Is a Photo Worth 1,000 Likes?
The Influence of Instagram at National Parks

Ashley C. Lowe Mackenzie
University of Hawaii at Manoa
alowemac@hawaii.edu

Steven J. Dundas
Oregon State University
steven.dundas@oregonstate.edu

Working Paper

(Preliminary Draft: Please Do Not Cite Without Permission)

September 15th, 2023

Abstract:

The rise of social media platforms has led to many facets of society that may now be impacted, directly or indirectly, by online behavior and content. One hypothesis posited in the news media is that social media, Instagram in particular, is driving increases in visitation to public lands. Using millions of georeferenced posts to Instagram from 2012 to 2019, we test this hypothesis at National Parks in the United States. We find national parks experienced substantial growth in visitation after Instagram had gained influence as a social media app but the reasons for the increase vary across parks. A small number of National parks with significant viral content with high user engagement on Instagram saw increases in visitation attributable to the cumulative impact over time from this influential content whereas most parks with less viral content saw visitation increases associated with the volume of upload activity. Posts from avid users of Instagram have the potential to impact a broader set of parks through content with high engagement. We also use textual and sentiment analysis to show that engagement online is driven by expressing both positive and negative sentiment about the content or listing points of interest and by avid users of the app.

Introduction

Management of the nearly 640 million acres of public land in the United States (U.S.) has multiple objectives, including balancing the conservation of species and ecosystems with recreational use. Trends in recreational visits to U.S. National Parks (USNPs) have experienced various periods of growth over the last century (Fig. 1). There was a multi-decade steady increase in visitation following the end of World War II in the mid-1940s. This prolonged trend is often attributed to the growth in the U.S. population and rising levels of income, along with improvements in travel and accessibility.[1] By the early 2000s, visitation was stagnant, leading some researchers to investigate reasons for the lack of growth[2,3]. This period of stagnation was short-lived, and visitation rose sharply in the early years of the 2010s. From 2015 to 2019, the NPS reported record-breaking visitation every year with the initial growth beginning around 2013. This recent visitor overcrowding adds stress to staff, facilities, ecosystems and aging infrastructure [4,5]. This latter point is significant as the U.S. National Park Service (NPS) currently faces a $22.3 billion dollar backlog of deferred maintenance.[6]

A common culprit portrayed in the news media and land management circles for the recent rise in visitors (and the resulting resource degradation) to USNPs is the social media platform Instagram[7–10] Instagram began as a photo-sharing app and has shouldered blame for the growing crowds given the timing of the release app, the nature of the content (e.g., georeferenced images of iconic landscapes), the way content is delivered to its 1 billion active monthly users and reports by park managers reporting swarms of visitors to particular locations observed taking self-portraits, or selfies.[9–11] Data from Instagram and other social media platforms have also been demonstrated to provide practical benefits to land managers. Research has found social media data is a reliable proxy to understand the number of visitors to public lands[12–16], suggesting avenues to reduce costs of data gathering across managing agencies and extend visitor use estimates to wilderness areas where prior data collection was limited[12,14,17–20]. Linking within-app behavior from a social media has been less studied and is an important aspect in understanding social medias relationship with visitor behaviors.[21] Regional evidence of Instagram's influence suggest certain locations are more susceptible to social media influence on visitation is attributed through the mechanisms of online engagement (e.g. likes and comments) and not the general act of sharing of photos of the location online.[11] The locations within this analysis are the "crown jewels" of the nation's public lands requiring congressional approval that the areas hold national significances of natural, cultural or recreational resources. The influence of Instagram across national recognized locations can provide spatial relevant information. Furthermore, the mechanism driving engagement within-app is far less known[11]. The framework for the influence of Instagram is driven by the user base within the app (Fig 2) and could have various mechanisms for driving user offline decisions. It could be driven by a reduction in search and information costs[22] bringing an awareness to potential recreators of previously unknown beautiful locations or induce a previous recreators to make a visit or the mechanism could be more behavioral such as the bandwagon effect (or herd behavior), whereas an individual's demand for a commodity is increased due to the fact that others are also consuming the same good.[23]

In this article, we explore social media's influence on National Parks and address overarching question about how online engagement influenced visitation across national iconic locations. To do so, we examine the following questions. First, does Instagram matter? We examine the timing of the impact and whether contemporaneous upload activity or engagement within the park matter in understand the parks visitation trends. Our findings suggest National Parks experienced an increase in visitation after Instagram gained influence, but only certain National Park can attribute the visitation growth experienced to engagement online. Parks which had a relatively viral moments experienced an increase in attributable to the cumulative impact from all current and past influential content whereas parks which did not experience a viral moments only say a contemporaneous effect of activity online. Second, what increases engagement within the app? Here we examine the type of information which increases engagement. We find specific points of interest within each park, expressing an opinion about the experience and the type of Instagram user increases engagement within the app. Furthermore, we find avid user who are highly active on the app. Lastly, we ask if the type of engagement impacts parks differently? We find parks which had the highest viral moment were impacted by the pure high engagement online but that avid users who posted frequently under geotags increased visitation to parks who had a slightly less viral moment. The work provides evidence the mechanism driving the increase in visitation is a behavioral effect such as the bandwagon effect. This work adds on to the ongoing conversations of how online activity is impacting offline activity.

Understanding driving factors of visitation across parks can help park managers develop effective management plans and strategies. Previous research important indicators in visitation trends include recessional indicators and other cost to visit finds (fee, gasoline),[3,24–26] changes in demographic[25,27] designations of a National Park [24,28,29] and residual impacts on travel from national terrorist attacks.[2,25]

Does Instagram Matter?

To determine whether Instagram impacted visitation to National Parks, we first examine the timing of the increase based off of Instagram's history.[11] We examined the systematically impact on visitation patterns at the time of the launch of the app (October 2010) and find that the launch of Instagram does not systematically change visitation across parks (Figure 3.A). Other important factors included the growing percent of the aging population and extreme temperatures (Figure 3.B**)**. The result from the launch is unsurprising given a launch of a social media app would likely not have a systematic impact on visitation. Instead, the offline impact from Instagram may correspond to when the app had gained influence within the social media landscape. A regional analysis of Instagram found April 2012 as a monumental month in Instagram's history.[11] In April of 2012 Instagram was bought by Facebook, currently known as Meta, the operating system was expanded from iPhone to Android operating system and the app reported reaching 50 million active users per day. We find evidence across all National Parks within our study experienced a significant increase in visitation at this period (Fig. 3.A).

The timing based of off Instagram's influence suggests an a significant change in visitation trends, however it could also capture other influences of technological advances during this same time period To isolate Instagram from other time periods, we evaluate the unique uploads from individual user per day of geotag photos to each park.[12] Using the monthly count photo-user per day (PUD), we find the number of PUD accounts for increases in overall visitation at only the 10% significant level (Fig. 3.C). The treatment only controls for the contemporaneous act of uploading content to a park and considers all parks are experiencing similar activity levels from visitors within the app. However, content engagement and its feedback to users (Fig. 2) are important aspects of the functionality of the app.

The emergence of social media has given online users the ability to go viral. Going viral occurs when an image spreads quickly and widely online. The process of going viral increases visibility of the individual and the content uploaded. Generally, on Instagram viral content is quantified by the number of likes and comments an image posted received. To evaluate a park's viral quality, we examine the most viral moment. We isolate each park most viral content based on the level of engagement, e.g., the number of likes and comments it received (Fig 4). The parks vary greatly based on their most viral moment and does not reflect average visitation at the park. Though most parks had a viral moment, two parks are clear comparative outliers on Instagram, Zion and Yosemite. Another sixteen parks appeared to have had moderate viral moments whereas the remaining parks had a visible engagement but comparatively did not have as viral of a moment on Instagram as other National Parks. The average visitation growth of each engagement grouping differs between these groups (Fig. 5). The high and medium viral parks saw an increase in average visitation with a large increase associated with high engagement parks. The low engagement parks appear to have a stagnated average visitation trend. Based on engagement grouping, we find the PUD activity has a nonlinearity quality associated with the level of engagement (Fig. 3.C) such that the increase in monthly visitation increases as a parks viral quality decreases. PUD at parks which went viral do not capture contemporaneous visitation as well as parks experienced a lower viral moment.

Given the varying viral moments across parks, engagement may systematically impact locations differently. We examine whether high levels of engagement impacted the parks visitation. We consider content to have a high level of engagement by the number of likes each post received by subsetting content which reached the 90th and 95th percentiles based on likes from all photos uploaded using a National Park geotagged **(see Supplementary Material for full table)**.[11]The yearly threshold controls for the growth of Instagram use overtime (Methods). The highly engaging content is treated as a cumulative effect to account for the feedback loop Instagram may have on users (Fig. 2). Content which receives high levels of engagement is likely to be sticky and remain visible for longer periods of time online. This is especially the case for Instagram which overtime increasingly relied on algorithmic based content versus chronological ordering content. Algorithmic content is generated based on user behavior and engagement from others. The cumulative effect accounts for the past and present impact of the lingering impact from engaging content online. We found evidence that the cumulative effect from engaging content does impact parks but only at those parks which had a highly viral moment (Fig 3.D). Thus, when we consider by park group impact from highly engagement content we find the inverted effect of the contemporaneous result, i.e. viral parks are experience growth associated with highly engaging content. The parks which had the highest viral moment compared to others saw a 7.7% increase in visitation per month associated with the cumulative impact from all past and current highly engaging content uploaded to the park.

Spatial Impacts from Engagement and Avid Users

We find evidence engagement on Instagram has had offline impacts to visitation at certain parks. To better understand what increases engagement, we analysis the caption of all geotag photo in our analysis. For each park, we build park specific dictionaries to capture information presented in the caption. This includes whether the content included general location information (the state the park is located or the park name), points-of-interest listed by park managers as important locations within the park and the activities listed as advertised Things To Do on each individual parks website.[30] The sentiment of the text is analyzed to provide context on whether the user caption was positive, neutral or negative (Methods). Furthermore, we consider user behavior by accounting for the number of times the individual uploads content to any National Park geotags over the period. We isolate avid users of the app following the previous method to determine engagement (Methods). The avid user is of interest given our interest in engagement. The ability to go viral has led to a social media profession known as influencers. Influencers is a form of social media marketing which have the power to affect the buying habits of others through the content they upload. Though it is beyond the ability of the dataset to determine if the avid users are Influencers, avid users likely contain Influencers who focus on the great outdoors and whom has an esthetic involving the National Park.

Among all content geotagged at a National Parks, we find park specific points-of-interest, non-neutral sentiment and avid user increases engagement from online users (Fig 6.A). Park specific points-of interest increases the engagement rate by 1.2 times greater than content from users not discussing specific locational information. The engagement rate also increases if the caption expresses a positive or negative sentiment. The results from negative sentiment may seem less intuitive. Negative sentiment increases engagement more than positive sentiment. However, a random selection of negative comments finds the type of negativity is often used in either a descriptive, facetious, sarcastic or a pun of park names which have a negative sentiment such as Badlands and Death Valley (**see Supplementary Material tables**). Lastly, the highest determinate of engagement rate is from avid users of Instagram. Avid users are expected to have an engagement rate 5 times greater than all other users. This result finds supportive evidence the mechanism driving engagement is less an informational shock and more of a behavioral effect such as the bandwagon effect.

