University of California
Santa Barbara

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

A Thesis submitted in partial satisfaction
of the requirements for the degree

Master of Arts
in
Geography

by

Meilin Shi

Committee in charge:

Professor Krzysztof Janowicz, Chair
Professor Konstadinos Goulias
Professor Keith Clarke

June 2021

The Thesis of Meilin Shi is approved.

Professor Konstadinos Goulias

Professor Keith Clarke

Professor Krzysztof Janowicz, Committee Chair

June 2021

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

by

Meilin Shi

Acknowledgements

I would like to thank the members of my committee, Professor Konstadinos Goulias and Professor Keith Clarke, for their guidance throughout the work and reviews of this thesis. I would like to thank my advisor, Professor Krzysztof Janowicz, for his distinct insights and valuable input in this work, as well as the help and advice he has provided since my undergraduate years. I would like to thank Ling Cai and all other STKO Lab members who have offered me numerous feedback and suggestions on my thesis topic and helped me overcome the obstacles encountered in the research process. I am very grateful to my colleagues and friends in the Geography Department for their support and friendship, especially those I met when I was still an undergraduate at UCSB. A special thanks to my friends for their remote companionship in the COVID-19 era. Finally and most importantly, I would like to thank my parents for their unconditional love, support, and encouragement.

Abstract

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

by

Meilin Shi

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 work, 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 three 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.

Contents

Chapter 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 [1], especially for national parks in the United States, according to the visitation statistics by the *National Park Service*.[1] Identifying and evaluating the relevant determinants of tourist flows is important. This information can help with the planning and management of the national parks (i.e., infrastructure and services), 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 [2, 3]. The emergence of location-based social networks (LBSNs) and volunteered geographic information (VGI), such as Flickr, Instagram, Facebook etc., together with the 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 the travel decision-making process nowadays [4]. For example, places like Horseshoe Bend in Arizona, Devil's Bath-

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

tub in Virginia, Kanarraville Falls in Utah, 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 [5]. More specifically, geotagged photos posted by social media influencers (SMIs) can rapidly attract new visitors [6]. These influencers are usually users with a large number of followers and have established credibility in certain fields that can shape the attitudes of tourists and thus influence their travel preferences [7, 8]. 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 [9].

Moreover, existing work has shown that fine-grained spatio-temporal data collected from social media can be used to quantify visitation rates [10], to estimate visitor flows [11, 12], and to detect popular sub-regions and temporal activity patterns [13]. 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 examine the process of a hidden gem place like Horseshoe Bend getting more and more popular. As it is surrounded by many attractions (Glen Canyon, Antelope Canyon, Grand Canyon, etc.), we can estimate the interactions between Horseshoe Bend and its nearby attractions, 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 [14], which was originally used to calculate the probability of customers shopping at each retail store and capture the visiting behavior of customers. 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 three 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 thesis is organized as follows. Chapter 2 introduces related work on tourism, geo-social media, and the Huff model. Chapter 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 Chapter 4 together with the model calibration method. In Chapter 5, we present the model calibration results and explain the spatial and temporal variability of the parameters used in the socially aware Huff model. Finally, we summarize our findings and discuss future directions in Chapter 6.

Chapter 2

Related Work

This chapter reviews the existing research on the involvement of location-based social networks (LBSNs) and volunteered geographic information (VGI) in tourism, as well as the variations of the Huff model with their applications in destination choice and travel pattern analysis. The involvement of geo-social media in tourism research is further categorized into two aspects: the use (i.e., the adoption of social media data for analysis purposes) and the role (i.e., the impact of social media in travel decision-making process) of social media in tourism.

2.1 Tourism and Geo-social Media

2.1.1 The Use of Social Media

The adoption of social media data in tourism research has an increasing popularity, because of its convenience in the collection process and broad spatial-temporal coverage, which is superior than the traditional survey data and travel logs [15]. Heikinheimo et al. [13] proved that geotagged Instagram posts can be used to detect popular sub-regions, visitor activities and their temporal patterns in Pallas-Yllästunturi National

Park, Finland, providing comparable information to that collected by surveys. Zheng et al. [16] used Flickr data to discover regions of attractions (RoAs), explored tourists movement patterns in relation to the RoAs, and investigated topological characteristics of travel routes by different tourists. Hu et al. [17] extracted popular attractions and tour routes using a graph-based network in New York City from Twitter data. Work by Ji et al. [18] investigated the extraction of landmarks for city scene summarization with geotagged Flickr photos. Majid et al. [19] proposed a context-aware personalized travel recommendation system and evaluated it based on a Flickr data set. Li et al. [20] used Flickr data to compare the spatial overlap of destinations between tourists and locals in ten US cities. Similarly, Maeda et al. [21] examined the preference of destinations between domestic tourists and foreign tourists in Japan using Twitter and Foursquare data. There are also many other applications of geotagged social media data, such as inbound tourism flows [22, 23, 24] and sentiment analysis in tourism [25, 26, 27].

2.1.2 The Role of Social Media

In the past decade, social media has evolved into an important player in tourism advertising and promotion [28, 29]. Litvin et al. [30] showed that travelers are increasingly influenced by electronic Word-of-Mouth (eWOM) from social media. Parsons [31] echoed the similar idea that Instagram influences tourist decision-making process, especially for younger generations. Meanwhile, Pop et al. [32] evaluated the credibility of the content generated by social media influencers (SMIs) and indicated that consumer trust in SMIs has a positive effect on travel decision-making. Jalilvand and Samiei [33] examined the influence of eWOM and showed that it has a significant impact on tourist attitudes towards visiting Isfahan, Iran. Likewise, Shafiee et al. [34] investigated the effect of destination image on tourism satisfaction and showed that positive image of the destination

on social media can lead to the intention of revisiting. Whereas, Hernandez-Mendez et al. [35] argued that eWOM has a biased target population and limited influence on tourist decision-making behavior. Tham et al. [36] also revealed that the role of social media appears to have only moderate-low influence on destination choice through interviews conducted with tourist decision-makers in Australia. 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 preferences of tourists.

2.2 Huff Model

There have been many research efforts towards tourist destination choice and sequential tourist flows [37, 38, 39]. The Huff model [14] is one of the options, though it was originally developed to predict retail sales and consumer behavior. The Huff model estimates the probability of a consumer patronizing retail stores based on two factors: attractiveness of a store and the travel cost, which can also be applied to tourism research. Misui and Kamata [40] adopted the Huff model to show the effect of travel time on visiting probability to spa destinations in Japan. Similarly, Nicolau [41] studied tourist sensitivities to distance and price for destination choice in Spain using a national tourist choice behavior survey data. In addition to attractiveness and travel cost, there are more factors that can influence tourist choice of destination. For example, as tourists have a greater tendency to take multi-destination tours to maximize the benefits [42], Yang et al. [39] applied 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. [43] included weekday and weekend variations when calculating visiting probability of shopping areas using taxi trajectory data in Shenzhen and New York. Liang et al. [44] proposed a T-Huff model to include temporal dynamics of human mobility patterns and proved that it outperforms the original static Huff model when estimating temporal store visits using SafeGraph POI visits data. Likewise, on the basis of previous work, we consider multi-destination travel behavior and include the temporal factor in our study, given the availability of social media data.

