AGILE: GIScience Series, 2, 14, 2021. https://doi.org/10.5194/agile-giss-2-14-2021
Proceedings of the 24th AGILE Conference on Geographic Information Science, 2021.
Editors: Panagiotis Partsinevelos, Phaedon Kyriakidis, and Marinos Kavouras.
This contribution underwent peer review based on a full paper submission.

A Socially Aware Huff Model for Destination Choice in Nature-based Tourism

Meilin Shi[a,b](corresponding author), Krzysztof Janowicz[a,b], Ling Cai[a,b], Gengchen Mai[a,b] and Rui Zhu[a,b]

meilinshi@ucsb.edu, janowicz@ucsb.edu, lingcai@ucsb.edu, gengchen_mai@ucsb.edu, ruizhu@ucsb.edu

[a] STKO Lab, Department of Geography, University of California, Santa Barbara, USA
[b] Center for Spatial Studies, University of California, Santa Barbara, USA

Abstract. Identifying determinants of tourist destination choice is an important task in the study of nature-based tourism. Traditionally, the study of tourist behavior relies on survey data and travel logs, which are labor-intensive and time-consuming. Thanks to location-based social networks, more detailed data is available at a finer grained spatio-temporal scale. This allows for better insights into travel patterns and interactions between attractions, e.g., parks. Meanwhile, such data sources also bring along a novel social influence component that has not yet been widely studied in terms of travel decisions. For example, social influencers post about certain places, which tend to influence destination choices of tourists. Therefore, in this paper, we propose a socially aware Huff model to account for this social factor in the study of destination choice. Moreover, with fine-grained social media data, interactions between attractions (i.e., the neighboring effects) can be better quantified and thus integrated into models as another factor. In our experiment, we calibrate a model by using trip sequences extracted from geotagged Flickr photos within two national parks in the United States. Our results demonstrate that the socially aware Huff model better simulates tourist travel preferences. In addition, we explore the significance of each factor and summarize the spatial-temporal travel pattern for each attraction. The socially aware Huff model and the calibration method can be applied to other fields such as promotional marketing.

Keywords. nature-based tourism, socially aware Huff model, tourist destination choice, geotagged social media, Flickr

1 Introduction

Nature tourism, i.e., tourism that is based on the natural attractions of an area, has gone through rapid growth over the past two decades (Balmford et al., 2009), especially for national parks in the United States, according to visitation statistics by *National Park Service*.[1] Identifying and evaluating relevant determinants of tourist flows is important. On the one hand to promote tourism, and on the other hand it helps to protect natural lands. Prior work on nature-based tourism relies on manually collected travel logs and survey data, which are time-consuming, labor-intensive, and limited in temporal coverage (Puustinen et al., 2009; Nahuelhual et al., 2013). The emergence of location-based social networks (LBSNs) and volunteered geographic information (VGI), such as Flickr, Instagram, Facebook etc., together with geotagging technology, provides more fine-grained spatial and temporal data, which equips us with a new lens to understand travel patterns as they relate to natural attractions.

Additionally, LBSNs play an increasingly important role in travel decision making process (Leung et al., 2013). For example, places like Horseshoe Bend, Devil's Bathtub, etc., once being hidden gems, are now receiving a large number of visitors annually. Social media has been regarded as the main culprit for the sudden and overwhelming popularity of these places (Djossa, 2019). More specifically, geotagged photos posted by social media influencers (SMIs) can rapidly attract new visitors (Glover, 2009). These influencers are usually users with a large number of followers and have established credibility in certain fields that can shape attitudes of tourists and thus influencing their travel preferences (Freberg et al., 2011; Li, 2016). Intuitively, a scenic photo posted by a user with 50k followers has a much broader potential influence than a user with 50 followers. Therefore, we argue that social factors brought by increasingly used social media need to be taken into account as a new norm to complement traditional destination choice models. To justify such an argument, we specifically explore this social effect in nature-based tourism destination choices, because tourists tend to share geotagged photos on social media platforms along their trips (Tasse et al., 2017).

Moreover, existing work has shown that fine-grained spatio-temporal data collected from social media can be used to quantify visitation rates (Wood et al., 2013), to estimate visitor flows (Orsi and Geneletti, 2013; Kim et al., 2019), and to detect popular sub-regions and temporal activity patterns (Heikinheimo et al., 2017). These studies illustrate the capability of using LBSNs to capture temporal variations in tourist visiting, with some places (e.g., dive resorts) being more attractive in summer and others (e.g., ski resorts) in winter. In addition, interactions between places (i.e., the neighboring effects) can be better quantified with social media data. For example, we can estimate the interactions between Horseshoe Bend and those attractions in its surroundings (Glen Canyon, Antelope Canyon, Grand Canyon, etc.), based on which we can further explore how they affect potential travel decisions of tourists to Horseshoe Bend, thereby uncovering interesting travel patterns that are difficult to detect using traditional data.

To explore how social factors and neighboring effects contribute to tourist destination choice in natural attractions, we propose a socially aware version of the well-known Huff model (Huff, 1964), which was originally used to calculate the probability of a customer shopping at each retail store.

The contributions of this work are as follows:

- We propose a socially aware Huff model, which incorporates both social factors and neighboring effects, to estimate the probability of tourists visiting specific places.
- The proposed method is calibrated on a data set containing 10-year geotagged Flickr photos in two national parks, whose results outperform the baseline Huff model.
- We explore the spatial and temporal variability of model parameters that are associated with attractiveness, distance, and neighboring effect in the socially aware Huff model.

The remainder of this paper is organized as follows. Section 2 introduces related work on tourism, geo-social media, and the Huff model. Section 3 briefly introduces the Flickr data set used for the study and explains the trip reconstruction process. A socially aware Huff model is introduced in section 4 together with the model calibration method. In section 5, we present the model calibration results and explain the spatial and

[1] https://irma.nps.gov/Stats/

temporal variability of the parameters used in the socially aware Huff model. Finally, we summarize our findings and discuss future directions in section 6.

2 Related Work

2.1 Tourism and Geo-social Media

The use and role of social media has been widely discussed in tourism research (Leung et al., 2013). Work by Zheng et al. (2012) used Flickr data to discover regions of attractions (RoAs) and explored tourists movement patterns in relation to the RoAs. Similarly, Hu et al. (2019) extracted popular attractions and tour routes using a graph-based network in New York City from Twitter data. Majid et al. (2013) proposed a context-aware personalized travel recommendation system and evaluated it based on a Flickr data set. Li et al. (2018) used Flickr data to compare the spatial overlap of tourists' and locals' destinations in ten US cities. Work by Mou et al. (2020) analyzed spatio-temporal distribution and changes of inbound tourism flow in Shanghai with Flickr data.

