METHODOLOGICAL ADVANCES FOR UNDERSTANDING SOCIAL CONNECTIVITY AND ENVIRONMENTAL IMPLICATIONS IN MULTI-USE LANDSCAPES

by

Matthew Clark

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Biology

Boise State University

August 2019

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

Matthew Clark

Thesis Title: Methodological Advances for Understanding Social Connectivity and Environmental Implications in Multi-Use Landscapes

Date of Final Oral Examination: 14 June 2019

The following individuals read and discussed the thesis submitted by student Matthew Clark, and they evaluated his presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Vicken Hillis, Ph.D.	Chair, Supervisory Committee
Trevor T. Caughlin, Ph.D.	Member, Supervisory Committee
Marie-Anne de Graaff, Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Vicken Hillis, Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

DEDICATION

Dedicated to all the people in my life who refuse to take themselves too seriously…and

to Senna, my first and last kiss.

ACKNOWLEDGEMENTS

My graduate adviser Dr. Vicken Hillis is far too nice. From late night manuscript edits, to hours of always-productive conversation, and even letting me indefinitely "borrow" your climbing gear, sincerely thank you. I would also like to thank the rest of the Human-Environment Systems Faculty at Boise State, as well as my committee members for their unwavering support and enthusiasm. Lastly, I would like to thank my lab mates putting up with me over the last two years and pretending they enjoyed learning 'R' every Friday morning.

ABSTRACT

Integrated social-ecological systems research is challenging; complicated feedback and interactions across scales in multi-use landscapes are difficult to decouple. Novel methods and innovative data sources are needed to advance social-ecological systems research. In this thesis, we use network science as a means of explicitly assessing feedback between social and ecological systems, and internet search data to better predict visitation in protected areas. This thesis seeks to provide empirical examples of emerging social-ecological systems science methods as a precedent for resource managers on-the-ground, as well as extending the line of scientific inquiry on the subject.

In the first chapter of this thesis, we used an online survey to gather information on the collaborative network and current projects of 169 wetland management organizations in the state of Montana. We used this information along with geographic analyses to delineate the flow of information between managers and ecological connectivity of projects, characterizing the social-ecological network of wetlands and wetland management within the state. We demonstrate that just 2 key organizations facilitate landscape scale information sharing, while most stakeholders collaborate on the basis of project difficulty and proximity <10km. This chapter contributes to an emerging body of literature on social-ecological networks, a promising frontier for integrating social and environmental sciences, specifically addressing feedbacks within and between the two systems.

For the second part of this thesis, we apply novel data to a classic natural resource management problem. In recent years, visitation to U.S. National Parks has been increasing, with the majority of this increase occurring in a subset of parks. Improved visitation forecasting would allow park managers to more proactively plan for such increases and subsequent visitor-related challenges. In this study, we leverage internet search data that is freely available through Google Trends to create a forecasting model. We compare this Google Trends model to a traditional autoregressive forecasting model. Overall, our Google Trends model accurately predicted 97% of the total visitation variation to all parks one year in advance from 2013-2017 and outperformed the autoregressive model by all metrics. While our Google Trends model performs better overall, this was not the case for each park unit individually; the accuracy of this model varied significantly from park to park. This project applies a contemporary social science data set to a traditional natural resource management problem, demonstrating the potential for social-ecological systems research to provide real-world solutions in multi-use landscapes. Both chapters of this thesis explicitly address feedbacks between social and ecological systems, a key advance for social-ecological systems science.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: NETWORK GOVERNANCE OF NATURAL RESOURCES: MAKING COLLABORATION COUNT

Abstract

In contemporary multi-use landscapes, management of ecological resources is essential for environmental and societal well-being. Management efficacy is often constrained by the capacity of individual organizations to act at the scale of ecological processes. Ecological processes function at landscape scales, while management of natural resources consists of an overlapping patchwork of jurisdiction and influence. Collaboration is a common prescription for the cohesive management of ecological resources at the landscape scale, but collaboration is costly. Land management organizations must decisively pick and prune their collaborations with other stakeholders to best match the ecological connectivity of the landscapes they manage. Empirical studies have demonstrated the utility of social-ecological networks to quantify fit in coupled natural and human systems and make concrete prescriptions about collaborative resource management. Social-ecological network science characterizes resource and management systems as an interconnected network of nodes (organizations, resource patches) and ties (collaboration, connectivity, management). Previous studies have used single distance thresholds to define ecological connectivity and estimate ecological outcomes at the whole system scale. With this research, we explore the potential biases that can be introduced into social-ecological network analyses by setting single connectivity thresholds and demonstrate the utility of incorporating ecological outcomes