Chapter 3

Data and Trip Reconstruction

In this chapter, we introduce the data set used for the study in and elaborate on how we reconstruct trips and calculate visiting probabilities from the geotagged photos.

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.[1] Three national parks - Acadia National Park, Yosemite National Park and Yellowstone National Park - have been selected from the top ten most visited national parks in the United States over the past decade, as reported by the *National Park Service*. These geotagged Flickr photos were collected from January 1, 2010 to December 31, 2019. Each photo is associated with its metadata including photo ID, owner ID, date taken, 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 Table 3.1.

[1] `https://www.flickr.com/services/api/`

Table 3.1: The numbers of geotagged Flickr photos and unique users retrieved for this study.

Park	Number of photos	Number of users
Acadia National Park	34,933	1,879
Yosemite National Park	50,384	3,653
Yellowstone National Park	67,896	2,599

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 [45], 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 work, we adopt HDBSCAN [46, 47], 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*[2] in the three national parks. Based on this comparison, *MinClusterSize* is set to 1% of the total number of photos, which is 349, 504 and 679 for Acadia, Yosemite and Yellowstone National Parks, respectively.

After applying the HDBSCAN algorithm, 13 clusters are extracted from Acadia National Park and 21 clusters from both Yosemite and Yellowstone National Parks. 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. Figure 3.1, Figure 3.2

[2] `https://www.tripadvisor.com/`

and Figure 3.3 display the distribution of geotagged photos, clustering results, as well as the identified attraction names in the three national parks.

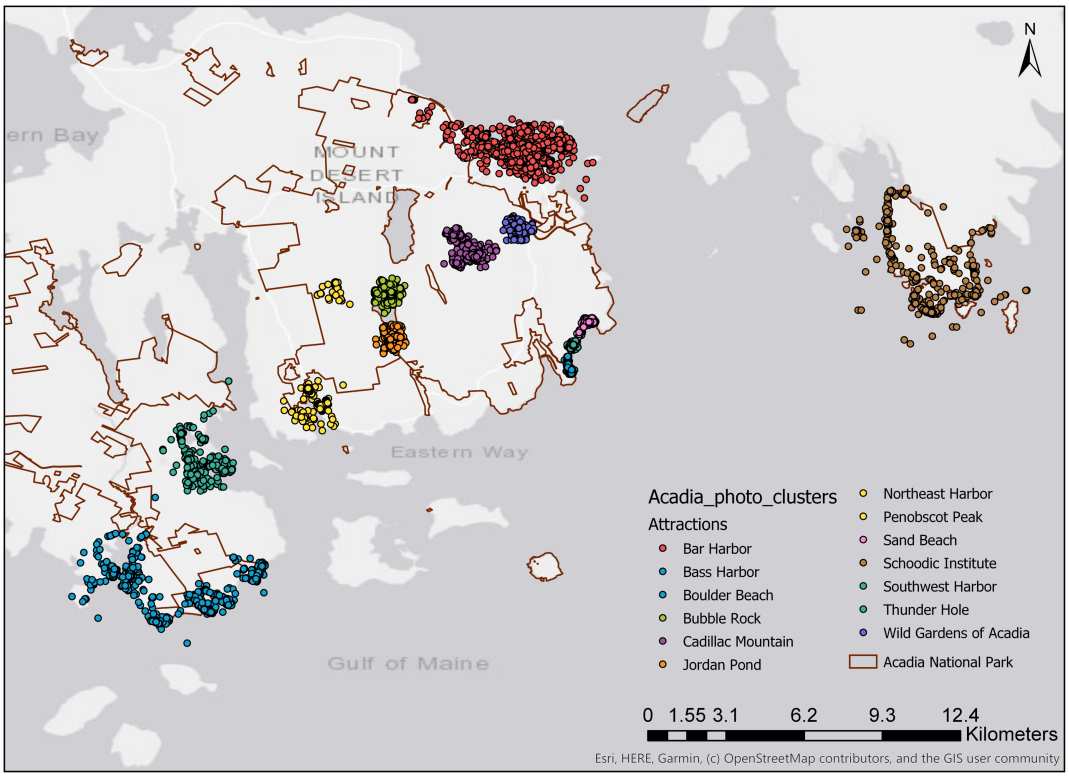


Figure 3.1: Photo clusters detected by HDBSCAN in Acadia National Park.

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

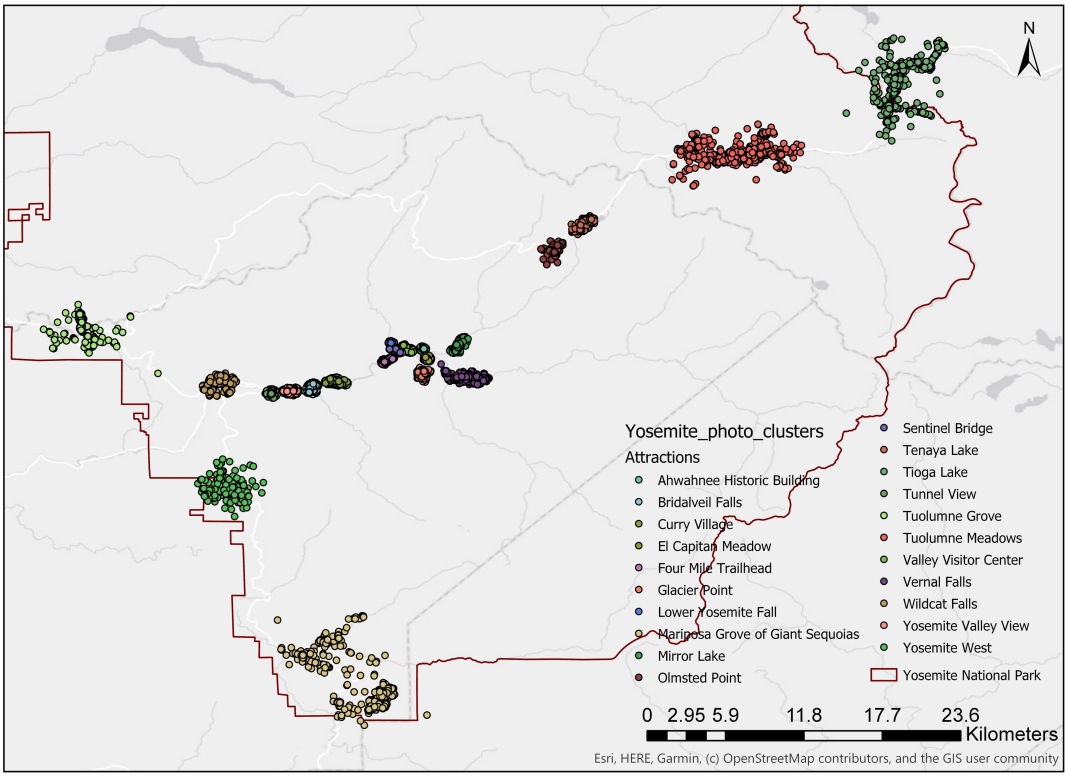


Figure 3.2: Photo clusters detected by HDBSCAN in Yosemite National Park.