In the past decade, social media has evolved into an important player in tourism advertising and promotion (Bakr and Ali, 2013). Litvin et al. (2008) showed that travelers are increasingly influenced by electronic Word-of-Mouth (eWOM) from social media. Parsons (2017) echoed the similar idea that Instagram influences tourist decision-making, especially for younger generations. Jalilvand and Samiei (2012) examined the influence of eWOM and showed that it has a significant impact on tourist attitudes towards visiting Isfahan, Iran. Tham et al. (2020), however, conducted interviews with tourist decision-makers in Australia and revealed that social media's role appears to have only moderate-low influence on destination choice. In line with such research, we include a social influence factor and examine the impact of social media on destination choices. We quantify the social impact that influencers could bring to a place by measuring the place attractiveness given the travel preference of tourists.

2.2 Huff Model

There have been many research efforts towards tourist destination choice and sequential tourist flows (Nicolau and Más, 2008; Wu et al., 2012; Yang et al., 2013). The Huff model (Huff, 1964) is one of them, though it was originally developed to predict retail sales and consumer behavior. The Huff model estimates the probability of consumer patronizing retail stores based on two factors: attractiveness of a store and travel cost, which can also be applied to tourism research. Misui and Kamata (2016) adopted the Huff model to show the effect of travel time on visiting probability to spa destinations in Japan. Similarly, Nicolau (2008) studied tourist sensitivities to distance and price for destination choice in Spain using a national tourist choice behavior survey data. Yang et al. (2013) conducted a logistic model to study the inter-dependencies among destination choices when two or more destinations are included in a trip, accounting for the future dependency in the multi-destination choice behaviors. Recently, more work has shown the importance of the temporal factor that is missing in the original Huff Model. Gong et al. (2020) included weekday and weekend variations when calculating visiting probability of shopping areas using taxi trajectory data in Shenzhen and New York. Liang et al. (2020) proposed a T-Huff model and proved that it outperforms the original static Huff model when estimating temporal store visits using SafeGraph POI visits data. Likewise, we include the temporal factor in our study given the availability of social media data.

3 Data and Trip Reconstruction

In this section, we introduce the data set used for the study in section 3.1 and explain how we reconstruct trips from the geotagged photos step by step in section 3.2.

3.1 Data and Study Area

In this study, we collected geotagged Flickr photos of tourist attractions within national parks using Flickr's public API.[2] Two national parks - Acadia National Park and Yosemite National Park - have been selected from the top ten most visited national parks in the United States over the past decade, as reported by *National Park Service*. These geotagged Flickr photos were collected from January 1, 2010 to December 31,

[2] https://www.flickr.com/services/api/

2019. Each photo is associated with its metadata including photo ID, owner ID, taken date, latitude, longitude, title, and the number of views. The total numbers of the geotagged Flickr photos and unique users in the data set are summarized in Tab. 1.

Table 1 The numbers of geotagged Flickr photos and unique users retrieved for this study.

Park	Number of photos	Number of users
Acadia NP	34,933	1,879
Yosemite NP	50,384	3,653

3.2 Trip Reconstruction

3.2.1 Identifying attractions using HDBSCAN

Spatial clustering is widely applied to point pattern analysis such as hot spot detection. One of the most popular clustering methods is DBSCAN (Ester et al., 1996), which is a density-based clustering algorithm. It requires two parameters: search radius (ϵ) and minimum number of points (*minPts*) within the search radius. Despite its broad applications, it is difficult to determine the ϵ in the original DBSCAN algorithm due to varying density distributions of points. In this paper, we adopt HDBSCAN (Campello et al., 2013; McInnes et al., 2017), a hierarchical density-based clustering algorithm, which addresses the aforementioned issue by using flexible ϵ values to identify attractions from the geotagged Flickr photos.

To identify a proper value for *MinClusterSize*, we compare different clustering results with the geographic distribution of top attractions listed on *TripAdvisor*[3] in the two national parks. Based on this comparison, *MinClusterSize* is set to 1% of the total number of photos, which is 349 and 504 for Acadia and Yosemite National Parks, respectively.

After applying the HDBSCAN algorithm, 13 clusters are extracted from Acadia National Park and 21 clusters from Yosemite National Park. We calculate the centroids of the clusters and label each cluster with the nearest attraction listed on *TripAdvisor* or *Google Maps* to its centroid coordinates. Fig. 1 shows the distribution of geotagged photos, clustering results, as well as the identified attraction names of Acadia and Yosemite National Parks.

[3] https://www.tripadvisor.com/

3.2.2 Extracting trip sequences from geotagged photos

With attractions being identified, each photo in the cluster is labeled with an attraction name (or cluster ID). To extract trip sequences, we first group all photos by their owner ID and then sort them by the date taken. We consider a trip as a temporally-ordered sequence of photographed locations taken by the same user. Given the possibility that one user could make several trips to the area over the years, we set a time threshold λ_t to distinguish these trips. If the time difference between two consecutive photos from the same user is larger than λ_t, we separate them into two different trips. Here, we set λ_t to 4 days, which is the average length of a stay in both national parks according to *National Park Statistics*.[4,5]

Thus, if a user took a photo at attraction A, attraction B, and then attraction C within the λ_t constraint, we are able to capture this trip sequence as [A, B, C] based on the timestamp of each geotagged Flickr photo. For our data set, 1,949 trip sequences were extracted from the clustered geotagged photos in Acadia National Park, and 3,426 trip sequences from Yosemite National Park.

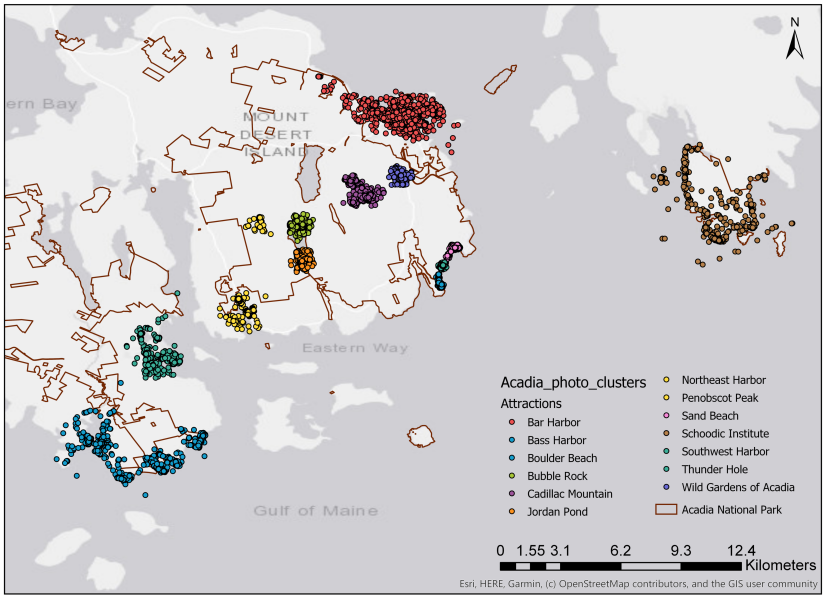
3.2.3 Calculating visiting probabilities from trip sequences

With the trip sequences extracted, we are able to construct a flow matrix based on the trip segments from all trip sequences. For example, [A, B] and [B, C] are two trip segments from the trip sequence [A, B, C]. The visiting probability is calculated proportional to the total number of outgoing trips for each attraction in the flow matrix. A monthly visiting probability matrix is also calculated in order to capture temporal factors in later computations. Fig. 2 visualizes the overall trip flows in the two national parks using flowmap.blue.[6] Further details of each attraction are provided in Tab. 8 and Tab. 9.

