6-17-2019

Analysis of Flickr, Snapchat, and Twitter use for the modeling of visitor activity in Florida State Parks

Hartwig H. Hochmair
University of Florida

Levente Juhasz
Florida International University, ljuhasz@fiu.edu

Analysis of Flickr, Snapchat, and Twitter use for the modeling of visitor activity in Florida State Parks

Hartwig Hochmair
University of Florida
3205 College Ave
Davie, FL, USA
hhhochmair@ufl.edu

Levente Juhász
Florida International University
11200 SW 8th Street
Miami, FL, USA
ljuhasz@fiu.edu

Abstract

Spatio-temporal information attached to social media posts allows analysts to study human activity and travel behavior. This study analyzes contribution patterns to the Flickr, Snapchat, and Twitter platforms in over 100 state parks in Central and Northern Florida. The first part of the study correlates monthly visitor count data with the number of Flickr images, snaps, or tweets, contributed within the park areas. It provides insight into the suitability of these different social media platforms to be used as a proxy for the prediction of visitor numbers in state parks. The second part of the study analyzes the spatial distribution of social media contributions within state parks relative to different types of points of interest that are present in a state park. It examines and compares the location preferences between users from the three different platforms and therefore can draw a picture about the topical focus of each platform.

Keywords: social media, Flickr, Twitter, Snapchat, points of interest.

1 Introduction

Data from social media and photo sharing Websites, including Twitter, Foursquare Swarm, Flickr, and Panoramio, have been widely used for the study of human mobility (Alivand and Hochmair 2013; Hawelka et al. 2014). The spatio-temporal distribution of shared geo-tagged images can help to identify tourist hotspots and to recommend tourist routes (Leung et al. 2016). Several studies correlated the number of shared photos with visitor counts. For example, using visitor count data of 38 National Parks in the western United States between 2007 and 2012 one study found that the number of Flickr photos posted monthly in a park can reliably indicate the number of visitors to a park in a given month (Sessions et al. 2016). Other online resources have so far been less explored for activity analysis. For example, Instagram images were used to identify frequently visited locations and most popular activities in the Pallas-Yllästunturi National Park, Finland (Heikinheimo et al. 2017). Another study compared spatial and temporal contribution patterns to Flickr, Twitter, and Snapchat in Florida, finding that Flickr contributions follows closely daylight hours, whereas Snapchat users are more active during evening or early morning hours and Twitter users post their tweets primarily during typical workday hours (Juhász and Hochmair 2019).

Like all crowd-sourced data, also social media and image sharing platforms exhibit user selection and geographical bias. Therefore, understanding differences in user contribution behavior to different platforms is necessary for the assessment of data validity, accuracy, and representativeness (Li et al. 2013).

This paper compares spatial contribution patterns to Flickr, Snapchat, and Twitter observed in state parks in Central and Northern Florida for varying time periods between July 2017 and October 2018. It analyses the following two aspects:

1. It computes the Pearson correlation between state park visitor numbers and the number of Flickr images, snaps, and geo-tagged tweets posted in these parks.
2. It compares the spatial distribution of posts on Flickr, Snapchat, and Twitter within state parks around different types of points of interest (POI).

The first aspect assesses whether the relative abundance of social media activities in state parks corresponds to observed visitor count patterns across these state parks. It examines therefore whether social media activities can be used as a proxy measure for state park visitor counts. If that is the case social media could be useful for park managers or state governments to estimate visitor counts at other locations, e.g. in wildlife management areas, or to fill in data gaps for certain time periods or parks where visitor count data are missing. The second aspect addresses the question of whether users of different social media platforms exhibit preferences for certain POI types in state parks, e.g. beaches or wedding facilities. Such information could help to customize promotions and advertising efforts of state parks in campaigns that are tailored to the preferences of the users base of the social media platform in question.

2 Study setup

2.1 Study area and observation dates

The study area comprises 142 state parks in Central and North Florida (Figure 1). The shapefiles of the park boundaries and the POIs inside the parks were downloaded from the Open Data portal of the Florida Department of Environmental Protection (DEP). The number of state parks included in the different analyses varied by analysis type.

Table 1: Observation dates of social media contributions and state park visitor counts for correlation analysis

Data source	Social media observation dates	# analyzed state parks	# Photos/posts in analyzed state parks	State park monthly visitor counts
Flickr	07/01/17-06/30/18	114	819	Jul '17-Jun '18
Snapchat	09/10/18-10/11/18	121	340	Sep-Oct '17
Twitter	09/01/18-10/31/18	121	610	Sep-Oct '17

Figure 1: State parks in Central and North Florida.