λ_t to 4 days, which is the average length of stay indicated for all three national parks according to the *National Park Service*[3,4] and the Yellowstone travel guide.[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, 3,426 trip sequences from Yosemite National Park, and 2,674 trip sequences from Yellowstone National Park.

[3] `https://www.nps.gov/acad/planyourvisit/faqs.htm`
[4] `https://www.nps.gov/yose/learn/management/statistics.htm`
[5] `https://yellowstone.net/intro/top-10/`

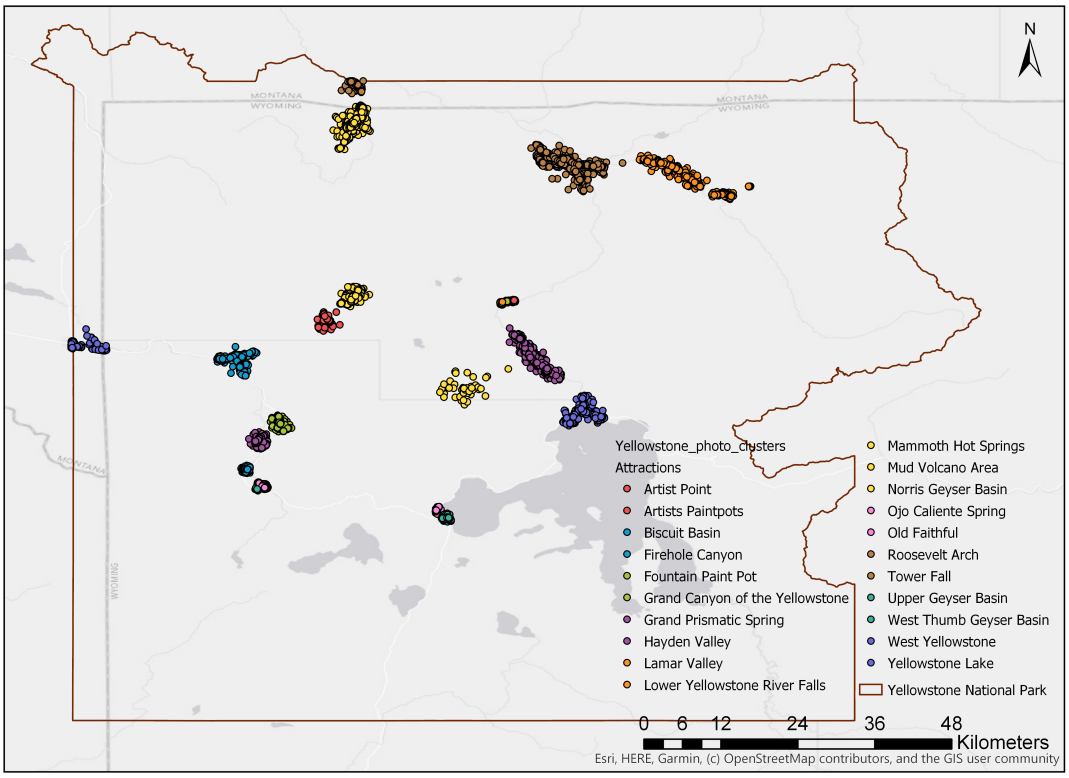
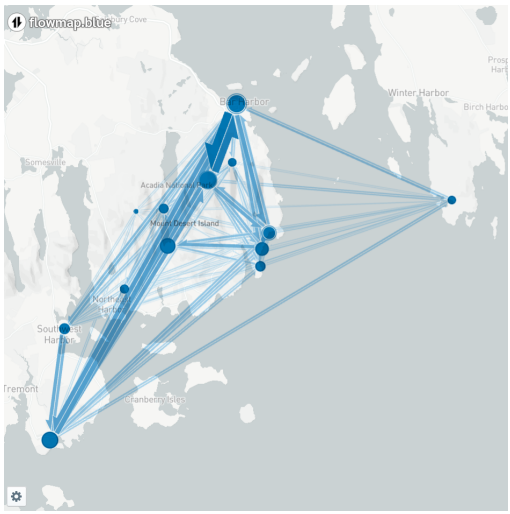


Figure 3.3: Photo clusters detected by HDBSCAN in Yellowstone National Park.

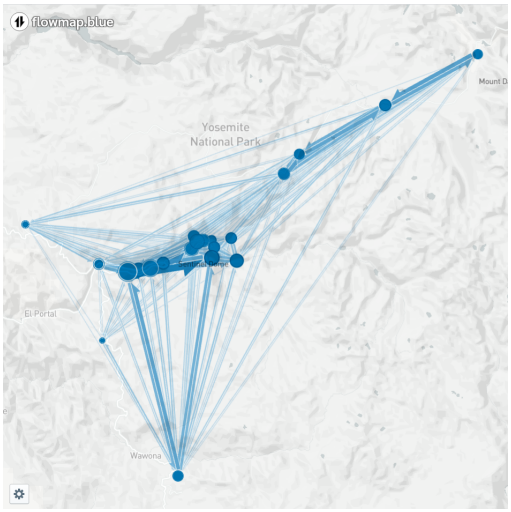
3.2.3 Calculating visiting probabilities from trip sequences

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. Figure 3.4 visualizes the overall trip flows in the three national parks using flowmap.blue.[6] Further details of each attraction are provided in Appendix A.2.

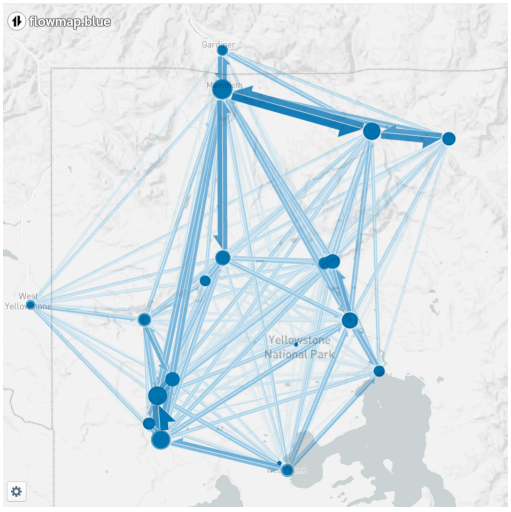
[6] `https://flowmap.blue/`



(a) Acadia National Park



(b) Yosemite National Park



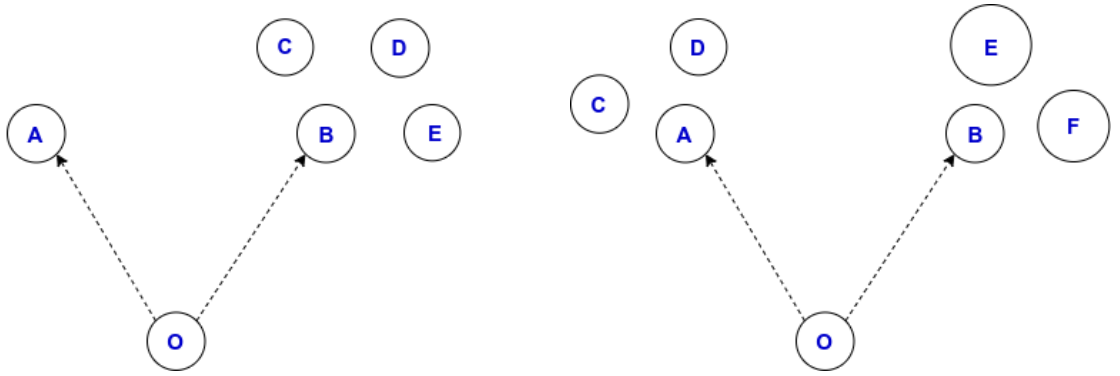
(c) Yellowstone National Park

Figure 3.4: Flow map visualization of trips in the three 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.

