks. It also compares the spatial distribution of the reviewer and visitor origins (at the U.S. state level) before and during the COVID-19 pandemic and the associated mean distances between reviewer/visitor origins (generalized to state centroids) and Florida. A similar evaluation can be conducted using country level data to assess impact of the pandemic on the origins of international visitors. The periods of before and during the pandemic for TripAdvisor and Yelp are defined as (March 20 to December 31, 2019) and (March 20 to December 31, 2020) respectively. For SafeGraph, due to data abundance, the before and after periods of the pandemic were chosen to be shorter, namely as (March 1 to April 30, 2019) and (March 1 to April 30, 2020), respectively. The following three objectives were addressed in this study:

1. Determine the spatial distribution of origins of reviewers (TripAdvisor, Yelp) and visitors (based on SafeGraph data) of Florida state parks at the U.S. state level.
2. Use chi-square tests on 2x2 crosstab tables to determine the association between the ratio of in-state (local) vs. out-of-state park visitors/reviewers and pre-pandemic vs. during pandemic periods.

3. Model the distribution of the distances between the park visitor/reviewer origin state and Florida using a power law approximation.

2 Study setup

2.1 Study area

The study area comprises of 177 state parks in Florida whose locations are generalized in (Fig. 1). The shapefiles of the park boundaries were downloaded from the Open Data portal of the Florida Department of Environmental Protection (DEP). The number of parks analyzed (Table 1) for the three different data sources depended on data availability for each source.



Fig. 1 Florida state parks

2.2 Data collection

TripAdvisor and Yelp have APIs which limit the number of reviews that can be retrieved. Therefore, we developed a python script which scraped reviews and their respective user profile data from the TripAdvisor and Yelp Websites for the study period. SafeGraph data, which was downloaded from the SafeGraph website, included park locations and the distribution of U.S. states of origins (homes) of the park visitors. A summary of the data counts is captured in Table 1, separated by data source and pre-pandemic (2019) vs. pandemic (2020) periods.

Table 1. Total number of parks, online reviewers, and visitor origins summary

Data source	Number of state parks	Number of reviewers/visitors (2019)	Number of reviewers/visitors (2020)
TripAdvisor	135	2595	867
Yelp	95	520	384
SafeGraph	125	6048	4341

2.3 Data availability

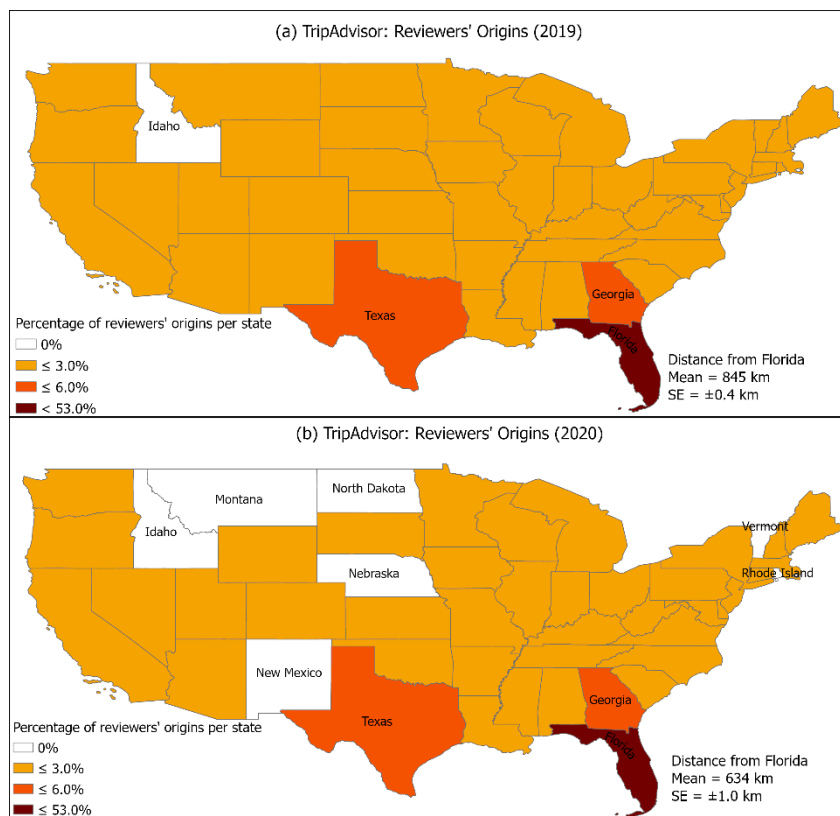
The datasets used for this research are available for download from <https://github.com/InnocensiaO/AGILE-2022>.