The first part of the study correlates the number of social media contributions (photos, tweets) with monthly state park visitor counts. Since Flickr and Twitter data come with user identifiers, Flickr and Twitter user counts in parks were also correlated with monthly state park visitor numbers. Visitor counts were provided by the Florida DEP for the fiscal year 2017/2018, which covered July 2017 through June 2018. Since some monthly Florida DEP visitor count data were missing not all state parks could be used for each correlation analysis. Table 1 lists for each social media platform observation date, number of state parks included in the correlation analysis, number of photos or posts located in the analyzed state parks within the analysis time frame, and the months used from the Florida DEP reference data source. An exact temporal match between observation data and reference visitor count data was only possible for Flickr. As opposed to this, availability of Twitter and Snapchat data was limited to several weeks or months at the end of 2018. Therefore, September/October 2018 Twitter/Snapchat count data had to be compared to 2017 state park visitor counts.

Figure 1 highlights one state park (Marjorie Harris Carr Cross Florida Greenway) which was found to be an outlier in the correlation analysis, with many more visitor counts observed than expected relative to social media posts. This is not a typical state park since it is elongated and stretches across two thirds of Central Florida. It was therefore excluded from statistical analysis.

For the second part of the study which analyzes the proximity of social media posts to POIs of different categories all 141 state parks (besides Marjorie Harris Carr Cross Florida Greenway) in Central and North Florida were considered. The observation time frame for Flickr and Snapchat data was the same as shown in Table 1, whereas for Twitter it was extended to 08/15/18-11/20/18.

2.2 Social media data collection

Twitter and Flickr provide data access through standard Application Programming Interfaces (APIs) (Juhász et al. 2016). For this study, the Twitter streaming API was used to continuously collect geotagged tweets with exact coordinates over a longer period. Tweets from users likely to be automated profiles were removed (Yang et al. 2019). Flickr photo locations were harvested on December 5, 2018 through the Flickr API. Since Snapchat does not provide an open API, a self-developed tool was used to continuously collect locations and approximate submission times of public posts (snaps) submitted to the "Our Story" feature of Snapchat (Juhász and Hochmair 2019). Since the available Snapchat data contains only the location of the snap and the timestamp of submission but no other metadata, this study focuses solely on the spatial (and to a limited extent also on the temporal) activity patterns of these three data sources.

3 Analysis results

3.1 Study area and observation dates

Figure 2 plots for 114 state parks and a one year time-period the correlation between Flickr photo counts and park visits (a) and between Flickr user numbers and park visits (b). The latter leads to a higher Pearson's r of 0.55 compared to the prior ($r = 0.47$). A possible explanation is that the latter method mitigates biases by individual park visitors who post a disproportionally large number of photos compared to the average Flickr user, e.g. by taking pictures of a plant collection. These correlation values are lower than those found between Flickr photo numbers and bed night numbers in European cities (Kádár 2014). A possible explanation is the small sample size of Flickr images in state parks compared to urban environments, leading to higher uncertainties in the correlations. A second explanation is that the actual composition of park visitors (those visitors who use social media and those that do not) varies by park due to other covariates (e.g. distance from city or the park size). Such a potential relationship needs to be explored in future work. Figure 2c and d plot correlations between monthly Flickr photo and user numbers and Park visitor numbers, which are low ($r < 0.6$). Also here, the small sample size of monthly posted images might be a possible explanation.

Figure 3 plots for 121 state parks and a two-month period the correlation between tweet count and park visits (a) and between Twitter user numbers and park visits (b). As with Flickr, the correlation with user numbers is higher. This indicates that mitigating the bias caused by exceptionally active social media users is important for obtaining a more accurate estimate of visitor numbers. The higher correlation of Twitter users with state park visitors ($r = 0.67$) than that for

Flickr users (r = 0.55) suggests that Twitter is a somewhat more useful resource to predict visitor numbers at given locations even outside urban environments, such as in state parks. Figure 3c plots snap counts against state park visitor numbers for about a 1-month period, which results in a lower Pearson's r value of 0.39. The shorter observation period compared to both Flickr and Twitter could play a role in this.