on the scale of individual patches opposed to the whole system. For this research, we delineate the social-ecological network of wetlands and wetland management in Montana, U.S. We address the current gaps in social-ecological network methodology in two key ways. We use a gradient of wetland connectivity to illustrate the possible ramifications of defining set connectivity thresholds in social-ecological network studies. We also incorporate a measure of wetland vegetation quality into our descriptive analysis to better understand the role of environmental condition in the system. Using these methodological advances, we discover that just two wetland management organizations in the system are responsible for ensuring efficient information diffusion and facilitating cohesive wetland management at the landscape scale. This project makes a methodological contribution to social-ecological network science broadly by exposing sources of potential bias and assessing outcomes at a finer scale than previous work.

Introduction

Ecological processes generally occur on a scale larger than any one entity can manage (Cadenasso 2003, Cowling, Egoh, Knight, O'Farrell, Reyers, Rouget & Wilhelm-Rechman 2008; Yarrow & Marín 2007). Because no single decision maker has the capacity to oversee entire ecoregions, the burden of management is spread among many stakeholders in an overlapping mosaic of jurisdictions that rarely coincide with ecological boundaries (Dallimer & Strange 2015; de Groot, Alkemade, Braat, Hein & Willemen 2010; Hamilton, Fischer & Ager 2019; Hein, van Koppen, de Groot & van Ierland 2006). In the American West, resource governance is further fragmented by a variety of social factors including historic land ownership, private interests, and government hierarchies (Andrews 2006; Kauffman 2002).

Within complex jurisdictional patchworks, research shows that collaboration between independent entities can lead to more efficient problem solving and improved environmental outcomes, as compared to siloed governance (Miller, Zhao & Calantone 2006; Scott 2015). The structure of collaboration, which organizations collaborate with which others, influences the ability of actors to solve complex problems (Mason & Watts 2012). Collaboration, notably, comes at a substantial cost for stakeholders in the form of staff time and financial investment (Koontz & Thomas 2006; March 1991). With these costs in mind, it follows that land management organizations should aim to maximize their environmental returns on investing in collaboration. Characterizing the tangible ecological impacts of specific collaborative arrangements and identifying worthwhile or deleterious collaborations, however, have proved difficult (Crona & Hubacek 2010).

The contribution any specific collaboration makes to address the cohesive management of a resource depends largely on the connectivity of the ecological system itself (Bodin, Alexander, Baggio, Barnes, Berardo, Cumming & Sayles 2019). For example, collaborative management of disconnected resources is superfluous, while collaborative management of highly connected resources is worthwhile. In addition to the management implications, ecological connectivity in general has considerable impact on the ecological condition of both terrestrial and aquatic resources (McRae, Hall, Beier & Theobald 2012; Wolf, Noe & Ahn, 2013). Species dispersal distances and community composition depend largely on ecological connectivity (Kareiva & Wennergren 1995; Ricketts 2001). The degree to which any given landscape is connected however can vary greatly depending on the species or mechanism of interest (Bunn, Urban & Keitt 2000; Laita, Kotiaho & Mönkkönen 2011). In wetland systems, surface water connectivity is highly indicative of wetland nutrient cycling, a key consideration for studying wetland vegetation composition (Cook & Hauer 2007). Defining ecological connectivity through hydrology however is likely less relevant when interested in avian dispersal. Ecological connectivity, specific to the species of interest therefore, is a key consideration when assessing fit of organizational collaborations to the resources they manage.