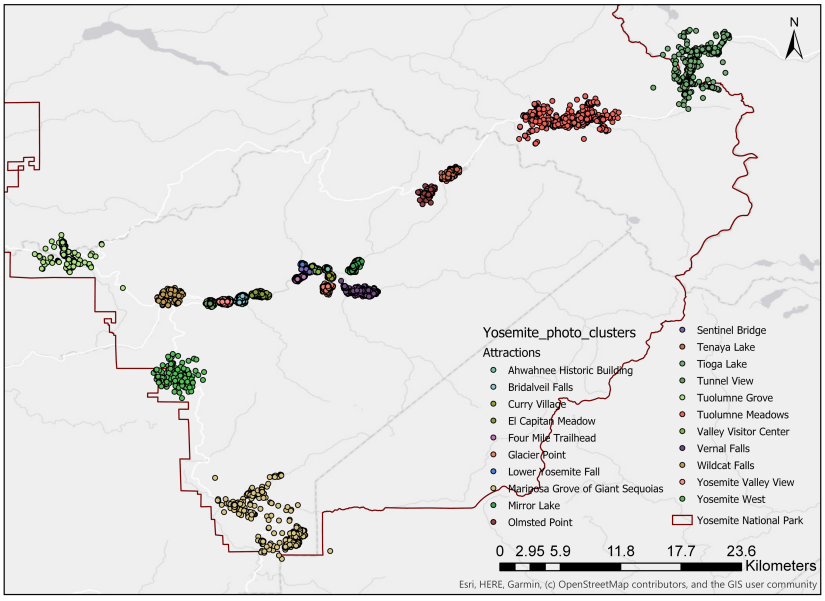
[4] https://www.nps.gov/acad/planyourvisit/faqs.htm

[5] https://www.nps.gov/yose/learn/management/statistics.htm

[6] https://flowmap.blue/

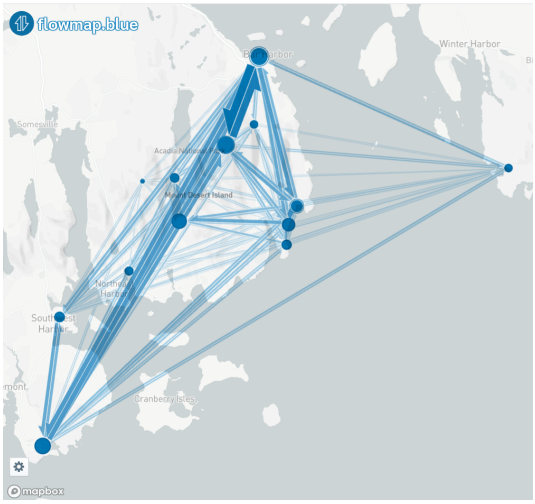


(a) Acadia National Park

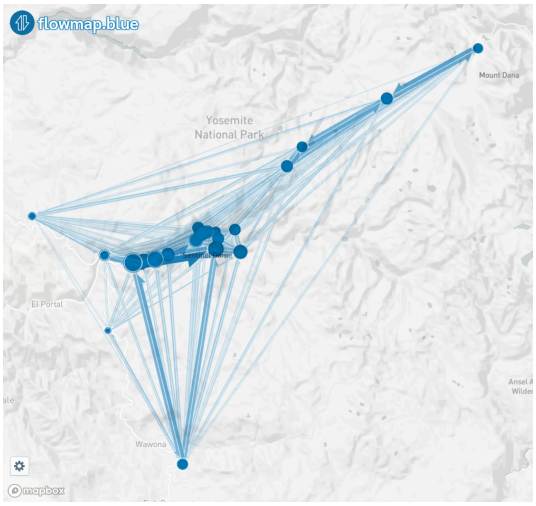


(b) Yosemite National Park

Figure 1. Photo clusters detected by HDBSCAN in the two national parks.



(a) Acadia National Park



(b) Yosemite National Park

Figure 2. Flow map visualization of trips in the two national parks. Attractions are represented as nodes. The size of nodes is determined by the total number of incoming and outgoing trips. The width of edges is determined by the number of trips.

4 A Socially Aware Huff Model

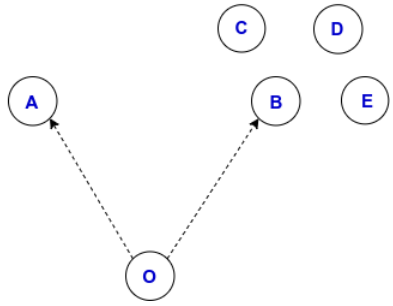
In this paper, we leverage the Huff model (Huff, 1964) with multi-destination travel behaviors being taken into account (Stouffer, 1940; Um and Crompton, 1990). Fig. 3 is used to illustrate the neighboring effect in a multi-destination trip. In Fig. 3(a), a tourist at Origin O has two destination choices A and B, with equal distance to origin O. In this case, destination B should be preferred since it has more future choices in its neighborhood compared with destination A. Furthermore, Fig. 3(b) illustrates the effect of attractiveness. When destination A and B have the same *number* of future choices in their neighborhood and the same distance to Origin O, then intuitively the destination with more *attractive* future choices in its neighborhood would be preferred. Orpana and Lampinen (2003) used the term "store centralities" to model the effect of its neighboring outlets on a store's utility. The term can be interpreted as the possibility of interaction between a store and its neighbors (Hansen, 1959). The "centrality" concept is applied to model tourist destination choice as well, with the assumption that people tend to travel to places with more attractive future choices.