Chapter 4

A Socially Aware Huff Model

In this work, we leverage the Huff model [14] with multi-destination travel behavior being taken into account [48, 49]. Figure 4.1 is used to illustrate the neighboring effect in a multi-destination trip. In Figure 4.1a, 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, Figure 4.1b 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 [50] 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 [51]. We can apply the "centrality" concept to model tourist destination choice as well, with the assumption that people tend to travel to places with more attractive future choices.



(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 4.1: Diagram of future choices in multi-destination travel behavior.

4.1 The Original Huff Model

The original Huff Model [14] 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:

$$P_{ij} = \frac{A_j^{\alpha} D_{ij}^{\beta}}{\sum_{j=1}^{n} A_j^{\alpha} D_{ij}^{\beta}} \tag{4.1}$$

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.

4.2 Socially Aware Huff Model

In this work, 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 Equation 4.1, the socially aware Huff model can be expressed as:

$$P_{ijt} = \frac{A_{jt}^{\alpha} D_{ij}^{\beta} C_{jt}^{\theta}}{\sum_{j=1}^{n} A_{j}^{\alpha} D_{ij}^{\beta} C_{jt}^{\theta}} \tag{4.2}$$

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 [52, 53]. 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:

$$A_{jt}^{(1)} = \log(M_{jt} + 1) \tag{4.3}$$

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

$$A_{jt}^{(2)} = \log(U_{jt} + 1) \tag{4.4}$$

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

$$A_{jt}^{(3)} = \log(M_{jt} \times \frac{1}{U_{jt}} \sum_{k=1}^{M_{jt}} V_{kjt} + 1) \tag{4.5}$$

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:

$$C_{jt} = \frac{\sum_{k=1}^{K} \frac{A_{kt}}{D_{kj}}}{\sum_{k=1}^{K} \frac{1}{D_{kj}}} \tag{4.6}$$

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[1] from *Google Maps*.

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

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:

$$y = \sum_{i=1}^{n} \beta_i x_i + \epsilon \tag{4.7}$$

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 Equation 4.2 is rewritten in a log-transformed-centered form, according to Nakanishi and Cooper [54], in order to obtain the least square estimate of parameters:

$$\ln(P_{ijt}/\tilde{P}_{it}) = \alpha_i \ln(A_{jt}/\tilde{A}_t) + \beta_i \ln(D_{ij}/\tilde{D}_i) + \theta_i \ln(C_{jt}/\tilde{C}_t) \tag{4.8}$$

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).

Chapter 5

Results and Discussions

5.1 Overall Calibration Results

In this section, we examine the overall calibration results for the three national parks and discuss the necessity of incorporating social factors, temporal 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 Equation 4.6. Values of K = 2, 3 and 5 are considered. In most cases, K = 2 gives the best performance, with the lowest mean squared error (MSE) and highest R^2. More details with the selection of K are shown in Appendix A.1. The following calibrations are subsequently all computed with K = 2 as the number of nearest neighbors in the centrality term C_{jt}.

Table 5.1: OLS regression results for different measurements of attractiveness

Park	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**
Yellowstone NP	$A_{jt}^{(1)}$	0.681	2255.9	40.0	2.061×10^{-9}
	$A_{jt}^{(2)}$	0.688	2234.7	18.8	8.272×10^{-5}
	$A_{jt}^{(3)}$	**0.693**	**2215.9**	**0**	**0.9999**

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

5.1.2 Social Influence

In Table 5.1, we show the calibration results using different measurements of the attractiveness factor, $A_{jt}^{(l)}$, expressed in Equation 4.3, Equation 4.4 and Equation 4.5. Based on R^2 and Akaike information criterion (AIC) [56], we observe that for all three 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). The results of AIC together with the Akaike weights (w_i) [57] can be used to conclude which model is significantly better than the others and the probability that model i is the best model [58]. 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.

Table 5.2: 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}
Yellowstone NP	SA model	**0.693**	**2215.9**	**0**	**0.7311**
	SA model w/o N	0.692	2217.9	2.0	0.2689
	SA model w/o T	0.687	2258.7	42.8	3.716×10^{-10}
	Huff model	0.686	2258.8	42.9	3.535×10^{-10}

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 [55].
Akaike weights $w_i = exp(-0.5 \times \Delta AIC_i) / \sum_{r=1}^{N} exp(-0.5 \times \Delta AIC_r)$

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 the SA model without the neighboring effect (SA model w/o N), the SA model without the temporal factor (SA model w/o T), and the original Huff model (Huff model), whose results are shown in Table 5.2. 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 and Yellowstone National Park. As for Yosemite National Park, the performance of the SA model and the SA model w/o N are similar in terms of R^2 and AIC, while both models fit to the data better than the SA model w/o T and Huff model. However, we cannot conclude that the SA model w/o N is significantly better than the SA model or vice versa based on ΔAIC_i and w_i values. The reason why we get similar performance for the SA and SA models w/o N may be due to the geographic distribution of attractions in Yosemite National

Park (see Figure 5.1, Figure 5.2). 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 and Yellowstone 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 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 three national parks using observed visiting probabilities calculated from the trip sequence data. Each attraction is treated as an origin to estimate the probability of visitors choosing their next destination 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, Yosemite and Yellowstone National Parks are shown in Table 5.3, Table 5.4 and Table 5.5. Only the significant origin attractions with more than 30 observed trips are included in the table.

Table 5.3: 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 $<$ 0.001; **p $<$ 0 .01; *p $<$ 0 .05.

5.2.1 Acadia National Park

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. Table 5.3, 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. The larger θ estimations for Northeast Harbor and Bubble Rock indicate 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, shown in Figure 5.1.