Lastly, we isolate the spatial impacts from avid users engaging content. Avid users posting highly engaging content in the 90th percentile make up 27% of all highly engaging content. We examine the cumulative impacts from highly engaging content from avid users across park groups. This analysis provides complimentary findings from our initial model (Figure 3.D) with one main difference. Here our findings suggest content from avid users which received high engagement had a cumulative impact on both the high and medium engagement parks (Fig 6.B). However, when we examine the remaining 73% of the engaging content from non-avid users, the result only holds for the parks experiencing highly viral moments. We find experiencing an extremely viral moment on the app may systematically influence visitation trends from the sheer engagement online whereas parks which had a milder viral moment can attribute increases to the type of user posting engaging content. The cumulative effect per month of all past and current engaging content from avid users increases visitation by 4.4 to 7.2% with the highest impacts observed at the parks which experiencing the most viral content online.

Discussion

Online activity has offline impacts[31] and National Parks visitation is no exception. Instagram has often been blamed for the recent surge of visitation and overcrowding issues and our findings support aspects of this claim. Once Instagram had gained influence, National Parks were experiencing significant growth in visitation and some of this growth can be attributed to Instagram but not all. Parks which did have a highly viral moment in the context of geotagged National Parks experienced a cumulative impact from all past and present highly engaging content. Furthermore, when we isolate avid user within the engaging content, the cumulative impact spilled over to parks experiencing moderate viral moments. The result suggests a positive nonlinear visitation relationship associated with parks generating engagement whereas the PUD account for a negative nonlinear relationship with visitation at lower engaging parks.

We also find changing national demographics played a large role in the recent increases in visitation. The percent of the national population which is 65 and older increased rapid over this time and as a covariate accounted for a large portion of the overall increase visitation. This demographic has increased levels of leisure time through higher rates of retirement as well as a lower cost associated with the visits to a national park. During this period, the NPS provided a lifetime pass to seniors $10 starting in 1994 and only increased it to $80 in 2017. The timing of Instagram's influence showed a across park increases in visitation yet our results including Instagram content does not fully account for the total visitation associated by the timing (Fig 3.A). We suggest other factors such as technology advances, additional social media sites and changing popularity of outdoor recreational lifestyles may have also influenced visitation. The general rise of smartphones and their sophisticated handheld GPS systems have increased security and access to outdoor area allowing recreators the ability and confidence to go further in the backcountry and find new locations to visits. Given our results on engagement, other social media sites with the ability to go viral such as Youtube and Tiktok also likely influenced viewer's trip decisions at various times. Furthermore, the suggestive evidence of a behavioral mechanism such as the bandwagon effect likely has influenced the growing popularity of alternative lifestyles. The rise of vanlife grew in popularity over this period and the growing movement of individuals living in converted vehicles part or full time often occurring on public lands. Outdoor recreational sports like climbing and thru hiking of long-distance trails grew in popularity after national releases of books such as Wild and Free Solo. Each of these types of lifestyles have their own online culture with some captured within the Instagram effect.

The policy implications of these findings suggest supportive evidence of actions taken by public land managers to address rising impacts from social media. Social media can make the outdoors less exclusive and increases diversity.[33] Our findings suggest accounts producing engaging content can have an impact on a public good. Currently, the NPS requires monetized social media accounts, no matter how large or small, to secure a commercial photography permit if the user post content is filmed on over 400 NPS locations across the nation. However, the permits require honest self-identification from visitors who are posting content under monetized accounts or enforcement by identification of individuals who are producing content online with monetized accounts. The act of enforcement is likely a cumbersome task for public mangers who are currently dealing with a backlog of deferred maintenance and repairs of $22.3 billion.[6] Other potential viable low-cost routes could include maintaining an active presence on social media sites informing users on those apps of these policies.

Another policy implication suggests a regulating route for social media sites. Social media is a relatively unregulated sector yet has powerful impacts on real world[31,32]. For many platforms, determining whether an account is monetized is private information (Tiktok, Instagram). Providing a route for public land managers to request which monetized accounts are posting content on public lands requiring permits could increase the ability to enforce compliance. Furthermore, providing a pathway for public land mangers understand engagement changes to the areas they manage online would increase preparedness for visitation changes or identifying and locating hotspots of activity. Policy focused on providing a pathway for access to in-app indicators for locations associated with public lands could increase the understanding of the new outdoor recreation paradigm.

---------Above 3,500 Words-----

Methods

Visitation Data

National Park visitation is obtained using publicly available *Visitor Use Statistics* from NPS Stats.[34] The database contains information on visitor trips to the over 400 park locations managed by the NPS. The type of a visitor trip can include recreational or non-recreational visits, the visit hours and whether the trip included overnight stays. My empirical setting focuses on the impacts to visitation to national parks (NPs), which includes sixty-three (63) locations under management by the NPS (Extended Fig 1). Out of the 63 NPs, our analysis is on parks within the lower forty-eight states that are easily assessable by car (excluding island and water-based NP). This reduces the set to forty-five (45) NPs. Five of the parks were designated as a NP after 2010 (post-Instagram), further reducing the set to forty (40) parks. There were a few parks designated during the sample period prior to 2010, which included Great Sands Dunes, Cuyahoga Valley, Black Canyon of the Gunnison, Death Valley, Joshua Tree, and Saguaro NP. For these parks, a control for designation is included for the period once it was designated. Data on national park entrance fees were obtained through email communication with NPS. The available data on entrance fee ranged from 1993 to 2019.[2,25] This reduces the empirical setting to a 27-year period of monthly observations for 40 national parks (see Supplementary Material for statistical summary).

Social Media Data

Instagram data was collected in late 2019. The National Park setting, over a similar timeframe, has over 8.3 million geotagged posts. The collection of all geotags associated under each National Park in our sample.[11] The collect all geotagged posts metadata containing each National Park's name and location. Individual posts outside the latitude and longitude coordinates of the park were identified and removed. The collection of metadata from public Instagram pages, producing information on each post indexed by identification number, date the photo was uploaded, the number of *likes* and comments, and the caption used. It did not collect demographic information on the user, or the image taken.

Photo user per day (PUD) is a unique upload from an account observed at a park on a given day. The total PUD per month accounts for the contemporaneous effect from user who are uploading content to the site on the given month the user decided to make a trip and share their content with others on the app. Viral parks were based off the most commented and most liked photo on single graphical plane (Fig. 4). The classification did not require content to be same photo. This figure provides potential grouping strategies by visualizing viral content of each National Park based on user activity within the app. The y-axis marks the highest liked content for each park and the x-axis marks the highest commented content to each park. The first grouping includes two parks Zion and Yosemite NPs. The next group contains sixteen more parks: Kings Canyon, Sequoia, Grand Canyon, Joshua Tree, Great Smoky Mountains, Everglades, Death Valley, Bryce Canyon, Carlsbad, Guadalupe Mountains, Glacier, Grand Teton, Mount Rainier, Olympic, Arches and Rocky Mountain. The remaining 22 parks are considered parks received minimal viral content.

Engagement was quantified by the number of *likes* and *comments* a photo receives. We measure cumulative influential Instagram posts with the inclusion of one more threshold to identify parks based on higher engagement:

$$InfPost_{pT} = \sum_{t=1}^{T} Content_{pt}^{l} \quad (1)$$

where $InfPost_{pT}$ is the cumulative (T) influential posts for park p, and $Content_{pt}^{l}$ is the content for park p in month t that meets threshold l. An influential post is quantified by surpassing a threshold, l, of the number of likes a post receives. The threshold is determined if the post is in the top 90th and 95th based on the total number of likes received on the entire set of photos geotagged per year (see Supplementary Material).

Covariates

Variables provided at monthly intervals is important to reduce the reliance of interpolating from yearly estimates. Macroeconomic variables were retrieved from the Federal Reserve Bank of St. Louis[35]. A seasonally adjusted national monthly income level measuring real disposable personal income (RDPI) was collected. This measure captures changes to after-tax income and has been previously used in NPS visitation research.[24] Unemployment is a reliable monthly estimate provided both at the national and state level, capturing consumer confidence overtime.[3,35] Population data is an annual estimate provided at a state level.[35,36] A linear interpolation was used to match the population to a monthly scale. For each variable provided at a state level, the park home state was matched to control for regional changes. To control for the changing demographics of the general population, the percent of the population 65 and above was calculated from the census bureau.

Gasoline prices can impact individual travel costs and have been found to be significant in aggregate visitation models.[3,25] National Parks are often accessed by car and gasoline prices may reflect effects of the recreationalist budget by capturing increases to transportation cost to these remote locations. Gasoline prices from the U.S. Energy Information Administration (EIA).[37] The EIA provides historical monthly average real gasoline price adjusted for inflation for the nation.

Daily observations of temperature (maximum, minimum, average) and precipitation for all parks were collected from PRISM Climate Group at Oregon State University.[38] Extreme heat and cold are likely to impact real-time decision of individuals and thus daily weather at each location. Exploiting this variation across space and time, we estimate a non-linear response function.[39,40] The temperature bins are set at a 10-degree Fahrenheit scale (i.e., 30.0 – 39.99°F, 40.0 – 49.99°F, etc.) This treatment of weather controls for real changes in locational conditions through time. We use maximum temperature instead of the daily mean temperatures for behavioral and spatial reasons. Information broadcasted on weather stations provided to potential recreators is reported as a daily high and low for a region. Therefore, recreators are likely reacting to the temperature maximum rather than the temperature mean. Another spatial concern arises given the national scope of this project. Locational variance in temperature is heterogenous across regions. Deserts areas can have higher highs as well as very cold nights – extremes that could get lost in an average measure. In such a case, a desert and a forested area with milder weather may appear to have a similar mean temperature distribution but a maximum temperature approach could account for and use the higher variance in deserts.

Statistical Analysis

The first modeling specification examines the impact of Instagram's launch on visitation using a binary indicator to denote the pre- and post-Instagram periods, $\boldsymbol{IG_{Launch_t}}$, and compare it to the time where Instagram would be likely to start influencing behavior in April 2012 ($\boldsymbol{IG_{Use_t}}$):

$$Ln(Visit_{pt}) = \beta_0 + \beta_1 \boldsymbol{IG}_{Launch_t} \boldsymbol{or}\ \boldsymbol{IG}_{Use_t} + \beta_2 T_{pt} + \beta_3 Precp_{pt} + \beta_4 X_t$$
$$\beta_5 Ln(Fee)_{pt} + \beta_6 Desig_{pt} + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \qquad (2)$$

Where T is a monthly count of observed temperature bins at each park, $Precp$ is the average percipation at each park per month, X is a vector of covariates at time t, $Ln(Fee)$ is the park specific log of the enterance fee, $Desig$ is a binary indicator for designation if a park within the dataset was designated as a National Park prior to Instagram (New designations were excluded) and $\rho_p, \tau_{y(t)}, \gamma_t$ control for park, year and month fixed effects.