3.2 Spatial association of social media activities with POI types

The park maps and shapefiles provided by the Florida DEP distinguish between over 60 types of POIs. Some of the POIs, although mapped as points, extend along linear features (e.g. biking trail) or across larger areas (e.g. birding). These types of spatially expansive POIs were removed before further analysis. Furthermore, we grouped similar POI categories (e.g. different types of camping facilities) into one type to simplify the analysis. After this process a total of 18 POI types remained. Figure 4 shows the location of POIs in two adjacent state parks, in which 11 of the 18 POI types are present. The map also shows the location of contributions from Snapchat, Flickr, and Twitter.

Since the targeted user base for Flickr, Twitter and Snapchat apps is different we hypothesize that this difference can be observed by different types of POIs around which

Figure 2: Correlation between annual Flickr photo counts and park visits (a), between annual Flickr photo user counts and park visits (b), between monthly Flickr photo counts and park visits (c), and between monthly Flickr photo user counts and park visits (d).

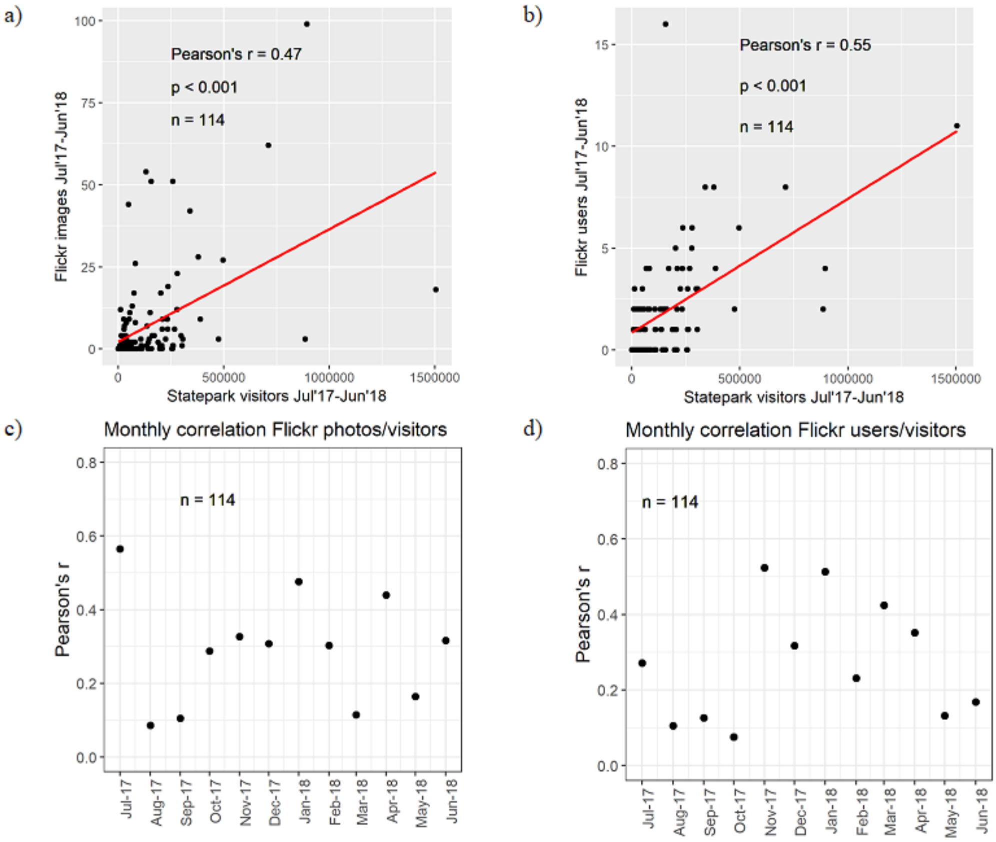
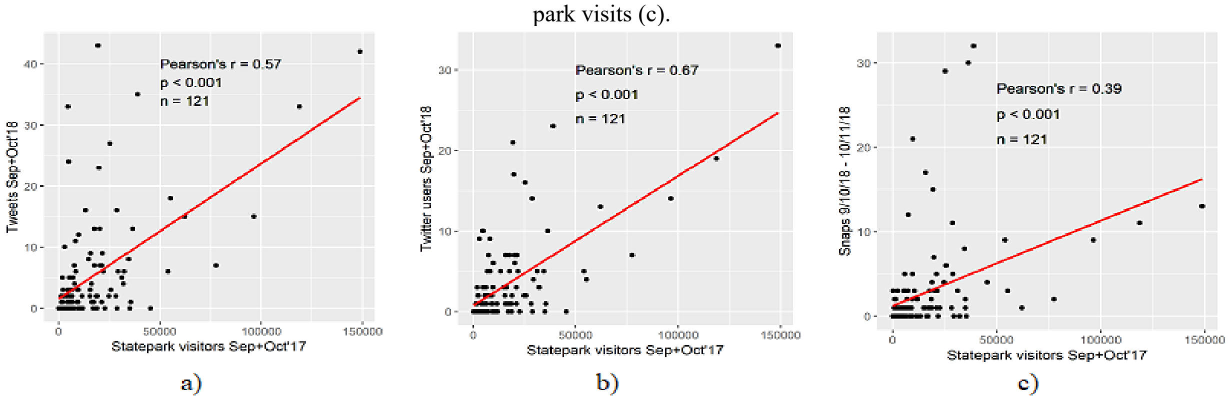
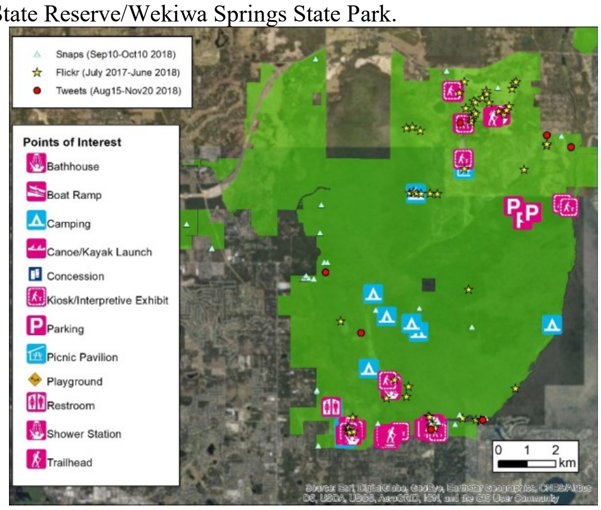