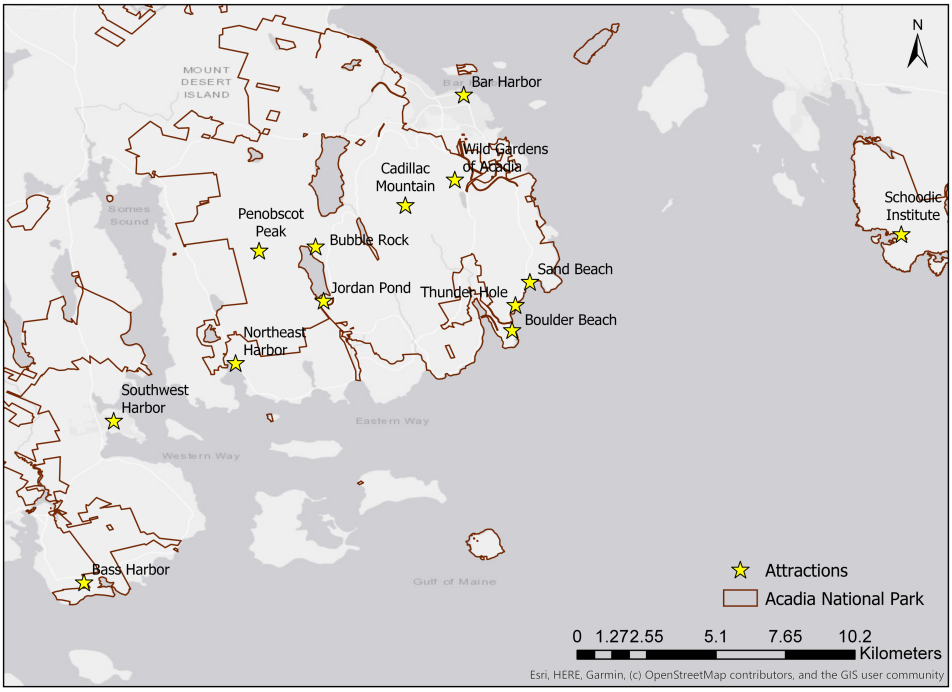


Figure 5.1: Geographic distribution of attractions in the Acadia National Parks

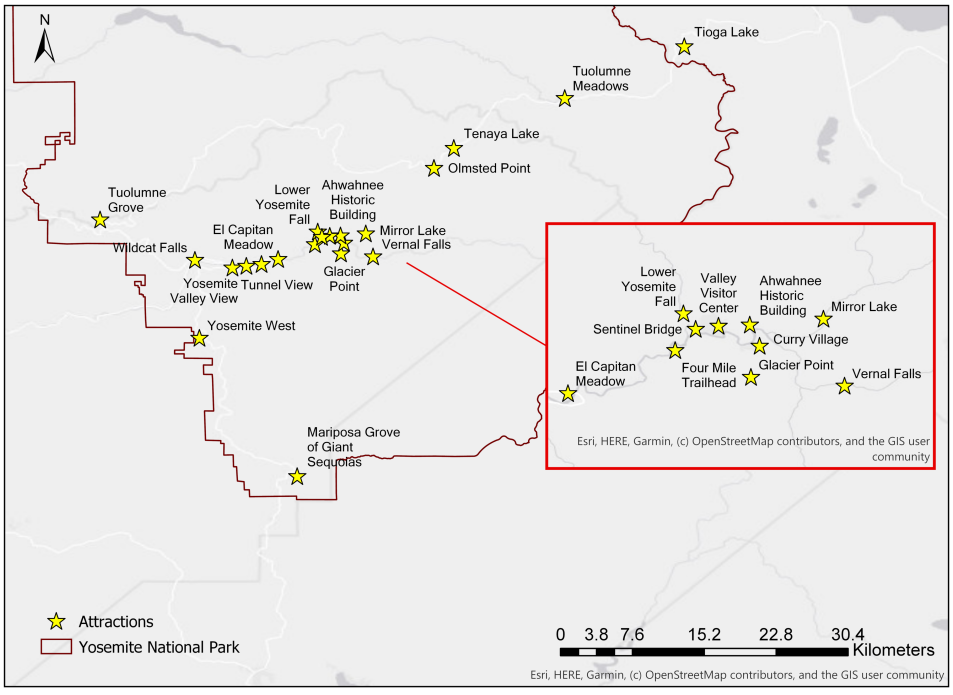


Figure 5.2: Geographic distribution of attractions in Yosemite National Park

Table 5.4: 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 $<$ 0.001; **p $<$ 0 .01; *p $<$ 0 .05.

5.2.2 Yosemite National Park

Table 5.4 includes the parameter calibration results for attractions in Yosemite National Park. We see relatively larger α estimations 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 Figure 5.2. 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, such as Wildcat Falls. Since Yosemite National Park is roughly 15 times larger than Acadia National Park in area, the absolute values of β estimations are generally smaller compared with those of attractions in Acadia National Park. This indicates

that visitors are less sensitive to distance and are willing to travel further in Yosemite National Park. The estimation of parameter θ is greater for dispersed attractions like Olmsted Point and Tuolumne Grove, as we can see in Figure 5.2. 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 shown in the map. In Table 5.4, we also see significant negative θ estimations, 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 attractions region. For origin attractions with θ estimations closer to 0, it means that neighboring effect is not an important factor for visitors to choose their next destination at these places.