Our model to understand general upload activity on overall visitation controls for unique photo user per day (PUD)[12] to examine the contemporaneous effect of individuals uploading content in a given month.

$$Ln(Visit_{pt}) = \beta_0 + \beta_1 \boldsymbol{IG}_{Launch_t} + \beta_2 \boldsymbol{PUD}_{p_G t} + \beta_3 T_{pt} + \beta_4 Precp_{pt} + \beta_5 X_t$$
$$\beta_6 Ln(Fee)_{pt} + \beta_6 Desig_{pt} + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \qquad (3)$$

where $\boldsymbol{IG_{Launch_t}}$ is a binary indicator to control for the period pre- and post-Instagram and PUD is the monthly uploads of unique users to Instagram per park at time t. An ideal experiment would have some parks were exposed to Instagram and not others. However, our viral group strategy, G, identifies which parks had the most viral moment across all parks in the study. Our grouping strategy will provide insight into the differential effect of the accuracy of monthly activity on Instagram at parks with varying levels of user activity within the Instagram app. We then examine how current and future visitation is impacted by the cumulative impact of highly engaging content.

$$Ln(Visit_{pt}) = \beta_0 + \beta_1 IG_{Launch_t} + \beta_2 PUD_{p_G t} + \beta_3 \boldsymbol{Influential}_{p_G t} +$$
$$\beta_4 T_{pt} + \beta_5 Precp_{pt} + \beta_6 X_t + \beta_7 Ln(Fee)_{pt} + \beta_8 Desig_{pt} + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt} \qquad (4)$$

where $Influential$ is determined by eq. 1 by park and viral group.

Textual Analysis

We found certain parks receiving a high level of engagement experienced cumulative impacts from highly engaging content whereas parks with lower engagement find the upload activity of PUD is contemporaneous. Our interest in parks based on engagement motives our examination of what type of information included in geotagged content is generating user engagement. Our interest is in a count variable of online users within the app liking a photo based off the unknown quality of the photo. Nevertheless, we do have some information on what is being discussed in the caption. Textual analysis provides a route to turn text into data by systematic quantifying textual information and extract meaning from fields of text.[41,42] The Instagram data on 40 national parks contains nearly 8.3 million observations of geotagged content uploaded to Instagram. Each observation contains a caption. Each caption contains a string of characters up to 2,200 characters in length.[1] The text material within these captions contains a range of information about the location, the experience, or what the individual did while visiting. Incorporating textual analysis by examining the information shared in the caption of each geotagged photo to estimate how and if the type of information explains the level of user engagement with the content.

The first step includes cleaning and simplifying the set of textual data. The text is processed by implementing tokenizing, stemming or lemmatizing to convert all letters to lower case and removing all stop words and punctuation from the text set.[41] The next step is to extract meaning from the remaining list of words. We both construct a dictionary for each park as well determine the sentiment of each post using natural language processing. The park-specific dictionaries focus on three information sets: 1) location, 2) specific points of interest and 3) activities associated with each park. The sentiment analysis utilizes a Python-based natural language processor designed specifically for social media microblogging for sites like Twitter and Instagram.[43] The following sections discuss how each dictionary was built and the sentiment analysis.

[1] Instagram allows 2,200 characters which are counted by the number of letter or symbols in a caption. The numbers of words are then bounded by the allotted characters used.

Park-Specific Dictionaries

Geotags included the name of the location under the user's name on each uploaded content under a given tag. Therefore, each post within the dataset already has the name of the park location embedded into the post. The user may still discuss the park within the caption using the name or the state where the park is located. This is evident when performing a word frequency analysis to generate word clouds at each park location. The top word is often the name of the park (Extend Fig 2). For the first park-specific dictionary, each caption is iterated over to determine if the user is discussing the park or the state in which the park is located. This is a simple determination of location information and does not provide insights into more nuanced locations within the park such as a location with a specific name. For example, Angels Landing is an iconic location within Zion National Park so much so that it has a dominant position in Zion's word cloud (**Extended Data Figure 2**fig 2). To determine more nuanced locational information, we utilize the NPS dataset of Points of Interest (POI)**.** The NPS POI dataset provides a list of useful names of location of interest as well as the location type such as a trailhead, viewpoints, historical building, restrooms, or parking lot. POI dictionary is created from the names and type for each park. Another park dictionary is created for the type of activities available at each park. Activities like hiking are commonly used in captions (**Extended Data Figure 2**fig 2). The activity list is generated from each park's "Things-to-Do" on their respective nps.gov websites. The activities listed on the 40 parks included auto touring, backpacking, backcountry camping, biking, birdwatching, boating, campground, caves, canyoneering, commercial tours, picnicking, fishing, hiking, hot springs, kayaking/rafting, historical places, horseback riding, lodge, museum, pet friendly, photography, ranger-led Programs, religious, climbing, stargazing, snowmobiling, sandboarding, swimming, sunset & sunrise, tidepools, visitor center, viewpoints, wildlife viewing, winter activities, wildflower viewing and waterfalls. To make sure activities were captured, words were added to certain activities to capture conversations associated with the activities. For example, parks considered "pet friendly" are specifically referring to allowing more access to dogs on trails and areas within the park, therefore the word dog was added while the word "friendly" was removed.

Sentiment Analysis

The next step is to extract meaning or feelings from the text. Sentiment analysis is a method to quantify the extent of negative, neutral or positive emotions from a text using a dictionary containing a list of positive and negative words. Extended Figure 2 highlights some of the positive sentiment common in two parks such as “beautiful” and “love.” The analysis counts the frequency occurrence of terms to construct a sentiment score (40). We utilize the open-source Python library package Natural Language Toolkit (NLTK) and the module Valence Aware Dictionary for sEntiment Reasoning (VADER).[43] NLTK uses advanced natural language processing (NLP) which focuses on processing large amounts of natural language data though statistical and machine learning. The VADER dictionary contains around 7,500 sentiment features in total and was developed to deal with the inherent nature of microblogging on social media sites like Instagram and Twitter (41). VADER blends sentiment lexicon approaches with grammatical rules for expressing polarity and intensity. A sentiment intensity scoring uses a scale from -4 (most negative) to +4 (most positive). Words with negative sentiment such as “horrible” have a score of -2.5 whereas “great” has a score of 3.1. Any word not listed in the dictionary will be scored as 0 or neutral. The scoring of each word was constructed by multiple human raters and then adjusted by the average rating score. Furthermore, VADER includes scoring of words frequently used in online conversations. VADER includes colloquialism such as the phrase “the bomb”, intense punctuation “!!!”, the uses of all caps for emphasis “this is VERY cool”, emoticons “☺”, as well as acronyms commonly used in online slang such as “LOL.” The sentiment of the sentence is calculated by summing up the values of all words within a caption and then is normalized between -1 to 1 using a function:

$$\frac{x}{\sqrt{x^2+\alpha}}, \tag{5}$$

where x is the sum of all the sentiment scores within the caption and α is a normalization parameter set to 15 within the VADER module.[43] Extended Table 1 provides a sample of the sentiment analysis results of all park captions.

Textual Analysis of Engaging Content

Our model to examine factors for engagement, i.e., the number if likes, is a non-negative count integer. The dependent variable acts as a count of the number of in-app individual users engaging with the content on Instagram. The distribution of the number of likes content receives is featured in extend figure 3. The engagement does not follow a normal distribution and is skewed right and the variance is 52 thousand times larger than the mean suggesting overdispersion in the distribution. Therefore, a negative binomial regression is the appropriate modeling framework (Wooldridge, 2010).

Engagement level of a photo is impacted by the timing of the upload, the image, the user and the information about the photo. We examine the factors impacting engagement by examining:

$$Likes_{pi} = \beta_0 + \beta_1 Location_{pi} + \beta_2 POI_{pi} + \beta_3 Things2Do_{pi} + \beta_4 Sentiment_i + U_i + \rho_p + \tau_{y(t)} + \gamma_t + \varepsilon_{pt}\,, \tag{6}$$

where $Location$ is a binary indicator for if the user i posted about park p by using the name or state where the park is located. POI is a binary indicator for if the user posted about points of interest determined by NPS. $Things2DO$ is a binary indicator for if the user posted about the activities listed on each park's website. $Sentiment$ is a categorical variable of the sentiment analysis which classified the caption as either positive, negative or neutral. ρ_p, $\tau_{y(t)}$, and γ_t represent park, year (y) and month fixed effects. Ideally, U_i would be a user fixed effect; however, we observe nearly 2.8 million unique individuals and therefore the computing effort is a cumbersome and lengthy task. Instead, we determine in-app behaviors among the individuals by examining how many times the user post an image under any geotag National Park. We do not have any personal information on each individual user nor do we know how many followers each user has but we do observe the number of times a user posts a photo using a National Park geotag. Those who post casually to Instagram or casually visited a National Park would appear nominally in our dataset. On the other hand, an avid National Park visitor and Instagram user would appear multiple times in the dataset. The average number of times an individual user posted at a National Park was 2 posts per user but the maximum number of posts uploaded for a user was 4,333. Therefore, U_i is a categorical count for each user determined by the number of times that user uploads content to Instagram using a National Park geotag.

Lastly, users who post a large volume of content associated with National Park geotags may indicate the users are influencers. Influencers are users on the platform who have a larger audience and typically promote lifestyles and brands. Using a similar method to isolate highly engaging content in eq. 1, users who posted in the top 90th percentile were identified. This required a count of 41 individual geotagged photos to NPS over the nine years. Modeling the avid users turns U_i in eq. 6 into a binary indicator given the categorical representation is perfectly correlated avid users and cannot be modeled together. The regression examines whether avid users are significant in generating engagement.

Avid users appear to have a higher rate of engagement. These user account for 27% of the content from the subset of highly engaging content determined by eq.1 and the threshold of 90^{th} percentile. Our last examination explores whether who is posting matters. Using eq. 4, the cumulative impacts from the viral grouping strategy compares across groups impact of the differing sets of engaging content. We first examine if the engaging content from avid users impacts the viral grouping strategy differently and compare it to engaging content not from avid users.