4.1 The Original Huff Model

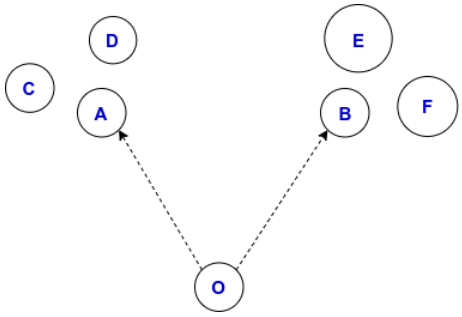
The original Huff Model (Huff, 1964) is designed to estimate the probability of customers at each origin patronizing a given store among all stores as their destination choices. It takes two factors into account: attractiveness and distance. Attractiveness can be computed as a function of many attributes of a store, including the store size, number of parking spaces, customer reviews, etc. The classic form of the Huff model can be expressed as:

$$P_{ij} = \frac{A_j^{\alpha} D_{ij}^{\beta}}{\sum_{j=1}^{n} A_j^{\alpha} D_{ij}^{\beta}} \tag{1}$$

where P_{ij} represents the probability of a customer at location i visiting store j; A_j is the measure of attractiveness of store j; D_{ij} is the distance between location i and store j; and n indicates the total number of stores in the data set. The parameters α and β ($\alpha > 0$, $\beta < 0$) are associated with the attractiveness and distance factors, respectively.



(a) Different number of future choices for destination A and B.



(b) Same number of future choices for destination A and B. Circle size represents the measure of attractiveness.

Figure 3. Diagram of future choices in multi-destination travel behavior.

4.2 Socially Aware Huff Model

In this paper, we propose a socially aware Huff model to include social factor and neighboring effect, based on the assumptions that: (1) People tend to choose more attractive travel destinations; (2) People tend to choose closer travel destinations; (3) People tend to choose travel destinations with more beneficial future choices. Based on the original Huff model shown in Eq. 1, the socially aware Huff model can be expressed as:

$$P_{ijt} = \frac{A_{jt}^{\alpha} D_{ij}^{\beta} C_{jt}^{\theta}}{\sum_{j=1}^{n} A_{j}^{\alpha} D_{ij}^{\beta} C_{jt}^{\theta}} \tag{2}$$

where P_{ijt} represents the probability of a tourist at location i visiting attraction j at time t; A_{jt} is the attractiveness of attraction j at time t; D_{ij} is the distance between origin i and attraction j; C_{jt} is the term used to describe the neighboring effect of attraction j, relative to other attractions at time t; and n indicates the total number of attractions in the area. The parameters α, β and θ are associated with the attractiveness, distance, and neighboring effect factors, respectively.

In the following, we explain how we quantify the three terms, i.e., A_{jt}, D_{jt}, and C_{jt}, mathematically. Previous research has shown that the number of geotagged photos and the number of unique users can be used to represent the attractiveness of a place (Kádár and Gede, 2013; Leung et al., 2017). Here, we include three proxies to estimate the attractiveness A_{jt} for later comparison. Log transformation is performed to address a right-skewed distribution of values. The three types of attractiveness $A_{jt}^{(l)}$, $l = 1, 2, 3$, can be expressed as:

$$A_{jt}^{(1)} = \log(M_{jt} + 1) \tag{3}$$

where M_{jt} is the number of photos at attraction j at time t.

$$A_{jt}^{(2)} = \log(U_{jt} + 1) \tag{4}$$

where U_{jt} is the number of unique users at attraction j at time t.

$$A_{jt}^{(3)} = \log(M_{jt} \times \frac{1}{U_{jt}} \sum_{k=1}^{M_{jt}} V_{kjt} + 1) \tag{5}$$

where V_{kjt} is the number of views for photo k at attraction j at time t. We use the product of the number of photos and the average number of photo views per user at attraction j to include a social influence factor. Given the fact that social media influencers (SMIs) have more followers than others, thus the photos they post would have more views and greater social impact, we include photo views per user here to account for potential existence of SMIs who upload photos at an attraction. We hypothesize that the attraction with more photo views per user is more attractive.

The term C_{jt}, measuring the neighboring effect, can be modeled as:

$$C_{jt} = \frac{\sum_{k=1}^{K} \frac{A_{kt}}{D_{kj}}}{\sum_{k=1}^{K} \frac{1}{D_{kj}}} \tag{6}$$

where K is the total number of nearest neighboring attractions being considered. C_{jt} reflects the assumption

that people tend to travel to places with more promising future choices in a multi-destination trip. We consider K-nearest neighbors of attraction j, calculating their attractiveness $A_{kt}^{(l)}$ at time period t, and weight $A_{kt}^{(l)}$ by their distance to attraction j, D_{kj}. A higher C_{jt} value is assigned to attractions with closer and more attractive neighbors. Finally, we define the term D_{ij} as the estimated driving distance using the Distance Matrix API[7] from *Google Maps*.

4.3 Calibration Method

Parameters of the Huff model need to be calibrated before further studying the travel patterns. Here, we use the linear regression calibration method - Ordinary Least Squares (OLS), which estimates one set of parameters α, β, and θ, that best fit the model based on observations. The estimation process is executed by minimizing the sum of squared residuals in a linear model. OLS calibration returns fixed values for the parameters and assumes that they are homogeneous across the study area. The general form of OLS regression can be expressed as:

$$y = \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{7}$$

where y is the dependent variable; x_i is the i^{th} independent variable; n is the number of independent variables; β_i is the regression coefficient for the i^{th} independent variable; and ϵ is the random error.

To conduct OLS, the socially aware Huff model in Eq. 2 is rewritten in a log-transformed-centered form, according to Nakanishi and Cooper (1974), in order to obtain the least square estimate of parameters:

$$\begin{aligned} \ln(P_{ijt}/\tilde{P_{it}}) &= \alpha_i \ln(A_{jt}/\tilde{A_t}) \\ &+ \beta_i \ln(D_{ij}/\tilde{D_i}) + \theta_i \ln(C_{jt}/\tilde{C_t}) \end{aligned} \tag{8}$$

where $\tilde{P_{it}}$, $\tilde{A_{jt}}$, $\tilde{D_i}$ and, $\tilde{C_t}$ are the means of P_{ijt}; A_{jt}; D_{ij} and C_{jt} over attraction j, respectively. For each origin attraction i, the model will estimate one best fit parameter set (α_i, β_i, and θ_i).

[7] https://cloud.google.com/maps-platform/routes

4.4 Software and Data Availability

Data used in this paper can be accessed with the public Flickr API.[8] The query used to access the data, code and interactive data visualization (Fig. 2) are available on GitHub.[9] The workflow underlying this paper was partially reproduced by an independent reviewer during the AGILE reproducibility review and a reproducibility report was published at https://doi.org/10.17605/OSF.IO/4CPM3.

5 Results and Discussions

5.1 Overall Calibration Results

In this section, we examine the overall calibration results for the two national parks and discuss the necessity of incorporating social factors and neighboring effects in the socially aware Huff model.

5.1.1 K-Nearest Neighbors

First we need to decide on the number of neighbors K in order to calculate the centrality C_{jt} in Eq. 6. Values of K = 2, 3 and 5 are considered. For both Acadia and Yosemite National Parks, K = 2 gives the best performance, with the lowest mean squared error (MSE) and highest R^2. More details are shown in Tab. 7. The following calibrations are subsequently all computed with K = 2 as the number of nearest neighbors in the centrality term C_{jt}.