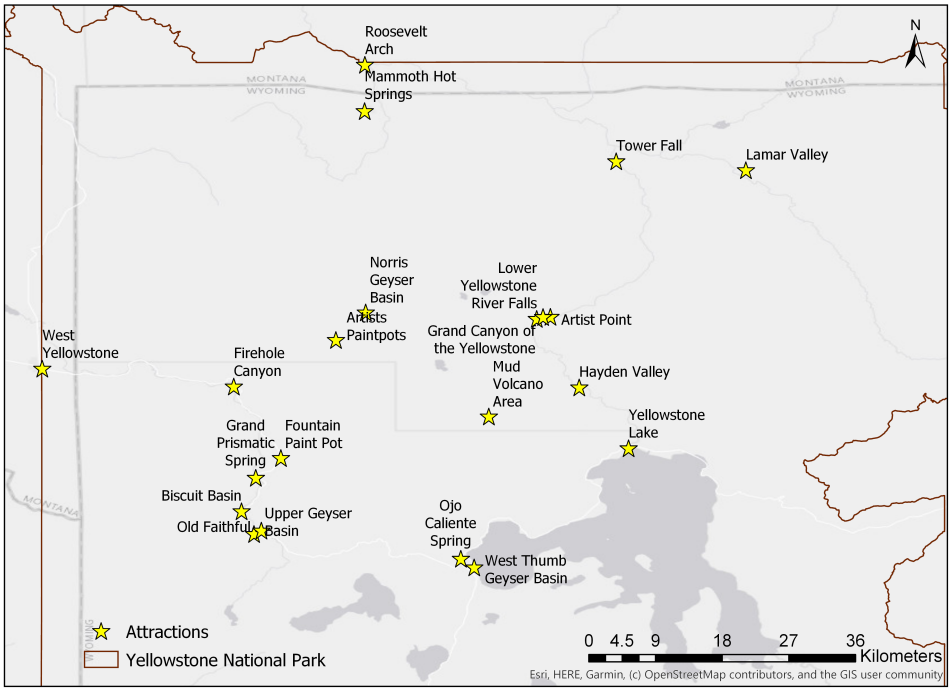


Figure 5.3: Geographic distribution of attractions in Yellowstone National Park

Table 5.5: OLS regression results for Yellowstone National Park

Origin Attraction	α	β	θ	MSE	R^2
West Yellowstone	1.4162***	-0.3506	0.0236	0.632	0.651
Roosevelt Arch	1.4039***	-1.1742*	0.3431***	0.361	0.885
Mammoth Hot Springs	1.4411***	-0.0555	-0.0836	0.311	0.745
Firehole Canyon	1.5694***	-0.3912***	-0.1206	0.244	0.816
Artists Paintpots	1.0360***	0.7745***	0.2500**	0.384	0.804
Norris Geyser Basin	1.3177***	-0.1650	0.0294	0.351	0.759
West Thumb Geyser Basin	1.4251***	0.4641	0.0625	0.483	0.740
Biscuit Basin	1.8398***	0.3523**	0.0054	0.491	0.793
Fountain Paint Pot	1.5536***	0.2773	0.1101	0.594	0.743
Grand Prismatic Spring	1.5115***	-0.1063	-0.1241	0.259	0.792
Yellowstone Lake	1.4656***	-0.2635	-0.0407	0.731	0.607
Old Faithful	1.3904***	-0.1682*	-0.1220	0.372	0.705
Tower Fall	2.6596***	-0.8268***	-0.6233***	0.324	0.822
Lamar Valley	1.3732***	-0.5286***	-0.0654	0.354	0.836
Hayden Valley	1.5363***	-0.1885	-0.1468	0.480	0.691
Lower Yellowstone River Falls	0.8800	-0.0979	0.1942	0.892	0.549
Artist Point	1.3240***	-0.2425**	-0.1704	0.337	0.672
Grand Canyon of the Yellowstone	1.8059***	-0.0418	-0.1378	0.567	0.710

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

5.2.3 Yellowstone National Park

Table 5.5 includes the parameter calibration results for attractions in Yellowstone National Park. We can see for attractions in Yellowstone National Park, the α estimations are relatively larger on average compared with those in Acadia and Yosemite National Parks. The larger α estimations for Tower Fall and Biscuit Basin indicate that visitors at these two attractions are more interested in the attractiveness of their next destinations, which are most likely Mammoth Hot Springs and Old Faithful, respectively. Compared with Yosemite National Park, we see larger absolute values of β estimations in Table 5.5. Even though Yellowstone National Park is three times the size of Yosemite National Park, visitors at many attractions in Yellowstone are more sensitive to distance (i.e., prefer closer next destinations), given the unique distribution of attractions in the upper and

lower loops of the park, shown in Figure. 5.3. The largest absolute value of β estimation for Roosevelt Arch indicates that the next destination for visitors at this attraction is most likely Mammoth Hot Springs, which is the nearest attraction. We also observe larger absolute values of β estimations for Tower Fall and Lamar Valley, indicating that there could be many interactions between these two attractions. In addition, we see a couple of positive β estimations for attractions like Artists Paintpots. This could be explained by its location in the middle of the park, serving as a connection between the upper and lower loops, so visitors at this attraction are willing to travel to further next destinations. The θ estimations for multiple attractions are close to 0, which indicate that visitors are less interested in the neighboring effect when choosing the next destinations. However, we can still infer from the larger θ estimations for Roosevelt Arch and Artists Paintpots that visitors at these two attractions prefer a next destination with more attractive and closer neighbors, such as the Grand Canyon of the Yellowstone, Lower Yellowstone River Falls, Artist Point cluster and the Grand Prismatic Spring, Biscuit Basin, Old Faithful cluster, as we can see in Figure. 5.3.

5.3 Temporal Variability of Parameters

To further explore the temporal variability of the model parameters, we divide the trips in Yosemite and Yellowstone National Parks to summer and non-summer based on the park travel recommendation.[1,2] Many of the attractions in the parks are seasonal, with roads and trails being closed due to snow in winter time. For example, Glacier Point in Yosemite National Park typically opens from May to November and the Grand Canyon of the Yellowstone is limited to snowmobile or snowcoach travel from early November to mid-April. According to most travel guides, the best time to visit Yosemite and

[1] `https://www.nps.gov/yose/planyourvisit/traffic.htm`
[2] `https://www.nps.gov/yell/planyourvisit/hours.htm`

Table 5.6: 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 $<$ 0.001; **p $<$ 0 .01; *p $<$ 0 .05. Summer months include May, June, July, August, and September.

Yellowstone National Parks is May to September. Hence, we use this time range to represent summer months here, and the rest as non-summer months. In Table 5.6 and Table 5.7, only origin attractions with more than 30 observed trips during both time periods are included.

From Table 5.6, 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 when choosing their next destinations 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

Table 5.7: OLS regression results for Yellowstone National Park Summer vs. Non–Summer months

Origin Attraction	Time of the year	α	β	θ	R^2
Mammoth Hot Springs	Summer	1.0645***	0.0120	0.0093	0.663
	Non-summer	1.7501***	-0.1434	-0.1715*	0.823
Firehole Canyon	Summer	1.4879***	-0.3617**	-0.1664	0.787
	Non-summer	1.5916***	-0.3432*	-0.0284	0.867
Grand Prismatic Spring	Summer	1.4223***	-0.1753**	-0.2511**	0.793
	Non-summer	1.6623***	-0.0513	-0.0381	0.864
Old Faithful	Summer	0.7861*	-0.2594**	-0.0453	0.687
	Non-summer	1.8930***	-0.0615	-0.1466	0.804

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

are attracted to closer attractions during summer months, and further attractions during non-summer months, which is potentially due to the closure of closer attractions in non-summer months. Meanwhile, we see 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 interested in clustered attractions (i.e., attractions with closer and more attractive neighbors) for their next destinations and a negative θ estimation means the opposite.

For Yellowstone National Park, we observe a much shorter list of attractions in Table 5.7. Many attractions are closed or have limited access in non-summer months. We can draw a similar conclusion that α estimations remain similar for different times of the year, except for Old Faithful. The α estimation is much smaller for Old Faithful in summer months, when Yellowstone National Park receives the most visitors.[3] As Old Faithful is ranked as the top 1 attraction in the Yellowstone travel guide,[4] visitors are less sensitive to attractiveness when choosing their next destinations after visiting Old Faithful. Similarly, we observe larger absolute values of β estimations in summer months

[3] https://www.nps.gov/yell/planyourvisit/summer.htm

[4] https://yellowstone.net/intro/top-10/

for Grand Prismatic Spring and Old Faithful. This can be explained by the closure of nearby attractions in non-summer months, thus visitors at these two attractions are attracted to further next destinations. We can conclude the same from the θ estimations. The negative values indicate that visitors tend to travel to dispersed attractions (i.e., attractions with further and less attractive neighbors), due to the limited access to most attractions in the park during non-summer months.

Chapter 6

Conclusions and Future Work

This work provides an extended Huff model to take social factors and neighboring effects into account. The model is calibrated and evaluated with an experiment using observed trip sequences extracted from geotagged Flickr photos taken in Acadia, Yosemite and Yellowstone National Parks over a ten-year period. This chapter summarizes the experiment results and presents directions for future areas of research.

6.1 Conclusions

In this work, we explore the visiting probabilities of attractions within three national parks using a socially aware Huff model, in which both social factors and neighboring effects are taken into account. 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. This confirms our assumption that the number of photo views would relate to the potential impact of social media influencers (SMIs), as they have more followers and their photos would have more views. 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 all three national parks.