Figure 3: Correlation between tweet counts and park visits (a), Twitter user counts and park visits (b), and snap counts and park visits (c).



contributions are posted in the different platforms. To explore this further, we computed for each platform across all state parks the percentage of posts that had a specific POI type within a 200 m radius. For this computation only those parks were considered which actually contained the POI of the analyzed type. POI types with a higher percentage value for a specific platform suggest that users of that platform found interest or use in posting in the vicinity of this POI type.

Figure 4: Social media contributions to Rock Springs Run State Reserve/Wekiwa Springs State Park.



This count process can be formalized as follows. If $\boldsymbol{s}_i$ is the set of social media posts from source s posted within state park i among all n analyzed state parks, and $\boldsymbol{L}_p$ is the set of POIs of type p within state park i, then the percentage of social media posts near a POI of type p across all analyzed parks can be computed as

$$Pct_{s,p} = \frac{\sum_i^n \#\left(d_{min}(\boldsymbol{s}_i, \boldsymbol{L}_p) < 200m\right)}{\sum_i^n \#\left(\boldsymbol{s}_i \mid \boldsymbol{L}_p \in Park_i\right)} \quad (1)$$

where $d_{min}(\boldsymbol{s}_i, \boldsymbol{L}_p)$ expresses the list of shortest distances from each post within set $\boldsymbol{s}_i$ in a park i to its nearest POI of type p. The # operator counts items satisfying a given condition.

Figure 5 shows the result of this process for the three data sources with POI types being sorted alphabetically. Using an unweighted mean across all POI types tweets have the highest percentage of posts located within 200 m of any POI (28.6%), followed by Flickr images (23.8%), and snaps (9.4%). This shows that tweets are frequently taken near dedicated POIs and less frequently taken off of marked or designated areas. A possible explanation is that Twitter is typically not used to post images, and hence tweets are not necessarily sent from nature spots away from marked POI areas. As opposed to this, Flickr and Snapchat are photo or video based. These are often taken at scenic sites further away from designated POI areas.

All platforms share some mundane POIs from whose vicinity is often posted from. One of these are concession buildings which are frequently visited to obtain tickets and permits. Similarly, parking lots and restroom areas experience above average posting rates, which is probably not due to the scenery of these POIs, but because visitors are gathering around these POIs upon their arrival or departure to take (group) pictures or send messages. Similarly, picnic and pavilion areas offer opportunities to share tweets and group images. Among the theme related POIs historic sites receive high post rates among Flickr and Twitter users, possibly due to their scenery and interesting history, which is potentially less relevant for the Snapchat community. Figure 6 shows an example of a Flickr image taken near a historic POI. Docks and piers as well as beaches with their interesting motifs and scenery are prominent spots for Flickr but less attractive for Twitter and Snapchat users. Weddings receive relatively high contribution rates on Twitter and Snapchat (but not so much on Flickr), reflecting the social event type of happenings at these locations.