5.1.2 Social Influence

In Tab. 2, we show the calibration results using different measurements of the attractiveness factor, $A_{jt}^{(l)}$, expressed in Eq. 3, Eq. 4 and Eq. 5. Based on R^2 and Akaike information criterion (AIC), we observe that for both national parks, the attractiveness $A_{jt}^{(3)}$ performs the best (highest R^2 and lowest AIC values), compared with the other two measurements (i.e., the number of photos and the number of unique users).

[8] https://www.flickr.com/services/api/

[9] https://github.com/meilinshi/Socially-aware-Huff-model

Table 2 OLS regression results for different measurements of attractiveness

Park	Measurement of Attractiveness	R^2	AIC	ΔAIC_i	w_i
Acadia NP	$A_{jt}^{(1)}$	0.743	724.6	14.9	5.810×10^{-4}
	$A_{jt}^{(2)}$	0.741	728.1	18.4	1.010×10^{-4}
	$A_{jt}^{(3)}$	**0.753**	**709.7**	**0**	**0.9993**
Yosemite NP	$A_{jt}^{(1)}$	0.715	2401.7	25.4	3.050×10^{-6}
	$A_{jt}^{(2)}$	0.717	2393.0	16.7	2.363×10^{-4}
	$A_{jt}^{(3)}$	**0.721**	**2376.3**	**0**	**0.9998**

ΔAIC_i is a measure of each model i to the model with the minimum AIC.
Akaike weights $w_i = exp(-0.5 \times \Delta AIC_i)/\sum_{r=1}^{N} exp(-0.5 \times \Delta AIC_r)$

Table 3 OLS regression results for different factors considered

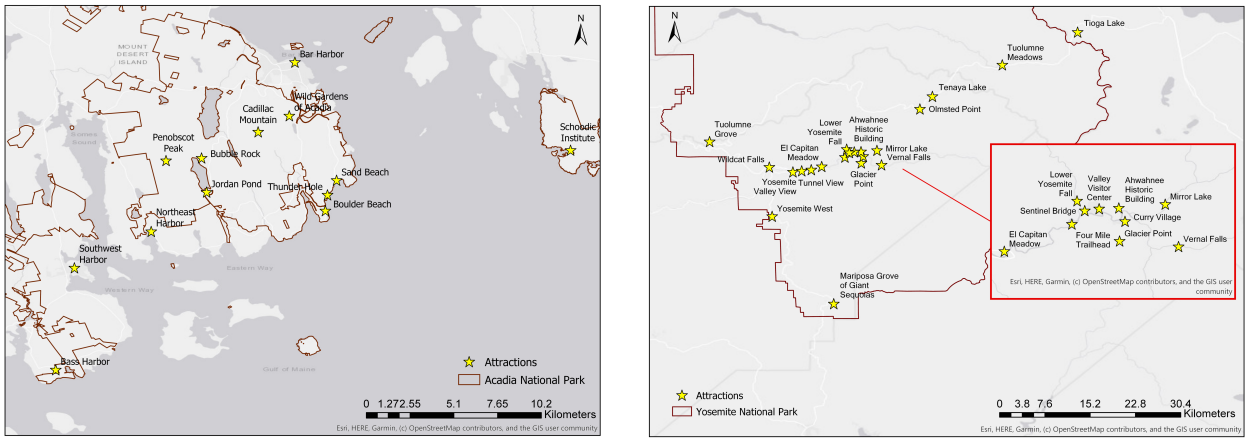
Park	Model	R^2	AIC	ΔAIC_i	w_i
Acadia NP	SA model	**0.753**	**709.7**	**0**	**0.9859**
	SA model w/o N	0.746	718.2	8.5	0.0141
	SA model w/o T	0.744	748.5	38.8	3.703×10^{-9}
	Huff model	0.738	755.8	46.1	9.624×10^{-11}
Yosemite NP	SA model	**0.721**	**2376.3**	**1.3**	**0.3430**
	SA model w/o N	**0.721**	**2375.0**	**0**	**0.6570**
	SA model w/o T	0.714	2412.4	37.4	4.969×10^{-9}
	Huff model	0.714	2410.6	35.6	1.222×10^{-8}

ΔAIC_i is a measure of each model i to the model with the minimum AIC. Models with $\Delta AIC_i < 2$ can also be considered to have substantial support (Burnham and Anderson, 2002).
Akaike weights $w_i = exp(-0.5 \times \Delta AIC_i)/\sum_{r=1}^{N} exp(-0.5 \times \Delta AIC_r)$

The results of Akaike weights (w_i), which can be interpreted as the probability that model i is the best model (Anderson et al., 2000), also show the same conclusion. Hence, we select $A_{jt}^{(3)}$ to estimate the attractiveness of an attraction, where the combination of photo views, the total number of photos, as well as the number of users are taken into account. The results indicate that including a social factor (i.e., the more photo views and potential social impact SMIs could bring to a place) can better simulate tourist preferences.

5.1.3 Temporal and Neighboring Effect Factors

To examine the overall performance of the temporal factor and neighboring effect in the socially aware Huff model (SA model), we compare it with SA model without the neighboring effect (SA model w/o N), SA model without the temporal factor (SA model w/o T), and the original Huff model (Huff model), whose results are shown in Tab. 3. The proposed SA model that includes both temporal factor and neighboring effect has the highest R^2 and lowest AIC values for Acadia National Park. As for Yosemite National Park, the performance of SA model and SA model w/o N are similar in terms of R^2 and AIC, while both models fit to the data better than SA model w/o T and Huff model. However, we cannot conclude that SA model w/o N is significantly better than SA model or vice versa based on ΔAIC_i and w_i values. The reason why we get similar performance for SA model and SA model w/o N may be due to the geographic distribution of attractions in Yosemite National Park (see Fig. 4). Most attractions are clustered (i.e., they have similar neighbors) at the center of the park, thus the neighboring effect may not be as significant as those of Acadia National Park. More details about this will be discussed in section 5.2. In general, the SA model w/o N performs better than the SA model w/o T, since the temporal factor provides more fine-grained data and for national parks, to include temporal variation when estimating visiting



(a) Acadia National Park

(b) Yosemite National Park

Figure 4. Geographic distribution of attractions in the two national parks.

patterns is crucial. Overall, the experimental results indicate the necessity to incorporate both social and temporal effects into the Huff model.