References

1. Clawson, M. Implications of Recreational Needs for Highway Improvements. in *Highway Research Board Bulletin* (1962).
2. Stevens, T. H., More, T. A. & Markowski-Lindsay, M. Declining National Park Visitation: An Economic Analysis. *J. Leis. Res.* **46**, 153–164 (2014).
3. Poudyal, N. C., Paudel, B. & Tarrant, M. A. A time series analysis of the impact of recession on national park visitation in the United States. *Tour. Manag.* **35**, 181–189 (2013).
4. Timmons, A. L. Too Much of a Good Thing: Overcrowding at America's National Parks. *Notre Dame Law Rev.* **94**,.
5. Robbins, J. How A Surge in Visitors Is Overwhelming America's National Parks. *Yale E360* https://e360.yale.edu/features/greenlock-a-visitor-crush-is-overwhelming-americas-national-parks (2017).
6. NPS. Deferred Maintenance and Repairs - Infrastructure (U.S. National Park Service). https://www.nps.gov/subjects/infrastructure/maintenance-backlog.htm.
7. Solomon, C. Is Instagram Ruining the Great Outdoors? *Outside Online* https://www.outsideonline.com/culture/opinion/instagram-ruining-great-outdoors/ (2017).
8. Simmonds, C., McGivney, A., Wilkinson, T. & Canon, G. Crisis in our national parks: how tourists are loving nature to death. *The Guardian* (2018).
9. Hegyi, N. Instagramming Crowds Pack National Parks : NPR. https://www.npr.org/2019/05/28/726658317/instagramming-crowds-pack-national-parks (2019).
10. Lake, Z. & Rulli, M. National parks officials grappling with high volume as Instagram tourism booms. *ABC News* https://abcnews.go.com/Lifestyle/instagram-tourism-booms-horseshoe-bend-national-parks-officials/story?id=64638198.

11. Lowe Mackenzie, A. L., Dundas, S. J. & Zhao, B. The Instagram Effect: Is Social Media Influencing Visitation to Public Land? *Land Econ.* 122920-0192R1 (2023) doi:10.3368/le.100.2.122920-0192R1.

12. Wood, S. A., Guerry, A. D., Silver, J. M. & Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **3**, 2976 (2013).

13. Sessions, C., Wood, S. A., Rabotyagov, S. & Fisher, D. M. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *J. Environ. Manage.* **183**, 703–711 (2016).

14. Fisher, D. M. *et al.* Recreational use in dispersed public lands measured using social media data and on-site counts. *J. Environ. Manage.* **222**, 465–474 (2018).

15. Wilkins, E. J., Wood, S. A. & Smith, J. W. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. *Environ. Manage.* **67**, 120–132 (2021).

16. Wood, S. A. *et al.* Next-generation visitation models using social media to estimate recreation on public lands. *Sci. Rep.* **10**, 15419 (2020).

17. Hausmann, A. *et al.* Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conserv. Lett.* **11**, e12343 (2018).

18. Walden-Schreiner, C., Leung, Y.-F. & Tateosian, L. Digital footprints: Incorporating crowdsourced geographic information for protected area management. *Appl. Geogr.* **90**, 44–54 (2018).

19. Walden-Schreiner, C., Rossi, S. D., Barros, A., Pickering, C. & Leung, Y.-F. Using crowd-sourced photos to assess seasonal patterns of visitor use in mountain-protected areas. *Ambio* **47**, 781–793 (2018).

20. Ghermandi, A. & Sinclair, M. Passive crowdsourcing of social media in environmental research: A systematic map. *Glob. Environ. Change* **55**, 36–47 (2019).

21. Miller, Z. D., Taff, B. D., Newman, P. & Lawhon, B. A Proposed Research Agenda on Social Media's Role in Visitor Use and Experience in Parks and Protected Areas. *J. Park Recreat. Adm.* (2019) doi:10.18666/JPRA-2019-9553.

22. Stigler, G. J. The Economics of Information. *J. Polit. Econ.* **69**, 213–225 (1961).

23. Leibenstein, H. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *Q. J. Econ.* **64**, 183 (1950).

24. McIntosh, C. R. & Wilmot, N. An Empirical Study of the Influences of Recreational Park Visitation: The Case of US National Park Service Sites. *Tour. Econ.* **17**, 425–435 (2011).

25. What Does the Future Hold for U.S. National Park Visitation? Estimation and Assessment of Demand Determinants and New Projections. *J. Agric. Resour. Econ.* (2020) doi:10.22004/ag.econ.298433.

26. Hobbs, K. G., Link, A. N. & Swann, C. A. The overcrowding of Zion National Park: is it a pricing problem? *J. Environ. Econ. Policy* **11**, 351–360 (2022).

27. Xiao, X., Lee, K. J. & Larson, L. R. Who visits U.S. national parks (and who doesn't)? A national study of perceived constraints and vacation preferences across diverse populations. *J. Leis. Res.* **53**, 404–425 (2022).

28. Weiler, S. A park by any other name: National Park designation as a natural experiment in signaling. *J. Urban Econ.* **60**, 96–106 (2006).

29. Weiler, S. & Seidl, A. What's in a Name? Extracting Econometric Drivers to Assess The Impact of National Park Designation*. *J. Reg. Sci.* **44**, 245–262 (2004).

30. NPS.gov Homepage (U.S. National Park Service). https://www.nps.gov/index.htm.

31. Allcott, H. & Gentzkow, M. Social Media and Fake News in the 2016 Election. *J. Econ. Perspect.* **31**, 211–236 (2017).

32. MacCarthy, M. Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry. SSRN Scholarly Paper at https://doi.org/10.2139/ssrn.3615726 (2020).

33. Flores, D. & Kuhn, K. Latino Outdoors: Using Storytelling and Social Media to Increase Diversity on Public Lands. *J. Park Recreat. Adm.* **36**, (2018).

34. STATS - Welcome to Visitor Use Statistics. https://irma.nps.gov/Stats/.

35. Federal Reserve Economic Data | FRED | St. Louis Fed. https://fred.stlouisfed.org/.

36. Bureau, U. C. Data. *Census.gov* https://www.census.gov/data.

37. Homepage - U.S. Energy Information Administration (EIA). https://www.eia.gov/.

38. PRISM Climate Group at Oregon State University. https://prism.oregonstate.edu/.

39. Dundas, S. J. & von Haefen, R. H. The Effects of Weather on Recreational Fishing Demand and Adaptation: Implications for a Changing Climate. *J. Assoc. Environ. Resour. Econ.* **7**, 209–242 (2020).

40. Dundas, S. J. & von Haefen, R. H. The importance of data structure and nonlinearities in estimating climate impacts on outdoor recreation. *Nat. Hazards* **107**, 2053–2075 (2021).

41. Gentzkow, M., Kelly, B. & Taddy, M. Text as Data. *J. Econ. Lit.* **57**, 535–574 (2019).

42. Dugoua, E., Dumas, M. & Noailly, J. Text as Data in Environmental Economics and Policy. *Rev. Environ. Econ. Policy* **16**, 346–356 (2022).

43. Hutto, C. & Gilbert, E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proc. Int. AAAI Conf. Web Soc. Media* **8**, 216–225 (2014).

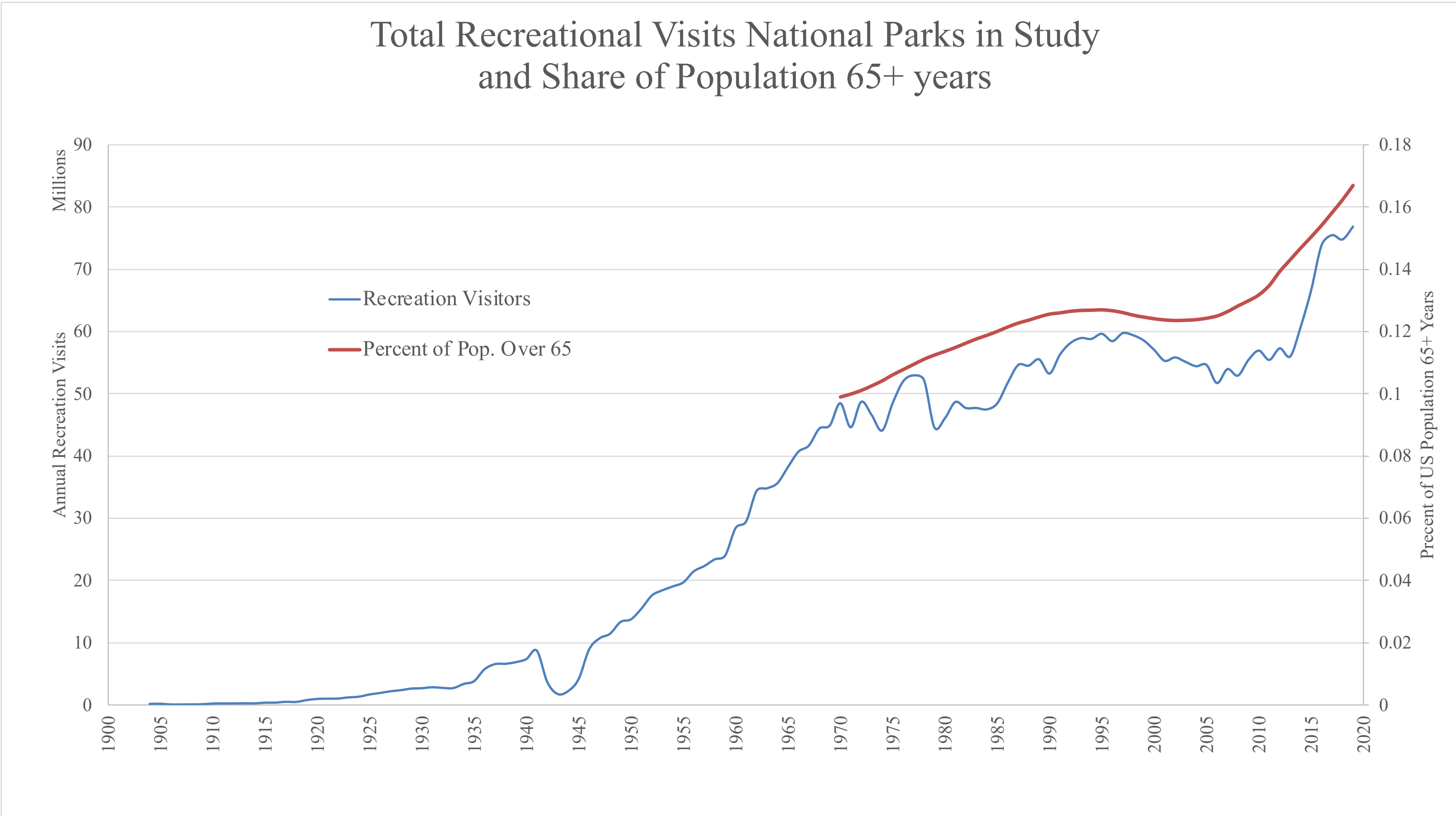


Figure 1: Total and average recreational visits at U.S. National Parks 1904 to 2020. Left-side y-axis scale is total recreation visitors and right-side y-axis scale average recreation visitors.