3 Analysis of results

The spatial granularity of the origin information provided in TripAdvisor and Yelp user profiles was inconsistent and occurred at the city or state level. Therefore, to retain a sufficient sample size, the distances were calculated between the reviewer/visitor origin state centroid coordinates and the Florida state centroid location, which were the basis for a power-law approximation. For the

latter, logarithmic binning was used to generate log-log plots. For reviewers/visitors from the state of Florida, the distance from home (origin) to park was zero. To avoid complexities during the log transformation, a constant value of ten was used instead of zero in the power-law model. The following figures show the proportion of origin (home) states of park reviewers and visitors for 2019 and 2020 based on TripAdvisor (Fig. 2), Yelp (Fig. 3) and SafeGraph (Fig. 4) data, respectively.

In 2019, TripAdvisor reviewers from all states in the contiguous US except for Idaho reviewed Florida parks, whereas in 2020 no reviews were shared from seven states. The mean distance between reviewer/visitor origin state centroid and Florida dropped from 845 km before the pandemic in 2019 to 634 km during the pandemic in 2020.

**Fig. 2** Florida state park reviewers' origins on TripAdvisor in 2019 (a) and 2020 (b)

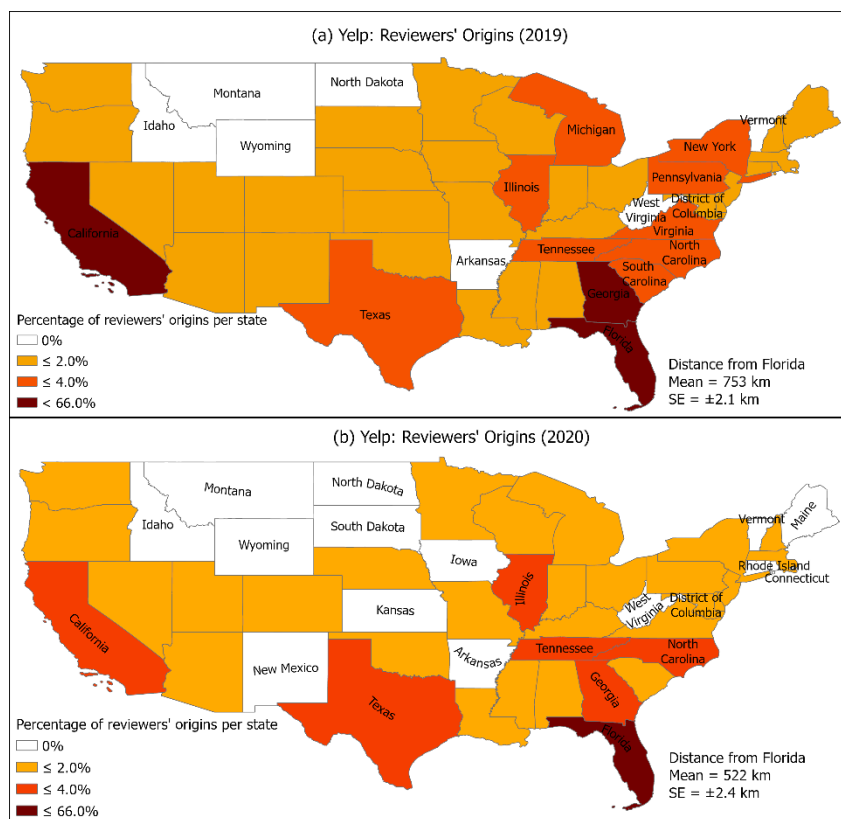


Fig. 3 Florida state park reviewers' origins on Yelp in 2019 (a) and 2020 (b)

Similarly, in Yelp the number of states from which state park reviews were posted reduced during the pandemic, namely from 41 states in 2019 to 34 states in 2020. The mean distance between reviewer/visitor origin state centroid and Florida dropped from 753 km before the pandemic in 2019 to 522 km during the pandemic in 2020.

SafeGraph detected Florida state park visitors from all 49 states in the contiguous US both in 2019 and 2020.

Despite this similarity, the proportion of local (Florida) residents among all visitors of Florida state parks increased from 51.1% (pre-pandemic) to 59.4% (during pandemic). The mean distance between visitor origin state centroid and Florida dropped from 673 km before the pandemic in 2019 to 568 km during the pandemic in 2020.