5.2 Regional Variability of Parameters

After examining the globally fitted parameters for the entire park, we further explore the regional variability of the parameters. The intuition underlying this experiment is that the relative impacts of attractiveness (α), distance (β), and neighboring effect (θ) can be different across regions in the park. Therefore, calibration is conducted for each attraction in the two national parks using observed visiting probabilities calculated from the trip sequence data. Each attraction is treated as an origin to estimate how visitors choose their next destination starting from this origin attraction. The OLS calibration gives one set of parameters (α, β, and θ) per origin, reflecting how attractiveness, distance and neighboring effect, respectively, contribute to the visiting probabilities. The results for attractions within Acadia and Yosemite National Parks are shown in Tab. 4 and Tab. 5. Only these significant origin attractions with more than 30 observed trips are included in the table.

In general, a large absolute estimation for α, β, or θ indicates a significant influence of attractiveness, distance, or neighboring effect to the destination choices, respectively. Tab. 4, demonstrates the parameter estimations for attractions in Acadia National Park. Here, we see a relatively small estimation of α for Cadillac Mountain, which is the top 1 traveler favorite attraction in Acadia National Park ranked by *TripAdvisor*. This means that compared with visitors at Boulder Beach, Sand Beach, and Jordan Pond etc., after visiting Cadillac Mountain, they are less interested in the attractiveness of an attraction to visit the next destination. The absolute values of β estimations are greater for Bass Harbor and Bubble Rock, which indicate that visitors tend to choose closer next destinations starting from these two attractions. A larger θ estimation indicates a higher probability of visitors choosing destinations with closer and more attractive neighbors, most likely clustered attractions, such as the Boulder Beach, Thunder Hole and Sand Beach cluster (see Fig. 4(a)).

Tab. 5 includes the parameter calibration results for attractions in Yosemite National Park. We see relatively larger estimations of α for attractions clustered at the center of the park, El Capitan Meadow, Lower Yosemite Fall, Sentinel Bridge, etc., which are shown in the red box of Fig. 4(b). The largest absolute value of β estimation for Tuolumne Grove indicates that visitors at this attraction are very likely to choose a closer next destination. Since Yosemite National Park is roughly 15 times larger than Acadia National Park in area, absolute values of β estimations are generally smaller compared with those of attractions in Acadia National Park. This indicates visitors are less sensitive to distance and are willing to travel further in

Table 4 OLS regression results for Acadia National Park

Origin Attraction	α	β	θ	MSE	R^2
Bass Harbor	0.8863*	-0.8608*	0.1155	0.273	0.731
Northeast Harbor	0.1684	-0.1137	0.4079*	0.208	0.838
Bar Harbor	1.3226***	0.2359	-0.0224	0.322	0.739
Cadillac Mountain	0.6601	-0.0218	0.1907	0.360	0.707
Bubble Rock	0.1722	-0.4898*	0.3994*	0.322	0.791
Jordan Pond	1.5456***	-0.0670	-0.0133	0.203	0.886
Boulder Beach	2.0496***	0.3590*	-0.3446	0.334	0.751
Thunder Hole	1.2109**	-0.1355	0.0232	0.301	0.782
Sand Beach	2.1061***	0.0374	-0.3318	0.397	0.742

Significance level: ***p $\leq$ 0.001; **p $\leq$ 0 .01; *p $\leq$ 0 .05.

Table 5 OLS regression results for Yosemite National Park

Origin Attraction	α	β	θ	MSE	R^2
Mariposa Grove of Giant Sequoias	1.6864***	-0.1023	-0.1060	0.433	0.743
Tioga Lake	1.4482*	0.0467	0.0381	0.770	0.615
Tuolumne Grove	0.7325*	-1.1656**	0.4270***	0.368	0.850
Tuolumne Meadows	0.8987**	-0.4552***	0.0985	0.388	0.776
Olmsted Point	0.2791	-0.0827	0.3661**	0.399	0.731
Tenaya Lake	0.9528*	-0.3699***	0.0838	0.495	0.786
Wildcat Falls	1.6585***	-0.0451	-0.0818	0.355	0.799
Mirror Lake	1.8888***	-0.0490	-0.2065	0.343	0.802
Vernal Falls	0.9077***	-0.0532	0.0236	0.294	0.666
El Capitan Meadow	1.1824***	-0.1147	0.0205	0.208	0.827
Tunnel View	0.5004	-0.2491	0.1451	0.368	0.646
Bridalveil Falls	1.3182***	0.0177	-0.1471	0.216	0.736
Yosemite Valley View	0.9467**	-0.2522**	0.0218	0.253	0.844
Glacier Point	1.3761***	0.0131	0.0563	0.500	0.703
Curry Village	1.3791***	-0.0800	-0.1009	0.281	0.781
Four Mile Trailhead	1.5207***	-0.0087	-0.1623*	0.168	0.836
Ahwahnee Historic Building	1.0631***	-0.1871*	0.0486	0.333	0.795
Valley Visitor Center	1.0149***	-0.0792	-0.0690	0.197	0.743
Lower Yosemite Fall	1.2195***	-0.0022	-0.0135	0.200	0.803
Sentinel Bridge	1.2127***	-0.1188**	-0.1572**	0.182	0.827

Significance level: ***p $\leq$ 0.001; **p $\leq$ 0 .01; *p $\leq$ 0 .05.

Yosemite National Park. The estimation of parameter θ is greater for dispersed attractions like Olmsted Point and Tuolumne Grove, as can be seen in Fig. 4(b). This means visitors at these two attractions are more attracted to clustered attractions (i.e. attractions with closer and more attractive neighbors), most likely the Tunnel View and Glacier Point clusters as shown in the map. In Tab. 5, we also see significant negative θ values, which reveal that visitors tend to travel to less clustered attractions (i.e., attractions with further and less attractive neighbors), especially for visitors at Four Mile Trailhead, Bridalveil Falls, Sentinel Bridge, etc., that are already in a clustered region. For origin attractions with θ closer to 0, it means that neighboring effect is not an important factor for visitors to choose their next destination at these places.

5.3 Temporal Variability of Parameters

To further explore the temporal variability of the model parameters, we divide the trips in Yosemite National

Table 6 OLS regression results for Yosemite National Park Summer vs. Non-Summer months

Origin Attraction	Time of the year	α	β	θ	R^2
Wildcat Falls	Summer	1.7046*	0.0519	-0.1768	0.744
	Non-summer	1.4516*	-0.1670	0.0945	0.859
Vernal Falls	Summer	0.8962**	-0.1116*	-0.0355	0.679
	Non-summer	0.9788***	0.0379	0.0978	0.693
El Capitan Meadow	Summer	1.2480***	-0.0869	0.0372	0.918
	Non-summer	1.0437**	-0.2023	0.0010	0.768
Tunnel View	Summer	0.4185	-0.4669*	0.1651	0.756
	Non-summer	0.7634	-0.0111	0.1197	0.607
Bridalveil Falls	Summer	1.4542***	0.1098	-0.1979	0.729
	Non-summer	0.9029	-0.1969	-0.0592	0.755
Curry Village	Summer	1.3597***	-0.0117	-0.0348	0.800
	Non-summer	1.2933***	-0.2013**	-0.1963*	0.800
Valley Visitor Center	Summer	0.8253***	-0.0782	-0.0404	0.705
	Non-summer	1.1169***	-0.0841	-0.0643	0.784
Lower Yosemite Fall	Summer	1.1861***	0.0315	0.0102	0.841
	Non-summer	1.2295***	-0.0319	-0.0319	0.787
Sentinel Bridge	Summer	0.7097*	-0.2088**	-0.0307	0.779
	Non-summer	1.4738***	-0.0374	-0.1927**	0.875

Significance level: ***p $\leq$ 0.001; **p $\leq$ 0 .01; *p $\leq$ 0 .05. Summer months include May, June, July, August, and September.