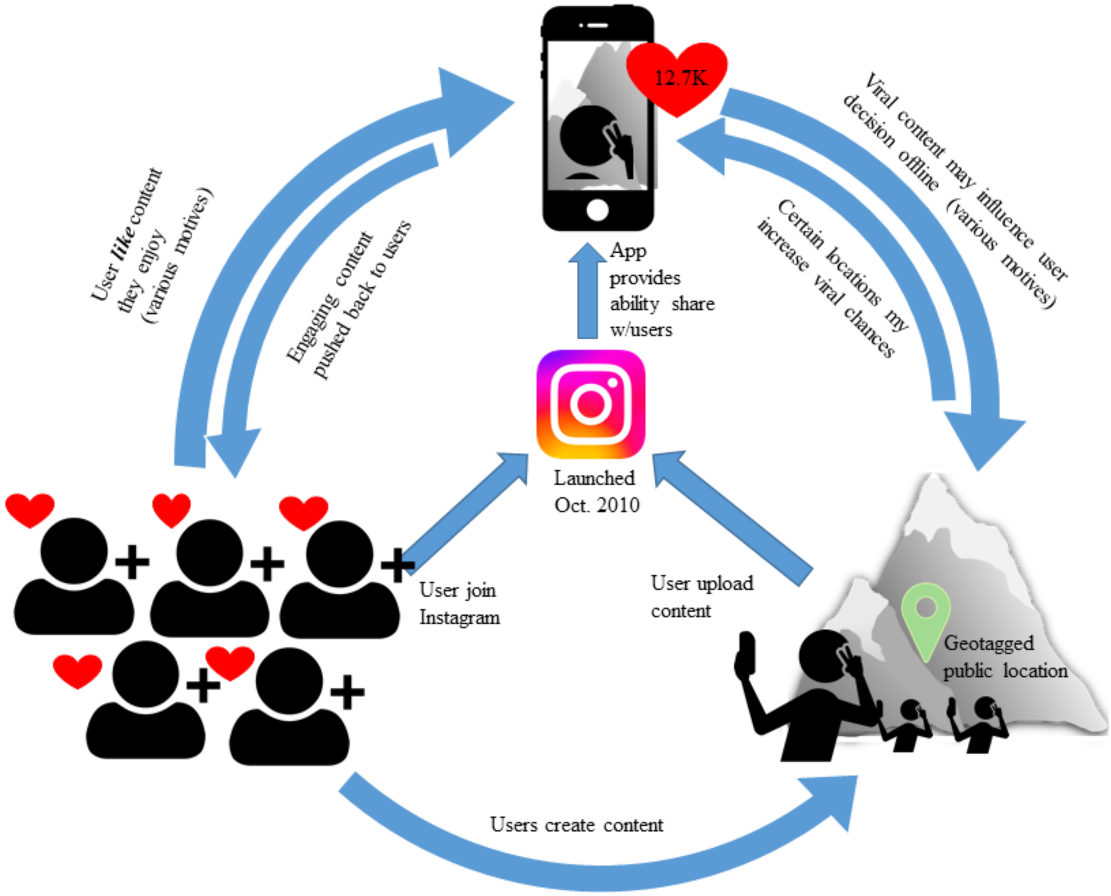


Figure 2: Conceptual diagram of potential mechanism of Instagram's influence of visitation to public lands. When social media sites are launched it takes time to develop a user base who upload content and engage with others content. Instagram was launched in October 2010 and reached 50million users by April 2012. As content is uploaded to the platform certain content is engaged with at higher levels than other. There are many motivations to engage with content highly and this influential content likely influences some decision of users. In context of visitation to public lands, viral content could influence those exposed to the image to want to see the location. Content with locational information, such as GPS through geotagging, provides information on where to go to get a similar photo and the quality of the location.

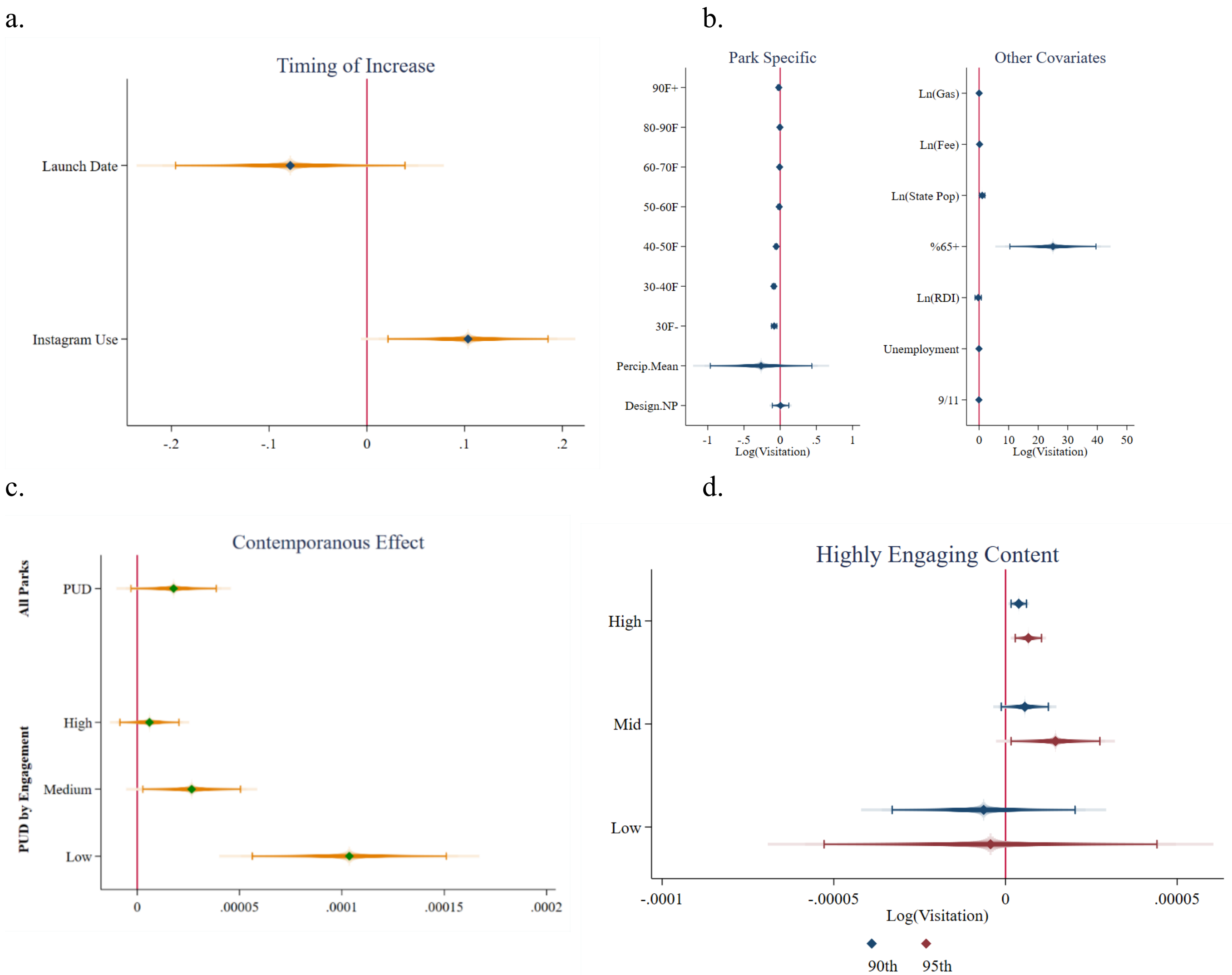


Figure 3: Factors impacting visitation trends. Panel (**a**) examines the timing of the recent increase in visitation. Panel (**b**) plots all covariates estimated as controls for the changes in visitation at national parks. Panel (**c**) plots the contemporaneous effect of users uploading geotagged content to Instagram per month for all parks and through groups determined by viral moments plotted in figure 4. Panel (**d**) defining engaging content by eq.1 plots the cumulative impact of engaging content by park group in figure 3.

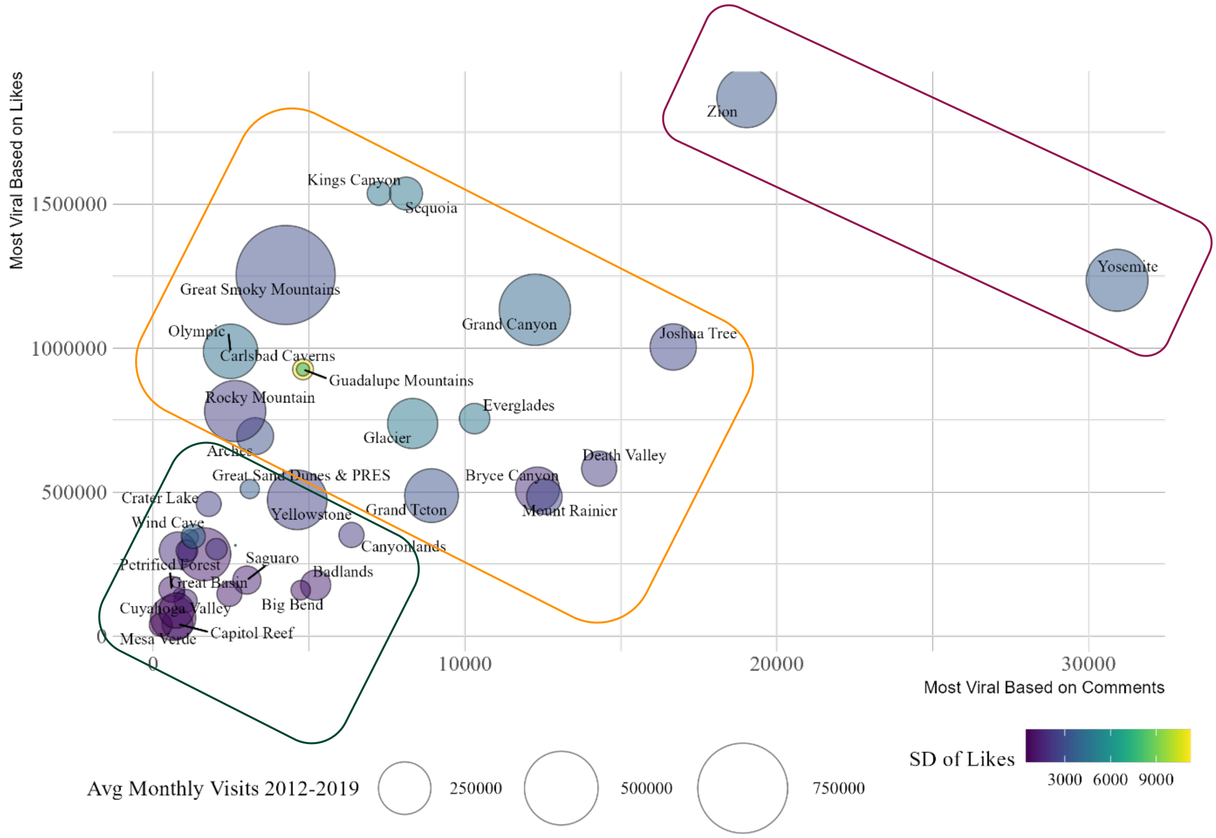


Figure 4: Modeling Viral Moments and Visitation Indicators

Y-axis is the max number of likes that the most liked photo from each park received. X-axis is the max number of comments the most commented photo from each park received. The size of the circle indicates average monthly visitation to park. Color represent the standard deviation (SD) of the number of likes a park recieves on all photos uploaded.The dark purple has a lower SD and the yellow color have higher SD. Caverns & Guadalupe Mountains Parks had the highest SD suggesting the likes received on photos at these parks are more varied . Parks fan out from the origin as their highest viral content recieves more engagement. Three grouping stragies appear. The High Engagement Group had the most viral content on the app; Zion & Yosemite. The Middle Engagement Group had simi viral content; Kings Canyon, Sequoia, Grand Canyon, Joshua Tree, Great Smoky Mountians, Everglades & Death Valley, Olympic, Carlsbad Caverns, Guadalupe Mountains, Glacier, Grand Teton, Bryce Canyon, Mount Rainer, Rocky Montain and Arches**.**