We further explore the visiting patterns of each attraction within the three national parks based on model parameters of attractiveness, distance, and neighboring effect factors. We have shown that there is a regional and temporal variability of the model parameters. In general, visitors in Acadia National Park are more sensitive to the distance factor and neighboring effects when choosing their next destinations, while visitors in Yosemite and Yellowstone National Parks are more sensitive to the attractiveness factor. For temporal variability, the attractiveness parameter remains similar for both summer and non-summer time periods. However, the absolute value of the distance parameter is smaller and the neighboring effect parameter is closer to 0 in non-summer time, potentially due to closure of many attractions in the parks.

6.2 Future Work

This research can be extended in several aspects. First of all, 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 that affect the destination choice of tourists, such as online reviews, accommodation availability, travel expenses, etc. More information can be included in future work. In addition, prior work has shown that for long distance travel, the purely tourism trips are not the majority [59]. People usually combine sightseeing trips with activities such as visiting friends and family, while our data set does not have this information. The purpose of trips can be integrated into future research by extracting information from

the titles and tags of the geotagged photos.

Second, taking the social factor alone as an example, traveling under social influence or taking copycat photos [60] has become an emerging trend. In this work, we quantify the potential impact of social media influencers (SMIs) from the number of photo views, which is added to the evaluation process of place attractiveness. In the future, we can consider adopting a more comprehensive measurement to evaluate the social impact of the geotagged photos from the SMIs and better capture their relationship with the travel intentions.

Third, this work only studies existing attractions within national parks. We can examine both intra- and inter-park travel behavior in larger regions of interest in the future, such as Horseshoe Bend and its nearby attractions (Glen Canyon, Antelope Canyon, Grand Canyon, etc.). Moreover, future work can apply this model to discover emerging travel destinations. With more comprehensive measurement of the social impact, we can adjust the weights on the attractiveness factor and examine the time series of place interactions to uncover hidden gem places in the region of interest. In this way, we can further predict the popularity of an emerging travel destination in comparison with the known profile of hidden gem places like Devil's Bathtub in Virginia, Kanarraville Falls in Utah, etc.

Appendix A

Supplementary Data

A.1 Regression results with selection of K in K-NN

A.2 Attractions Summary

A.3 Software and Data Availability

Data used in this paper can be accessed with the public Flickr API.[1] The query used to access the data, software information, code and interactive data visualization (Fig. 3.4) related to Acadia and Yosemite National Parks are available on GitHub.[2] The workflow underlying this thesis 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`.

[1] `https://www.flickr.com/services/api/`
[2] `https://github.com/meilinshi/Socially-aware-Huff-model`

Table A.1: 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
Yellowstone National Park	All time	2	0.532940	0.693
		3	**0.532272**	**0.694**
		5	0.533368	0.693
	Summer months	2	0.369846	0.666
		3	0.368514	0.667
		5	**0.367671**	**0.668**
	Non-summer months	2	0.599632	0.785
		3	**0.598383**	**0.785**
		5	0.601543	0.784

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

Table A.2: 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 A.3: 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

Table A.4: Summary of attractions in Yellowstone National Park

Attraction	Number of photos	Outgoing trips	Incoming trips
West Yellowstone	982	117	94
Roosevelt Arch	734	161	148
Mammoth Hot Springs	7608	613	610
Firehole Canyon	1248	260	225
Artists Paintpots	1046	149	156
Norris Geyser Basin	2945	324	324
Ojo Caliente Spring	1101	12	15
West Thumb Geyser Basin	2176	234	194
Biscuit Basin	1157	182	200
Fountain Paint Pot	2194	286	289
Grand Prismatic Spring	4196	500	519
Mud Volcano Area	1248	14	14
Yellowstone Lake	1088	158	167
Upper Geyser Basin	1044	22	19
Old Faithful	6046	568	531
Tower Fall	2648	420	440
Lamar Valley	1537	231	248
Hayden Valley	2801	346	375
Lower Yellowstone River Falls	1053	206	205
Artist Point	1402	286	300
Grand Canyon of the Yellowstone	738	143	159

Bibliography

[1] A. Balmford, J. Beresford, J. Green, R. Naidoo, M. Walpole, and A. Manica, *A Global Perspective on Trends in Nature-Based Tourism*, *PLoS Biology* **7** (2009), no. 6 e1000144.

[2] J. Puustinen, E. Pouta, 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** (2009), no. 1 18–31.

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

[4] D. Leung, R. Law, H. Van Hoof, and D. Buhalis, *Social media in tourism and hospitality: A literature review*, *Journal of travel & tourism marketing* **30** (2013), no. 1-2 3–22.

[5] C. Djossa, *When not to geotag while traveling*, *National Geographic* (2019).

[6] P. Glover, *Celebrity endorsement in tourism advertising: Effects on destination image*, *Journal of Hospitality and Tourism Management* **16** (2009), no. 1 16–23.

[7] K. Freberg, K. Graham, K. McGaughey, and L. A. Freberg, *Who are the social media influencers? a study of public perceptions of personality*, *Public Relations Review* **37** (2011), no. 1 90–92.

[8] Z. Li, *Psychological empowerment on social media: who are the empowered users?*, *Public Relations Review* **42** (2016), no. 1 49–59.

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

[10] S. A. Wood, A. D. Guerry, J. M. Silver, and M. Lacayo, *Using social media to quantify nature-based tourism and recreation*, *Scientific reports* **3** (2013), no. 1 1–7.

[11] F. Orsi and D. Geneletti, *Using geotagged photographs and gis analysis to estimate visitor flows in natural areas*, *Journal for Nature Conservation* **21** (2013), no. 5 359–368.

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

[13] V. Heikinheimo, E. D. Minin, H. Tenkanen, A. Hausmann, J. Erkkonen, and T. Toivonen, *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** (2017), no. 3 85.

[14] D. L. Huff, *Defining and Estimating a Trading Area*, *Journal of Marketing* **28** (1964), no. 3 34–38.

[15] J. Li, L. Xu, L. Tang, S. Wang, and L. Li, *Big data in tourism research: A literature review*, *Tourism Management* **68** (2018) 301–323.

[16] Y.-T. Zheng, Z.-J. Zha, and T.-S. Chua, *Mining travel patterns from geotagged photos*, *ACM Trans. Intell. Syst. Technol* **3** (2012).

[17] F. Hu, Z. Li, C. Yang, and Y. Jiang, *A graph-based approach to detecting tourist movement patterns using social media data*, *Cartography and Geographic Information Science* **46** (2019), no. 4 368–382.

[18] R. Ji, Y. Gao, B. Zhong, H. Yao, and Q. Tian, *Mining flickr landmarks by modeling reconstruction sparsity*, *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* **7** (2011), no. 1 1–22.

[19] A. Majid, L. Chen, G. Chen, H. T. Mirza, I. Hussain, and J. Woodward, *A context-aware personalized travel recommendation system based on geotagged social media data mining*, *International Journal of Geographical Information Science* **27** (2013), no. 4 662–684.

[20] D. Li, X. Zhou, and M. Wang, *Analyzing and visualizing the spatial interactions between tourists and locals: A flickr study in ten us cities*, *Cities* **74** (2018) 249–258.