Park to summer and non-summer based on the park travel recommendation.[10] A couple of attractions in the park are seasonal, with many roads and trails being closed due to snow in winter. For example, Tuolumne Meadows typically opens from late May or June to November and Glacier Point typically opens from May to November. According to most travel guides, the best time to visit Yosemite is May to September. Hence, we use this time range to represent summer months here, and the rest as non-summer months. In Tab. 6, only origin attractions with more than 30 observed trips during both time periods are included.

Based on Tab. 6, we observe that the α estimations mostly stay the same for different times of the year. However, visitors at Bridalveil Falls are more sensitive to the attractiveness factor in summer months, while visitors at Sentinel Bridge are more interested in the attractiveness factor in non-summer months. Attractions like Vernal Falls, Tunnel View, Sentinel Bridge, etc., show larger absolute values of β estimations in summer. This means visitors at these attractions are attracted to closer attractions during the summer months, and further attractions for non-summer months, which is potentially due to closures of closer attractions in non-summer months. Meanwhile, we see also larger absolute θ estimations for Curry Village in non-summer months and Bridalveil Falls in summer months. A positive θ estimation means visitors are more attracted to clustered attractions (i.e., attractions with closer and more attractive neighbors) and negative value means the opposite.

[10] https://www.nps.gov/yose/planyourvisit/traffic.htm

6 Conclusions and Future Work

In this work, we explore the visiting probabilities of attractions within two national parks using a socially aware Huff model, in which both social factors and neighboring effects are taken into account. To calibrate the model parameters, we use the observed trip sequences extracted from the geotagged Flickr photos within Acadia and Yosemite National Parks from the year 2010 to 2019. For the social factor, we have shown that incorporating the number of photo views when evaluating the attractiveness of a place achieves a better result than simply using the number of photos or number of users alone. The calibration results also demonstrate that the socially aware Huff model considering a temporal factor in place attractiveness and neighboring effects is more accurate than the original

Huff model in predicting visiting probabilities of attractions within both national parks.

We further explore the visiting patterns of each attraction within the two national parks based on model parameters of attractiveness, distance, and neighboring effect factors. In general, visitors in Acadia National Park are more sensitive to the distance factor and neighboring effects when choosing their next destination, while visitors in Yosemite National Park are more sensitive to the attractiveness factor. We have also shown that there is a regional and temporal variability of the model parameters.

This work is based on our three assumptions that: (1) People tend to choose more attractive travel destinations; (2) People tend to choose closer travel destinations; (3) People tend to choose travel destinations with more beneficial future choices. In fact, there could be many other factors contributing to the visiting probability of a place. Taking the social factor alone as an example, traveling under social influence or taking copycat photos (Picheta, 2021) has become an emerging trend. In future work, we plan to consider more comprehensive measurements to evaluate the social impact of geotagged photos in order to better capture the travel patterns. Furthermore, this work is only studied using existing attractions within national parks. We plan to look at more than just two parks in the future, explore new methods to discover emerging travel destinations and study their interactions with the surroundings.

References

Anderson, D. R., Burnham, K. P., and Thompson, W. L.: Null hypothesis testing: problems, prevalence, and an alternative, The journal of wildlife management, pp. 912–923, 2000.

Bakr, G. and Ali, I. E. H.: The role of social networking sites in promoting Egypt as an international tourist destination, South Asian Journal of Tourism and Heritage, 6, 169–183, 2013.

Balmford, A., Beresford, J., Green, J., Naidoo, R., Walpole, M., and Manica, A.: A Global Perspective on Trends in Nature-Based Tourism, PLoS Biology, 7, e1000 144, 2009.

Burnham, K. P. and Anderson, D. R.: A practical information-theoretic approach, Model selection and multimodel inference, 2, 2002.

Campello, R. J., Moulavi, D., and Sander, J.: Density-based clustering based on hierarchical density estimates, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp. 160–172, Springer, Berlin, Heidelberg, 2013.

Djossa, C.: When not to geotag while traveling, National Geographic, 2019.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise., in: Kdd, vol. 96, pp. 226–231, 1996.

Freberg, K., Graham, K., McGaughey, K., and Freberg, L. A.: Who are the social media influencers? A study of public perceptions of personality, Public Relations Review, 37, 90–92, 2011.

Glover, P.: Celebrity endorsement in tourism advertising: Effects on destination image, Journal of Hospitality and Tourism Management, 16, 16–23, 2009.

Gong, S., Cartlidge, J., Bai, R., Yue, Y., Li, Q., and Qiu, G.: Geographical and temporal huff model calibration using taxi trajectory data, GeoInformatica, pp. 1–28, 2020.

Hansen, W. G.: How Accessibility Shapes Land Use, Journal of the American Planning Association, 25, 73–76, 1959.

Heikinheimo, V., Minin, E. D., Tenkanen, H., Hausmann, A., Erkkonen, J., and Toivonen, T.: 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, 85, 2017.

Hu, F., Li, Z., Yang, C., and Jiang, Y.: A graph-based approach to detecting tourist movement patterns using social media data, Cartography and Geographic Information Science, 46, 368–382, 2019.

Huff, D. L.: Defining and Estimating a Trading Area, Journal of Marketing, 28, 34–38, 1964.

Jalilvand, M. R. and Samiei, N.: The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB), Internet Research, 22, 591–612, 2012.

Kádár, B. and Gede, M.: Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography, Cartographica: The International Journal for Geographic Information and Geovisualization, 48, 78–88, 2013.

Kim, Y., ki Kim, C., Lee, D. K., woo Lee, H., and Andrada, R. I. T.: Quantifying nature-based tourism in protected areas in developing countries by using social big data, Tourism Management, 72, 249–256, 2019.

Leung, D., Law, R., Van Hoof, H., and Buhalis, D.: Social media in tourism and hospitality: A literature review, Journal of travel & tourism marketing, 30, 3–22, 2013.