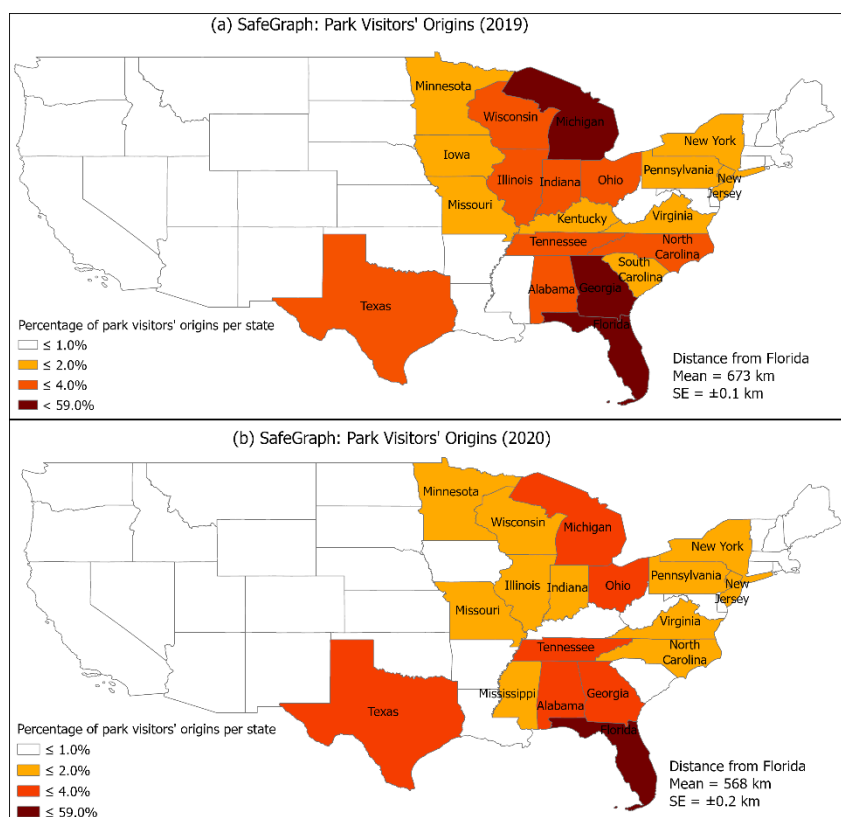


Fig. 4 Florida State Park visitors' origins based on SafeGraph data in 2019 (a) and 2020 (b)

For all three data sources, the share of reviewers and visitors of Florida state parks located in Florida increased during the pandemic. The increase ranged between 8.3% (SafeGraph) and 10.6% (TripAdvisor). (Table 2). These results are in line with previous studies which showed local residents visited parks more during the COVID-19 pandemic (Volenc et al. 2021). A Chi square test of independence ($df = 1$) showed that for each data source the proportion of local vs. out-of-state reviewers/visitors was significantly associated with the analysis period as indicated by the χ^2 and p values in (Table 2).

This means that the pandemic caused significant changes on the origins of reviewers and visitors of Florida state parks and thus affected travel behaviour. Whereas the latter is more obvious from SafeGraph data which provides information on visitation pattern, we also assume that the majority of TripAdvisor and Yelp users who provided a review visited those parks as well. Table 2 also reveals a general decline in reviews and visits during the pandemic, which reflects overall reduced travel during that period, as has been found in numerous other studies (Yabe et al., 2020).

Table 2. Chi-square test cross-table for local versus out-of-state reviewers/visitors

Data source	2019		2020		χ^2	p
	Local visitors/ reviewers	Out of state visitors/ reviewers	Local visitors/ reviewers	Out of state visitors/ reviewers		
TripAdvisor	1109 (42.7%)	1486 (57.3%)	462 (53.3%)	405 (46.7%)	29.2	< 0.001
Yelp	288 (55.7%)	232 (44.3%)	253 (66.2%)	131 (33.8%)	10.1	0.001
SafeGraph	3092 (51.1%)	2956 (48.9%)	2578 (59.4%)	1763 (40.6%)	69.6	< 0.001

The distances spanning between a park visitor/reviewer origin state and the Florida centroid were fitted to a power law distribution (Clauset et al., 2009) for the three data

sources and two analysis periods, and visualized as log-log plots (Fig. 5). The charts reflect that the frequency of reviews or visits of Florida state parks decreases with an increase in separation distance between origin state and

Florida. Although a larger percentage of reviewers and visitors came from Florida during the pandemic than before (compare Table 2), there was no consistent

decrease in the alpha value (i.e., the slope) associated with the power law across the three data sources.