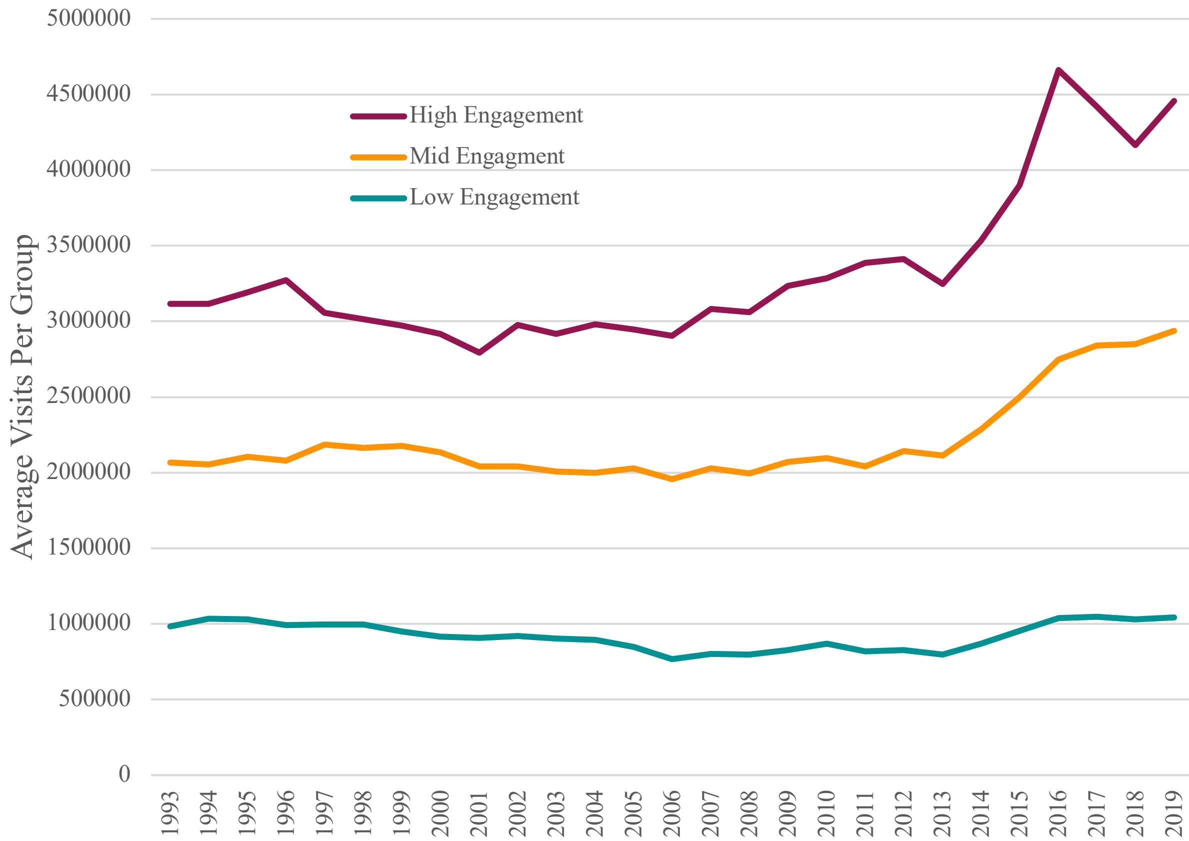
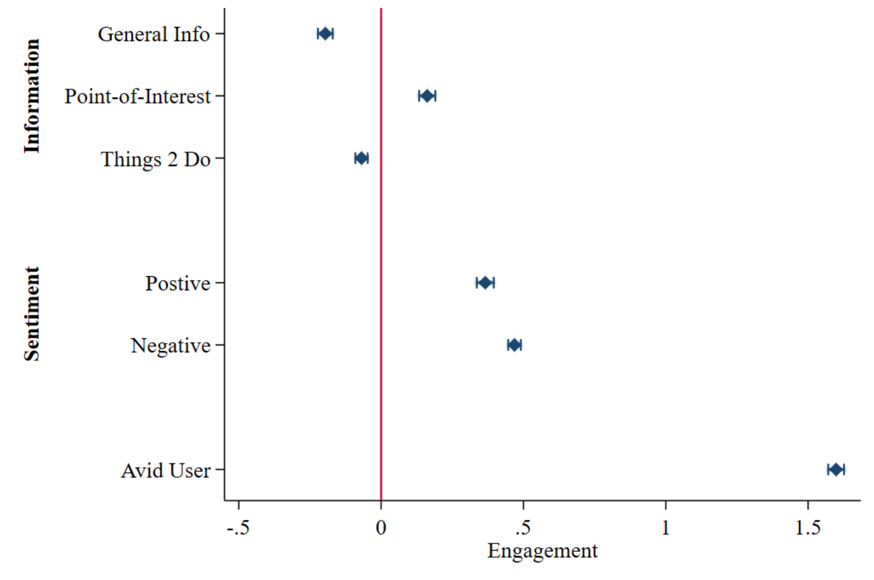


Figure 4: Visitation Growth by Viral Group Y-axis is the average visits per engagement group. X-axis is the time period covered in the analysis. The high engagement group sees more on average visitation than groups and saw sharp increase post Instagram. The middle engagement group had some viral content and saw on average growth post Instagram. The low engagement group saw very little growth on average.

a.



b.

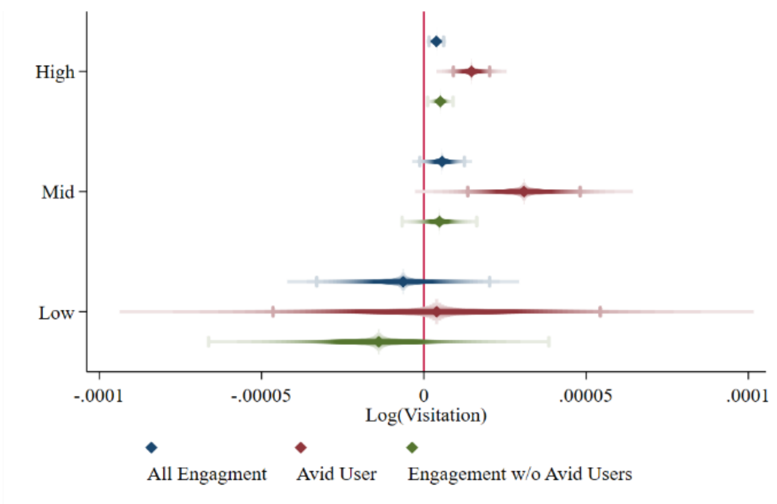
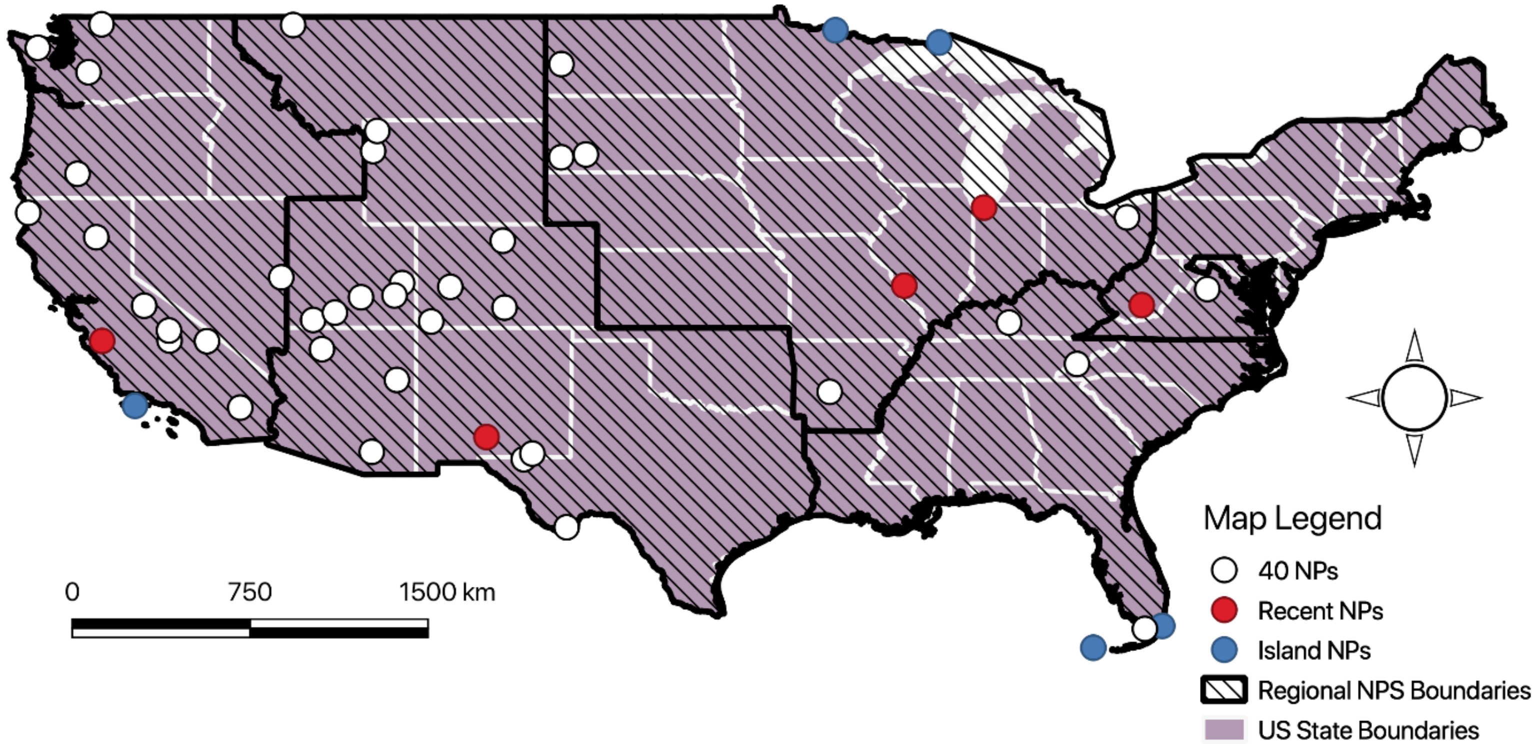
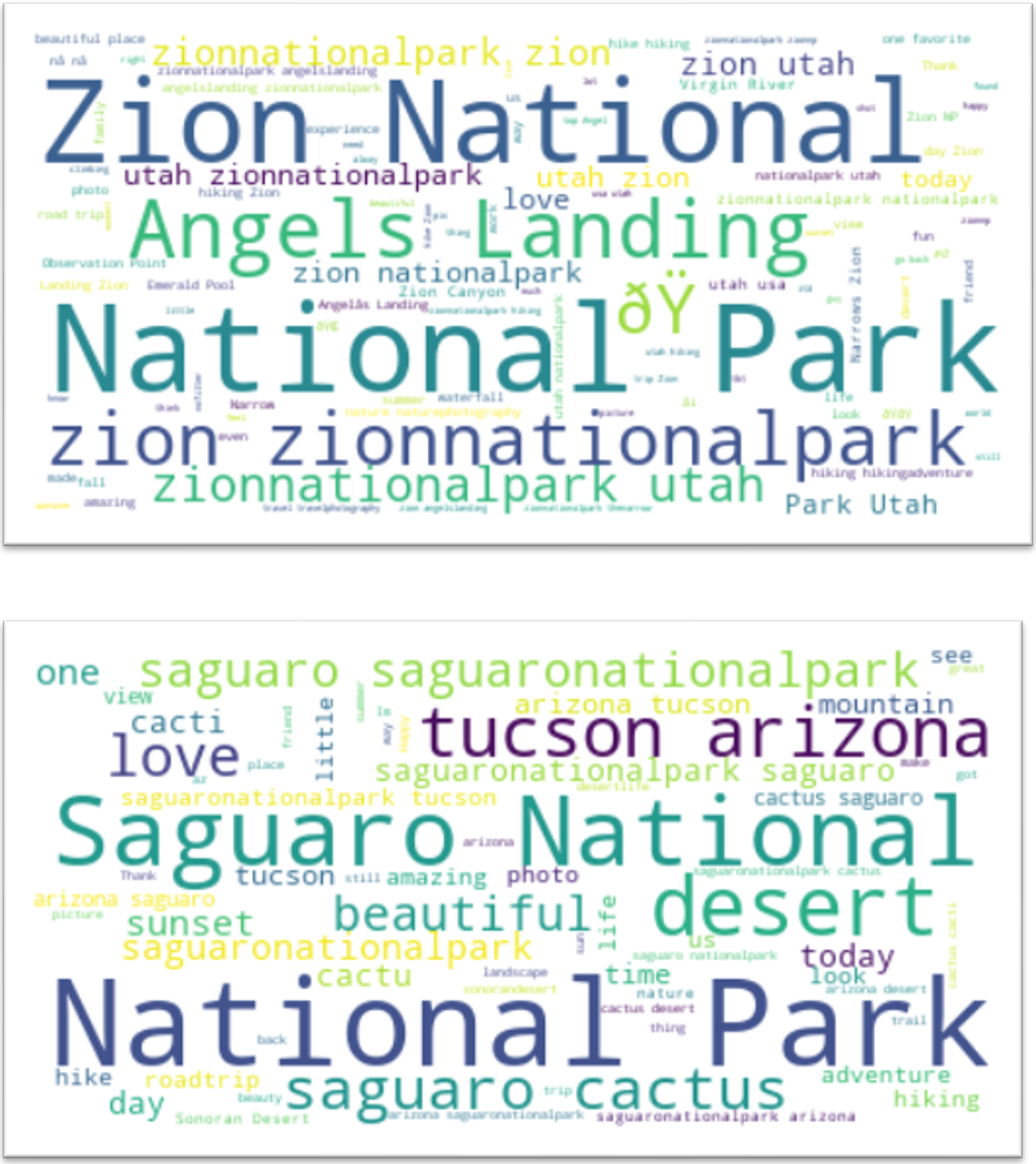