Leung, R., Vu, H. Q., and Rong, J.: Understanding tourists' photo sharing and visit pattern at non-first tier attractions via

geotagged photos, Information Technology and Tourism, 17, 55–74, 2017.

Li, D., Zhou, X., and Wang, M.: Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities, Cities, 74, 249–258, 2018.

Li, Z.: Psychological empowerment on social media: who are the empowered users?, Public Relations Review, 42, 49–59, 2016.

Liang, Y., Gao, S., Cai, Y., Foutz, N. Z., and Wu, L.: Calibrating the dynamic Huff model for business analysis using location big data, Transactions in GIS, 24, 681–703, 2020.

Litvin, S. W., Goldsmith, R. E., and Pan, B.: Electronic word-of-mouth in hospitality and tourism management, Tourism management, 29, 458–468, 2008.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., and Woodward, J.: A context-aware personalized travel recommendation system based on geotagged social media data mining, International Journal of Geographical Information Science, 27, 662–684, 2013.

McInnes, L., Healy, J., and Astels, S.: hdbscan: Hierarchical density based clustering, The Journal of Open Source Software, 2, 205, 2017.

Misui, Y. and Kamata, H.: Where do Spa tourists come from?-An application of Huff model to Japanese spa destination, Tech. rep., University of Massachusetts Amherst, 2016.

Mou, N., Yuan, R., Yang, T., Zhang, H., Tang, J. J., and Makkonen, T.: Exploring spatio-temporal changes of city inbound tourism flow: The case of Shanghai, China, Tourism Management, 76, 103 955, 2020.

Nahuelhual, L., Carmona, A., Lozada, P., Jaramillo, A., and Aguayo, M.: Mapping recreation and ecotourism as a cultural ecosystem service: An application at the local level in Southern Chile, Applied Geography, 40, 71–82, 2013.

Nakanishi, M. and Cooper, L. G.: Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach, Journal of Marketing Research, 11, 303, 1974.

Nicolau, J. L.: Characterizing Tourist Sensitivity to Distance, Journal of Travel Research, 47, 43–52, 2008.

Nicolau, J. L. and Más, F. J.: Sequential choice behavior: Going on vacation and type of destination, Tourism Management, 29, 1023–1034, 2008.

Orpana, T. and Lampinen, J.: Building Spatial Choice Models from Aggregate Data, Journal of Regional Science, 43, 319–348, 2003.

Orsi, F. and Geneletti, D.: Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas, Journal for Nature Conservation, 21, 359–368, 2013.

Parsons, H.: Does social media influence an individual's decision to visit tourist destinations? Using a case study of Instagram., Ph.D. thesis, Cardiff Metropolitan University, 2017.

Picheta, R.: New Zealand tells tourists to stop copying other people's travel photos, CNN, 2021.

Puustinen, J., Pouta, E., Marjo Neuvonen, and Tuija Sievänen: Visits to national parks and the provision of natural and man-made recreation and tourism resources, Journal of Ecotourism, 8, 18–31, 2009.

Stouffer, S. A.: Intervening Opportunities: A Theory Relating Mobility and Distance, American Sociological Review, 5, 845, 1940.

Tasse, D., Liu, Z., Sciuto, A., and Hong, J.: State of the geotags: Motivations and recent changes, in: Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, 2017.

Tham, A., Mair, J., and Croy, G.: Social media influence on tourists' destination choice: importance of context, Tourism Recreation Research, 45, 161–175, 2020.

Um, S. and Crompton, J. L.: Attitude determinants in tourism destination choice, Annals of Tourism Research, 17, 432–448, 1990.

Wood, S. A., Guerry, A. D., Silver, J. M., and Lacayo, M.: Using social media to quantify nature-based tourism and recreation, Scientific reports, 3, 1–7, 2013.

Wu, L., Zhang, J., and Fujiwara, A.: A Tourist's Multi-Destination Choice Model with Future Dependency, Asia Pacific Journal of Tourism Research, 17, 121–132, 2012.

Yang, Y., Fik, T., and Zhang, J.: Modeling sequential tourist flows: Where is the next destination?, Annals of Tourism Research, 43, 297–320, 2013.

Zheng, Y.-T., Zha, Z.-J., and Chua, T.-S.: Mining travel patterns from geotagged photos, ACM Trans. Intell. Syst. Technol, 3, 2012.

Appendix

Regression results with different K values selected for K-NN

Attractions Summary

Table 7 OLS regression results for different K values selected for K-Nearest Neighbors

Park	Time of the year	K	MSE	R^2
Acadia National Park	All time	2	**0.351958**	**0.753**
		3	0.355672	0.750
		5	0.360020	0.747
	Summer months	2	**0.239118**	**0.780**
		3	0.241003	0.778
		5	0.241888	0.777
	Non-summer months	2	**0.470965**	**0.766**
		3	0.479230	0.762
		5	0.490113	0.757
Yosemite National Park	All time	2	**0.373372**	**0.721**
		3	0.373495	0.721
		5	0.373465	0.721
	Summer months	2	**0.310437**	**0.714**
		3	0.310878	0.714
		5	0.311528	0.713
	Non-summer months	2	**0.430608**	**0.735**
		3	0.430768	0.734
		5	0.431058	0.734

Summer months include May, June, July, August, and September for both parks.

Table 8 Summary of attractions in Acadia National Park

Attraction	Number of photos	Outgoing trips	Incoming trips
Schoodic Institute	1119	53	64
Bass Harbor	2298	260	288
Southwest Harbor	723	109	111
Northeast Harbor	605	67	76
Bar Harbor	6259	433	357
Wild Gardens of Acadia	550	60	66
Cadillac Mountain	3285	349	345
Penobscot Peak	776	16	15
Bubble Rock	703	83	89
Jordan Pond	1250	227	250
Boulder Beach	536	85	102
Thunder Hole	977	167	185
Sand Beach	1253	216	177

Table 9 Summary of attractions in Yosemite National Park

Attraction	Number of photos	Outgoing trips	Incoming trips
Mariposa Grove of Giant Sequoias	1787	135	135
Tioga Lake	1054	111	111
Tuolumne Grove	555	65	53
Tuolumne Meadows	1630	151	165
Yosemite West	674	35	31
Olmsted Point	890	168	165
Tenaya Lake	626	123	128
Wildcat Falls	724	147	110
Mirror Lake	875	134	150
Vernal Falls	2349	205	229
El Capitan Meadow	1010	175	197
Tunnel View	1987	489	414
Bridalveil Falls	1835	366	332
Yosemite Valley View	1469	231	249
Glacier Point	2165	250	288
Curry Village	855	140	157
Four Mile Trailhead	1269	246	225
Ahwahnee Historic Building	560	91	107
Valley Visitor Center	1788	201	197
Lower Yosemite Fall	869	164	167
Sentinel Bridge	1194	267	284