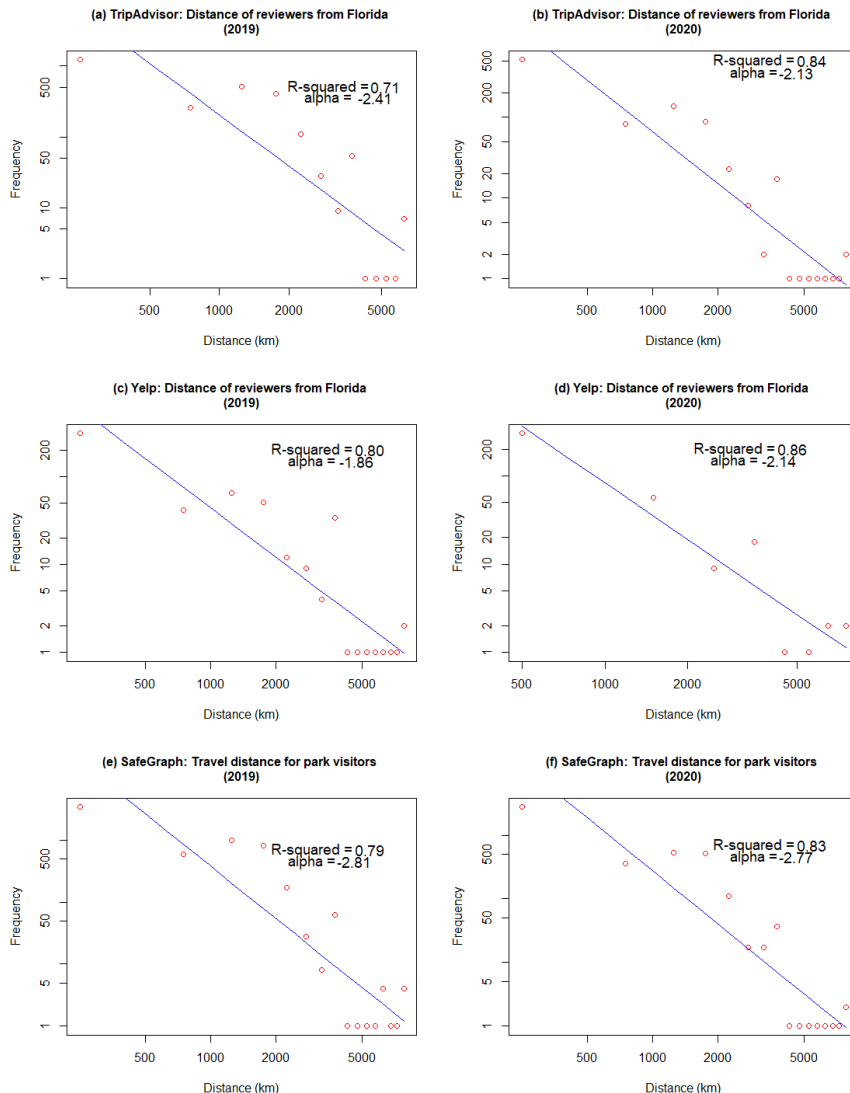


Fig. 5 Power-law approximation of origin – Florida distance distribution for TripAdvisor reviewers (a, b), Yelp reviewers (c, d) and visitors based on SafeGraph visitation patterns (e, f) in 2019 and 2020

4 Conclusions

The results show that there were changes in the distribution of the origins of online reviewers and visitors of Florida state parks from before to during the COVID-19 pandemic.

For TripAdvisor, Yelp and SafeGraph, Florida had the highest proportion of reviewer/visitor/ origins both in 2019 and 2020, which implies that the parks were mostly utilized by local residents. This share of local reviewers and visitors increased further in 2020. The difference in

percentage between local and out-of-state park visitors for the three datasets between before and during the pandemic periods was statistically significant, revealing that the COVID-19 pandemic impacted travel behaviour and visitor trip distances to Florida state parks.

The integrity of social media geodata is affected by various factors such as automated content generated through bots which on platforms such as Twitter is estimated to be at least 9% of its accounts (Varol et al., 2017). Data vandalism, and user privacy restrictions may also lead to incorrect origins (home) locations provided in a user profile. The reviewer origins collected from TripAdvisor and Yelp user profiles were not validated.