[21] T. N. Maeda, M. Yoshida, F. Toriumi, and H. Ohashi, *Extraction of tourist destinations and comparative analysis of preferences between foreign tourists and domestic tourists on the basis of geotagged social media data*, *ISPRS International Journal of Geo-Information* **7** (2018), no. 3 99.

[22] A. Chua, L. Servillo, E. Marcheggiani, and A. V. Moere, *Mapping cilento: Using geotagged social media data to characterize tourist flows in southern italy*, *Tourism Management* **57** (2016) 295–310.

[23] W. Chen, Z. Xu, X. Zheng, and Y. Luo, *Geo-tagged photo metadata processing method for beijing inbound tourism flow*, *ISPRS International Journal of Geo-Information* **8** (2019), no. 12 556.

[24] N. Mou, R. Yuan, T. Yang, H. Zhang, J. J. Tang, and T. Makkonen, *Exploring spatio-temporal changes of city inbound tourism flow: The case of shanghai, china*, *Tourism Management* **76** (2020) 103955.

[25] A. Hausmann, T. Toivonen, C. Fink, V. Heikinheimo, R. Kulkarni, H. Tenkanen, and E. Di Minin, *Understanding sentiment of national park visitors from social media data*, *People and Nature* **2** (2020), no. 3 750–760.

[26] Y. Yan, J. Chen, and Z. Wang, *Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations*, *Applied Geography* **123** (2020) 102306.

[27] W. Jiang, Z. Xiong, Q. Su, Y. Long, X. Song, and P. Sun, *Using geotagged social media data to explore sentiment changes in tourist flow: A spatiotemporal analytical framework*, *ISPRS International Journal of Geo-Information* **10** (2021), no. 3 135.

[28] B. Zeng and R. Gerritsen, *What do we know about social media in tourism? a review*, *Tourism management perspectives* **10** (2014) 27–36.

[29] G. Bakr and I. E. H. Ali, *The role of social networking sites in promoting egypt as an international tourist destination*, *South Asian Journal of Tourism and Heritage* **6** (2013), no. 1 169–183.

[30] S. W. Litvin, R. E. Goldsmith, and B. Pan, *Electronic word-of-mouth in hospitality and tourism management*, *Tourism management* **29** (2008), no. 3 458–468.

[31] H. Parsons, *Does social media influence an individual's decision to visit tourist destinations? Using a case study of Instagram.* PhD thesis, Cardiff Metropolitan University, 2017.

[32] R.-A. Pop, Z. Săplăcan, D.-C. Dabija, and M.-A. Alt, *The impact of social media influencers on travel decisions: The role of trust in consumer decision journey*, *Current Issues in Tourism* (2021) 1–21.

[33] M. R. Jalilvand and N. Samiei, *The impact of electronic word of mouth on a tourism destination choice: Testing the theory of planned behavior (TPB)*, *Internet Research* **22** (2012), no. 5 591–612.

[34] M. M. Shafiee, R. A. Tabaeeian, and H. Tavakoli, *The effect of destination image on tourist satisfaction, intention to revisit and wom: An empirical research in foursquare social media*, in *2016 10th International Conference on e-Commerce in Developing Countries: with focus on e-Tourism (ECDC)*, pp. 1–8, IEEE, 2016.

[35] J. Hernández-Méndez, F. Muñoz-Leiva, and J. Sánchez-Fernández, *The influence of e-word-of-mouth on travel decision-making: consumer profiles*, *Current issues in tourism* **18** (2015), no. 11 1001–1021.

[36] A. Tham, J. Mair, and G. Croy, *Social media influence on tourists' destination choice: importance of context*, *Tourism Recreation Research* **45** (2020), no. 2 161–175.

[37] J. L. Nicolau and F. J. Más, *Sequential choice behavior: Going on vacation and type of destination*, *Tourism Management* **29** (2008), no. 5 1023–1034.

[38] L. Wu, J. Zhang, and A. Fujiwara, *A Tourist's Multi-Destination Choice Model with Future Dependency*, *Asia Pacific Journal of Tourism Research* **17** (2012), no. 2 121–132.

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

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

[41] J. L. Nicolau, *Characterizing Tourist Sensitivity to Distance*, *Journal of Travel Research* **47** (2008), no. 1 43–52.

[42] C. Tideswell and B. Faulkner, *Multidestination travel patterns of international visitors to queensland*, *Journal of Travel research* **37** (1999), no. 4 364–374.

[43] S. Gong, J. Cartlidge, R. Bai, Y. Yue, Q. Li, and G. Qiu, *Geographical and temporal huff model calibration using taxi trajectory data*, *GeoInformatica* (2020) 1–28.

[44] Y. Liang, S. Gao, Y. Cai, N. Z. Foutz, and L. Wu, *Calibrating the dynamic Huff model for business analysis using location big data*, *Transactions in GIS* **24** (2020), no. 3 681–703.

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

[46] R. J. Campello, D. Moulavi, and J. Sander, *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.

[47] L. McInnes, J. Healy, and S. Astels, *hdbscan: Hierarchical density based clustering*, *The Journal of Open Source Software* **2** (2017), no. 11 205.

[48] S. A. Stouffer, *Intervening Opportunities: A Theory Relating Mobility and Distance*, *American Sociological Review* **5** (1940), no. 6 845.

[49] S. Um and J. L. Crompton, *Attitude determinants in tourism destination choice*, *Annals of Tourism Research* **17** (1990), no. 3 432–448.

[50] T. Orpana and J. Lampinen, *Building Spatial Choice Models from Aggregate Data*, *Journal of Regional Science* **43** (2003), no. 2 319–348.

[51] W. G. Hansen, *How Accessibility Shapes Land Use*, *Journal of the American Planning Association* **25** (1959), no. 2 73–76.

[52] B. Kádár and M. Gede, *Where Do Tourists Go? Visualizing and Analysing the Spatial Distribution of Geotagged Photography*, *Cartographica: The International Journal for Geographic Information and Geovisualization* **48** (2013), no. 2 78–88.

[53] R. Leung, H. Q. Vu, and J. Rong, *Understanding tourists' photo sharing and visit pattern at non-first tier attractions via geotagged photos*, *Information Technology and Tourism* **17** (2017), no. 1 55–74.

[54] M. Nakanishi and L. G. Cooper, *Parameter Estimation for a Multiplicative Competitive Interaction Model: Least Squares Approach*, *Journal of Marketing Research* **11** (1974), no. 3 303.

[55] K. P. Burnham and D. R. Anderson, *A practical information-theoretic approach*, *Model selection and multimodel inference* **2** (2002).

[56] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, in *Selected papers of hirotugu akaike*, pp. 199–213. Springer, 1998.

[57] E.-J. Wagenmakers and S. Farrell, *Aic model selection using akaike weights*, *Psychonomic bulletin & review* **11** (2004), no. 1 192–196.

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

[59] A. W. Davis, E. C. McBride, K. Janowicz, R. Zhu, and K. G. Goulias, *Tour-based path analysis of long-distance non-commute travel behavior in california*, *Transportation research record* **2672** (2018), no. 49 1–11.

[60] R. Picheta, *New zealand tells tourists to stop copying other people's travel photos*, *CNN* (2021).