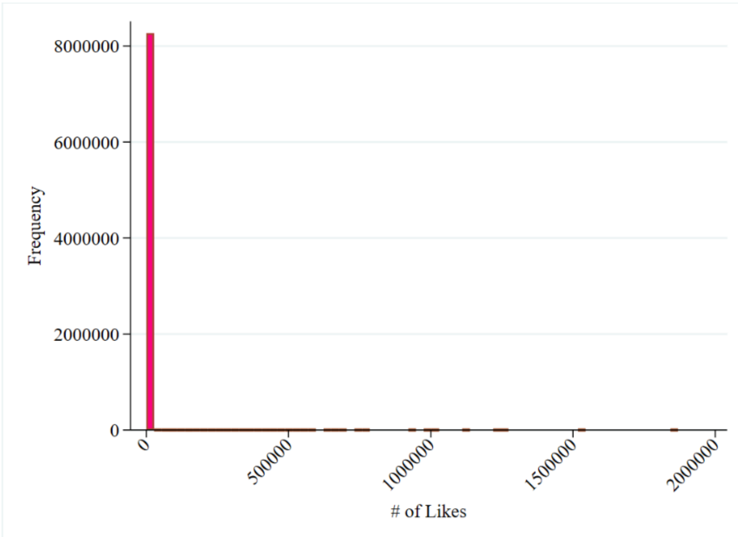
Figure 5: Factors for Engagement and the effect National Parks across Viral Moment Group. Panel (a) To access the factors which increased engagement we examine whether information, sentiment or user behavior are driving forces for engagement. The incident rate of the negative binomial count model suggests a 5 times higher than all other users. Panel (b) examines whether these factors matter across parks by examining all engagement from the 90th percentile of highly engaging content (blue), avid users highly engaging content (red) and the remaining content from all other users (green).



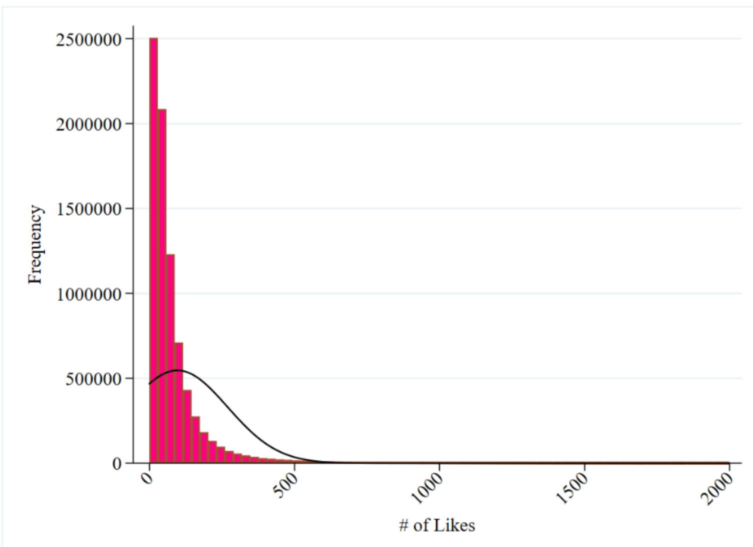
Extended Data Figure 1: Map of National Parks in the Continental United States of America. Map provides a description of the NPs within the study. The 40 NPs are in white circles. The red and blue circles are not included in the study. Red NPs were recently classified as NPs. Blue NPs are parks which require waterway travel to explore. Regional boundaries are outlined in black. This includes Pacific West, Intermountain, Midwest, Southeast, and Northeast. State boundaries are in white and filled in light purple



Extended Data Figure 2: Example of Wordclouds from 2 Parks. Zion Naitonal Park wordcloud of word frequency in captions (top). Saguaro National Park worldcloud of word frequency in captions (bottom). Font size is weighted on frequency. Symbool in Zion is an emoji code.



3



Extended Figure 3: Frequency of Content based on the # of Likes

Note: Figure on top shows the frequency of all content based on the level the content received in the forms of likes. Figure on the bottom shows distribution of frequency of content receiving at least 2000 likes. Black Line is the normal density

Extended Table 1 Descriptive Instagram Statistics for 40 National Park

Quantify Set	N	*Likes* to Reach 90th percentile	*Likes* to Reached 95th percentile	*Likes* to Reached 99th percentile
Entire Set of Geotag Photos				
2010-2019	8,310,295	207	347	2,047
Per Year Geotag Photos				
2010	10	11	12	13
2011	706	22	43	307
2012	9,383	39	67	269
2013	22,723	57	94	346
2014	90,157	95	158	737
2015	215,324	129	223	1,114
2016	1,087,549	163	284	1,460
2017	2,267,731	211	371	1,885
2018	2,462,000	227	425	2,436
2019	2,153,712	216	395	2,304

Extend Table 2 Textual Analysis and Sentiment Analysis w/ Natural Language Toolkit

NationalPark	Caption Uncleaned	LocationInfo	POI	Things-2-Do	Sentiment Analysis
Acadia	#view of #TheBowl on our #hike up! #amainezing #acadia #acadianationalpark #beautifulday #beehivetrail #ocean #instasky #sundayfunday #maine #newengland #ig_newengland	1	1	1	Neutral
Arches	Day trip to #Moab Delicate Arch is looking fine... #iphoneonly #landscape #nature #mountains #moab #arches	1	1	0	Positive
Badlands	Bad bad lands badlands	1	0	0	Negative
Bryce Canyon	a photo just doesn't do it justice	0	0	0	Negative
Everglades	#everglades #alligators #alligator #dead #vulture #florida	1	1	0	Negative
Glacier	⛰️You don't have to move mountains. Simply fall in love with life. ❤️ Be a tornado 🌪 of happiness ☺️, gratitude 🙏🏾 and acceptance🧚🏾‍♀️. You will change the world 🌎 just by being a warm, kind-hearted human being tbt to this past weekend in montana	1	1	0	Positive
North Cascades	All smiles on the ridge of Sharkfin Tower a few weeks back, smoky skies and all.	0	1	0	Positive
Yellowstone	another ridiculous sunset in the stone	0	0	1	Negative
Yosemite	Rockclimbing El Capitan	0	1	1	Neutral
Zion	i miss snow & zion	1	1	1	Negative

VADER

Note: Example of textual analysis results from both park-based dictionaries and identifying location, points of interest, things to do and the results of classification based on the sentiment analysis.