Figure 5: Percentage of social media points within 200 m from selected POI types.

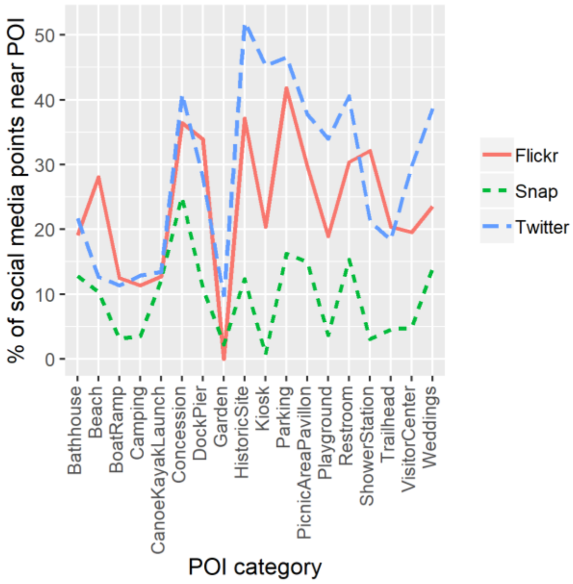


Figure 6: Flickr image of Fort Clinch State Park in Fernandina Beach, Florida.

Source: https://www.flickr.com/photos/67355751@N04/28199105129

In summary, the comparison of frequencies of social media posts around POIs between platforms gives some insight into commonalities of and differences between the different user communities of online platforms in terms of location preferences and topical interest.

4 Conclusions

The first part of the study revealed that the correlations between social media platform activities in Florida state parks and visitor counts are moderate. Therefore the number of social media posts contributed in a state park cannot be used as an accurate proxy for visitor counts. For Flickr and Twitter the correlations were higher for user numbers than for photo and tweet counts, indicating that participation inequality within social media platforms can lead to distortions in estimated visitor counts. Although Snapchat showed lower correlation numbers than both other sources, these presented results are based on a short-term sample only and hence need to be interpreted with caution. Proximity analysis of posts around POIs in state parks revealed certain differences in location preferences between users of these three platforms. This information could be used in customized promotion and advertising campaigns in the different platforms to attract a certain user base for park visits. For future work we plan to extend this analysis to longer observation windows and to integrate covariates into regression type analyses for a refined prediction of state park visitor numbers from social media contributions. We assume that a longer-term social media data set, a combination of these social media sources into one observable, and the consideration of various socio-economic and environmental predictor variables will help to increase the correlations between social media activity counts and visitor count numbers in state parks and hence improve the usability of such online data as a proxy for state park visitor counts.

References

Alivand, M. and Hochmair, H. H. (2013). Extracting Scenic Routes from VGI Data Sources. In D. Pfoser & A. Voisard (Eds.), *Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information (GEOCROWD)* (pp. 23-30). Orlando, FL: ACM.

Hawelka, B., Sitkoa, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., and Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41* (3), 260-271.

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., and Toivonen, T. (2017). User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. *ISPRS International Journal of Geo-Information*, *6* (3), 85.

Juhász, L. and Hochmair, H. H. (2019). Comparing the Spatial and Temporal Activity Patterns between Snapchat, Twitter, and Flickr in Florida. *GI_Forum*, *2019* (1).

Juhász, L., Rousell, A., and Arsanjani, J. J. (2016). Technical Guidelines to Extract and Analyze VGI from Different Platforms. *Data*, *1* (3), 15.

Kádár, B. (2014). Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*, *16* (1), 88-104.

Leung, R., Vu, H. Q., Rong, J., and Miao, Y. (2016). Tourists Visit and Photo Sharing Behavior Analysis: A Case Study of Hong Kong Temples. In A. Inversini & R. Schegg (Eds.), *Information and Communication Technologies in Tourism 2016* (pp. 197-209). Cham, Switzerland: Springer.

Li, L., Goodchild, M. F., and Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, *40* (2), 61-77.

Sessions, C., Wood, S. A., Rabotyagov, S., and Fisher, D. M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*, *183* (December), 703-711.

Yang, K.-C., Varol, O., Davis, C. A., Ferrara, E., Flammini, A., and Menczer, F. (2019). Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, *1* (1), 48-61.