Therefore, validation methods, such as analyzing the number of likes a reviewer received or checking whether a user's reviews are primarily posted for POIs near the user's origin location, will be applied as part of future work.

References

- Arkema, K. K., Fisher, D. M., Wyatt, K., Wood, S. A., and Payne, H. J.: Advancing Sustainable Development and Protected Area Management with Social Media-Based Tourism Data, 13, 2427, <https://doi.org/10.3390/su13052427>, 2021.
- Berryman, F.: Announcing SafeGraph Global Brands: <https://www.safegraph.com/blog/announcing-safegraph-global-brands>, last access: 19 April 2022.
- Cirer-Costa, J. C.: Turbulence in Mediterranean tourism, *Tourism Management Perspectives*, 22, 27–33, <https://doi.org/10.1016/j.tmp.2017.01.004>, 2017.
- Clauset, A., Shalizi, C. R., and Newman, M. E.: Power-law distributions in empirical data, 51, 661–703, 2009.
- Ebrahim, S. H., Ahmed, Q. A., Gozzer, E., Schlagenhaut, P., and Memish, Z. A.: Covid-19 and community mitigation strategies in a pandemic, *BMJ*, 368, m1066, <https://doi.org/10.1136/bmj.m1066>, 2020.
- Hasnat, M. M. and Hasan, S.: Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data, *Transportation Research Part C: Emerging Technologies*, 96, 38–54, <https://doi.org/10.1016/j.trc.2018.09.006>, 2018.
- Hecht, B., Hong, L., Suh, B., and Chi, E. H.: Tweets from Justin Bieber's heart: the dynamics of the location field in user profiles, in: *Proceedings of the SIGCHI conference on human factors in computing systems*, 237–246, 2011.
- Hecht, B. J. and Gergle, D.: On the “localness” of user-generated content, in: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, New York, NY, USA, 229–232, <https://doi.org/10.1145/1718918.1718962>, 2010.
- Juhász, L. and Hochmair, H.: Studying Spatial and Temporal Visitation Patterns of Points of Interest Using SafeGraph Data in Florida, *GI_Forum*, 1, 119–136, https://doi.org/10.1553/giscience2020_01_s119, 2020.
- Ma, S. and Kirilenko, A.: How Reliable Is Social Media Data? Validation of TripAdvisor Tourism Visitations Using Independent Data Sources, in: *Information and Communication Technologies in Tourism 2021*, Cham, 286–293, https://doi.org/10.1007/978-3-030-65785-7_26, 2021.
- Rossetti, M., Stella, F., and Zanker, M.: Analyzing user reviews in tourism with topic models, *Inf Technol Tourism*, 16, 5–21, <https://doi.org/10.1007/s40558-015-0035-y>, 2016.
- SafeGraph.: Privacy Policy: <https://www.safegraph.com/privacy-policy>, last access: 21 April 2022.
- Statista.: Yelp: cumulative number of reviews 2020: <https://www.statista.com/statistics/278032/cumulative-number-of-reviews-submitted-to-yelp/>, last access: 15 April 2022.
- Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., and Toivonen, T.: Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas, 7, 1–11, 2017.
- TripAdvisor.: Travelers Push Tripadvisor Past 1 Billion Reviews & Opinions! <https://tripadvisor.mediaroom.com/2022-02-01-Travelers-Push-Tripadvisor-Past-1-Billion-Reviews-Opinions>, last access: 15 April 2022.
- Varol, O., Ferrara, E., Davis, C., Menczer, F., and Flammini, A.: Online human-bot interactions: Detection, estimation, and characterization, in: *Proceedings of the international AAAI conference on web and social media*, 2017.
- Volenec, Z. M., Abraham, J. O., Becker, A. D., and Dobson, A. P.: Public parks and the pandemic: How park usage has been affected by COVID-19 policies, *PLOS ONE*, 16, e0251799, <https://doi.org/10.1371/journal.pone.0251799>, 2021.
- Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., and Ukkusuri, S. V.: Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic, *Sci Rep*, 10, 18053, <https://doi.org/10.1038/s41598-020-75033-5>, 2020.
- Zheng, X., Han, J., and Sun, A.: A Survey of Location Prediction on Twitter, 30, 1652–1671, <https://doi.org/10.1109/TKDE.2018.2807840>, 2018.