Extended Table 3 Visitation and Instagram Data with Grouping 1

	(1) N	(2) mean	(3) sd	(4) min	(5) max
Panel A: Entire Set of Parks Pre Instagram Launch (Oct 2010)					
Visits per Month	8,520	124,136	184,863	0	1,761,918
Panel B: Entire Set of Parks Post Instagram Launch					
Visits per Month	4,440	140,647	211,830	0	1,613,133
Panel C: High Activity Parks – Post Instagram Launch					
Visits per Month	220	323,238	181,842	56,972	780,728
Geotagged Content Uploads	220	6,768	9,867	0	39,757
Influential Posts top 90^{th}	220	960.5	1,403	0	4,936
Cumulative top 90^{th}	220	20,099	33,463	0	141,213
Influential Posts top 95^{th}	220	502.6	752.1	0	2,860
Cumulative top 95^{th}	220	10,530	17,961	0	76,658
Influential Posts top 99^{th}	220	112.9	178.5	0	751
Cumulative top 99^{th}	220	2,374	4,216	0	18,245
Panel D: Low Activity Parks - Post Instagram Launch					
Visits per Month	4,162	132,032	209,883	0	1,613,133
Geotagged Content Uploads	4,162	1,352	3,176	0	28,236
Influential Posts top 90^{th}	4,162	149.7	356.9	0	2,939
Cumulative top 90^{th}	4,162	3,334	8,822	0	90,106
Influential Posts top 95^{th}	4,162	73.43	170.8	0	1,346
Cumulative top 95^{th}	4,162	1,635	4,233	0	38,831
Influential Posts top 99^{th}	4,162	14.01	33.51	0	356
Cumulative top 99^{th}	4,162	311.1	804.7	0	6,649

Extend Table 4 Pre-Instagram Summary Statistics for Individual Parks

National Park	N	Annual Visitation 1993-2010	Acres	Trail Miles	# Campgrounds	# Things to do	# of POI
Acadia NP	216	204,641	49,088	178	4	21	132
Arches NP	216	69,664	76,519	16	1	16	121
Badlands NP	216	81,125	244,000	23	0	12	56
Big Bend NP	216	27,570	801,163	172	3	16	160
Black Cany. Gun. NP	216	16,278	30,750	17	3	17	83
Bryce Canyon NP	216	89,048	35,835	68	2	18	388
Canyonlands NP	216	34,651	267,640	186	2	14	234
Capital Reef NP	216	4,9802	241,904	124	1	14	140
Carlsbad Caverns NP	216	40,732	46,766	37	0	8	43
Crater Lake NP	216	37,052	183,224	90	2	17	102
Cuyahoga Valley NP	216	248,863	33,000	125	0	16	110
Death Valley NP	216	81,216	3,373,063	248	7	14	188
Everglades NP	216	82,823	1,508,976	311	2	13	989
Glacier NP	216	157,892	1,012,837	697	13	22	373
Grand Canyon NP	216	365,353	1,218,375	501	6	20	853
Grand Teton NP	216	214,435	310,000	263	8	21	1471
Great Basin NP	216	6,999	77,180	73	5	17	20
Great Sand Dunes NP	216	23,509	107,342	19	1	17	44
Great Smoky Mt. NP	216	785,060	522,427	843	10	12	2453
Guadalupe Mt. NP	216	16,522	86,416	84	3	10	84
Hot Springs NP	216	117,801	5,550	25	1	17	12
Joshua Tree NP	216	106,498	795,156	251	8	16	31
Kings Canyon NP	216	48,564	461,901	357	7	14	90
Lassen Volcanic NP	216	31,420	106,452	167	7	19	174
Mammoth Cave NP	216	128,488	52,830	79	3	15	286
Mesa Verde NP	216	46,370	52,485	26	0	12	177
Mount Rainier NP	216	105,592	236,380	230	3	17	210
North Cascades NP	216	2,109	504,654	166	0	16	2
Olympic NP	216	271,638	922,650	586	15	22	386
Petrified Forest NP	216	57,599	221,391	6	0	12	75
Redwood NP	216	34,309	131,983	256	4	12	228
Rocky Mountain NP	216	244,785	265,807	329	5	13	1384
Saguaro National	216	58,773	91,716	190	0	6	82
Sequoia NP	216	78,578	202,430	449	7	14	126
Shenandoah NP	216	116,056	199,173	521	5	14	293
Theodore Roos. NP	216	39,711	70,446	88	3	15	204
Wind Cave NP	216	58,246	10,522	30	1	4	15
Yellowstone NP	216	252,569	2,219,791	1,044	12	22	1357
Yosemite NP	216	298,791	747,956	813	13	24	429
Zion NP	216	208,974	148,016	96	3	16	626
Average		123,503	441,845	245	4	15	356

Note: **Things to Do** listed on website: auto touring, backpacking, backcountry camping, biking, birdwatching, boating, campground, cave, canyoneering, commercial tours, picnicking, fishing, hiking, hot Springs, kayaking/rafting, historical places, horseback riding, lodge, museum, pet Friendly, photography, ranger-led Programs, religious, climbing, stargazing, snowmobiling, sandboarding, swimming, sunset & sunrise, tidepools, visitor center, viewpoints, wildlife viewing, winter activities, wildflower viewing. **Campgrounds** listed on Website**.**

Extend Table 5 Post-Instagram Summary Statistics for Individual Parks

Nation Park	# Months w/ Instagram Data	Monthly Visitation	Influential Posts/Month (90th /yr)	Total Influential (90th/yr)	Maximum Likes	Total Uploads Unique Day & User
Acadia NP	97	242,801	192	18,613	284,010	209,634
Arches NP	105	114,132	291	30,566	694,191	266,708
Badlands NP	96	79,023	50	4,834	176,756	62,927
Big Bend NP	90	31,474	82	7,423	159,201	73,228
Black Canyon NP	82	20,518	28	2,288	341,852	29,805
Bryce Canyon NP	105	160,981	223	23,432	510,555	238,152
Canyonlands NP	94	51,462	139	13,037	350,921	110,029
Capitol Reef NP	72	77,802	5	370	41,656	3,819
Carlsbad Caverns NP	77	35,843	8	644	926,178	18,778
Crater Lake NP	99	50,345	125	12,383	458,431	126,203
Cuyahoga Valley NP	96	185,395	42	4,030	63,032	51,102
Death Valley NP	96	103,243	321	30,853	581,104	230,258
Everglades NP	101	83,117	77	7,775	755,131	86,925
Glacier NP	94	214,610	471	44,235	738,180	282,109
Grand Canyon NP	105	445,748	858	90,106	1,133,472	852,763
Grand Teton NP	99	253,765	414	40,951	487,677	268,444
Great Basin NP	79	10,191	13	1,033	146,313	12,241
Great Sand Dunes NP	99	29,579	138	13,662	510,526	113,129
Great Smoky Mts NP	91	884,092	204	18,557	1,254,079	202,704
Guadalupe Mts NP	87	14,465	15	1,325	926,105	16,886
Hot Springs NP	102	119,881	28	2,901	297,001	44,152
Joshua Tree NP	103	176,678	722	74,375	1,004,006	536,917
Kings Canyon NP	99	49,327	96	9,536	1,537,371	107,161
Lassen Volcanic NP	88	38,400	48	4,196	296,057	48,853
Mammoth Cave NP	95	44,765	5	447	122,220	12,944
Mesa Verde NP	89	45,252	10	881	38,770	19,185
Mount Rainier NP	90	106,762	276	24,859	484,129	193,423
North Cascades NP	87	2,218	172	14,948	315,014	77,874
Olympic NP	99	264,138	163	16,130	988,961	124,267
Petrified Forest NP	91	56,603	18	1,677	162,082	27,610
Redwood NP	100	37,508	70	7,037	301,780	82,960
Rocky Mountain NP	104	325,960	413	42,948	781,421	450,685
Saguaro NP	104	65,856	52	5,443	194,159	53,834
Sequoia NP	99	94,265	263	26,046	1,537,167	259,294
Shenandoah NP	100	108,522	87	8,740	88,544	112,847
Theodore Roosevelt NP	95	53,626	20	1,928	146,267	20,600
Wind Cave NP	88	48,629	6	546	346,087	7,411
Yellowstone NP	99	316,206	144	14,253	472,793	189,590
Yosemite NP	106	345,619	1,332	141,213	1,235,720	938,188
Zion NP	101	306,049	694	70,104	1,869,267	550,791
Average	95	142,371	207	20,858	568 954	177,861

Note: Number of observations per month impacted when the initial scraping of each park took place. Influential post per month are average. Unique uploads calculated for each park by user.

Extend Table 5 Textual Analysis Indicators by Park

Park Name	Location	POI	Things-2-Do	Positive	Negative	Neutral
Acadia	137,174	143,295	79,406	0.141	0.024	0.742
Arches	186,935	235,257	120,916	0.110	0.020	0.780
Badlands	44,136	41,950	24,767	0.104	0.031	0.759
Big Bend	53,980	65,088	32,754	0.077	0.018	0.905
Black Gunnison	22,223	18,173	12,637	0.109	0.022	0.772
Bryce Canyon	185,162	215,621	106,156	0.103	0.016	0.783
Canyonlands	74,638	94,777	40,372	0.104	0.021	0.798
Capitol Reef	2,583	3,387	1,554	0.115	0.018	0.788
Carlsbad Caverns	13,732	15,098	14,150	0.102	0.021	0.750
Crater Lake	93,056	98,109	48,285	0.117	0.021	0.776
Cuyahoga Valley	26,040	29,872	17,370	0.152	0.024	0.738
Death Valley	166,661	89,022	83,503	0.093	0.087	0.742
Everglades	63,822	81,986	28,215	0.100	0.019	0.784
Glacier	163,647	229,651	110,520	0.125	0.023	0.776
Grand Canyon	584,510	622,181	279,873	0.192	0.022	0.692
Grand Teton	175,729	237,309	120,042	0.141	0.021	0.772
Great Basin	9,248	9,289	7,497	0.175	0.028	0.730
Great Sand Dunes	86,797	86,923	39,805	0.165	0.028	0.725
Great Smoky Mt.	108,961	197,721	64,351	0.167	0.026	0.709
Guadalupe Mt.	12,998	14,936	8,099	0.132	0.032	0.753
Hot Springs	24,788	1,887	24,896	0.142	0.024	0.723
Joshua Tree	345,153	113,429	220,929	0.140	0.028	0.765
Kings Canyon	63,458	43,393	42,022	0.134	0.025	0.756
Lassen Volcanic	32,056	42,818	24,530	0.139	0.030	0.738
Mammoth Cave	10,095	11,820	10,144	0.132	0.025	0.746
Mesa Verde	14,314	17,324	5,131	0.117	0.027	0.736
Mount Rainier	134,591	42,062	89,649	0.156	0.025	0.741
North Cascades	42,774	272	28,797	0.151	0.031	0.764
Olympic	67,055	76,001	58,917	0.109	0.020	0.793
Petrified Forest	19,693	15,203	8,172	0.107	0.084	0.683
Redwood	47,707	24,424	21,639	0.130	0.025	0.748
Rocky Mt.	284,613	437,760	184,815	0.124	0.021	0.777
Saguaro	36,129	40,706	14,319	0.124	0.025	0.764
Sequoia	162,364	13,068	86,977	0.128	0.022	0.762
Shenandoah	60,164	83,705	57,374	0.148	0.025	0.746
Theodore Roos.	14,611	19,589	8,454	0.137	0.030	0.737
Wind Cave	5,569	4,968	5,195	0.121	0.024	0.763
Yellowstone	133,857	172,788	69,173	0.137	0.022	0.718
Yosemite	625,967	849,925	410,561	0.111	0.019	0.790
Zion	396,378	474,316	272,552	0.151	0.028	0.745
Average	118,334	125,378	72,113	0.129	0.027	0.757

Note: First 3 columns are sum of photos identified using park-based dictionaries. Positive and negative score is based on average of all photos. Interpretation of this number can assume a majority of post are neutral 75%, roughly 13% are positive and .2% are negative, the remaining percentage have *no caption.*