Social-ecological networks are a promising tool to assess the fit, or degree of alignment, between natural systems and the social institutions that manage them (Bodin 2017; Sayles & Baggio 2017; Treml, Fidelman, Kininmonth, Ekstrom & Bodin, 2015). This lens for studying coupled natural and human systems delineates two distinct, but connected networks of nodes representing organizations or ecological patches, and ties representing social collaboration, ecological connectivity, or management actions. Studying complex systems, like resource management in the American West, using a network approach allows for a nuanced understanding of the degree to which relationships dictate outcomes (Jackson 2010, Newman 2010; Tassier 2013). For example, Guerrero, Bodin, McAllister & Wilson (2015) used social-ecological networks to empirically assess the fit of a collaborative restoration initiative to the ecological connectivity of native vegetation in Western Australia. Similarly, Kininmonth, Bergsten & Bodin (2015) used this framework to demonstrate how Swedish municipalities can utilize coordinating third party actors to best manage interconnected wetlands.

Defining social connectivity in coupled natural and human systems is often unequivocal; people can report who they communicate with and document analysis can detail formal collaborations (Nkhata, Breen & Freimund 2008). Defining connectivity between discrete ecological resources such as wetlands, however, has proved more challenging (Leibowitz, Wigington, Rains & Downing 2008). When building networks of ecological connectivity, social-ecological network analyses commonly specify distance thresholds to define resources connectivity (Guerrero et al. 2015). As described above, describing ecological connectivity without considering the natural history of the species or mechanism in question likely constitutes a significant loss of valuable information. Additionally, we do not yet understand how setting different connectivity thresholds may bias the results of social-ecological network studies and generate misleading conclusions.

Furthermore, while social-ecological network measures have proved useful in quantifying the system-level fit of natural resource management, they have seldom been associated with ecological outcomes on the scale of each observation (i.e. the node level) (Barnes et al. 2019). For example, Bodin et al. (2014) used social-ecological network analysis to compare the fit of two distinct common-pool resource use systems, using the overall state of the resource as the outcome variable. Natural resource management and ecological research often focus on ecological outcomes at the scale of individual units or patches of interest. Hence, the ability to estimate the impact of network position on individual patches would greatly advance the utility of social-ecological network analysis.

Lastly, social-ecological network science theory and methodology have progressed rapidly since the framework was first proposed (Bodin & Tengö 2012). These advances, while impressive and worthwhile, have neglected the literature regarding complex problem solving in social networks. The capacity for a social network to rapidly diffuse important information to all actors is critical for comprehensively adapting to disturbances in coupled natural and human systems (Baggio & Hillis 2018). Failure to estimate the ability of the associated social systems to circulate beneficial information represents a missed opportunity to better understand and frame this emerging field.

In this study, we examine the social-ecological network structure of wetland management in Montana. We make three specific contributions that address the gaps described in the preceding paragraphs. First, we define ecological connectivity as a gradient of varying thresholds to both explore the utility of this method and to recognize the ramifications and potential biases of defining arbitrary thresholds. We also incorporate a measure of ecological condition at the node level to draw descriptive inference about the feedback between environmental health and social-ecological network structure. Finally, we examine how social-ecological network analyses can be better understood and corroborated by further exploring the capacity of the social network to rapidly diffuse information and solve complex problems.

We delineate the social-ecological network of wetland managers and wetlands in Montana, U.S. for this empirical research. While addressing the methodological gaps outlined above, we aim to answer several key research questions: To what degree is general or *any* collaboration associated with improved ecological condition? How readily and on what basis do wetland managers in the state collaborate? And lastly, what are the implications of these observed trends on the capacity of wetland managers to efficiently solve complex problems? While this research provides considerable insight for wetland management in the state of Montana, our aim is rather to make methodological advances and expand the line of inquiry for social-ecological network science broadly.

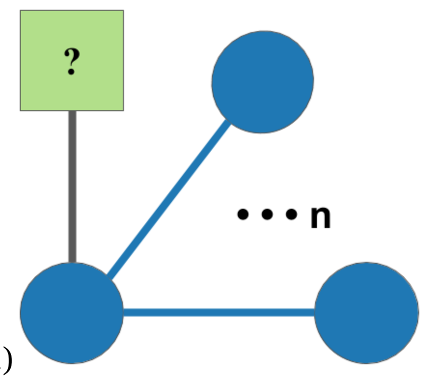
Methods and Data

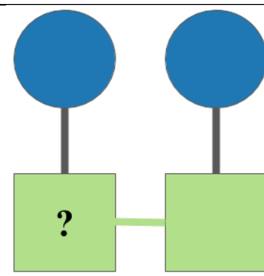
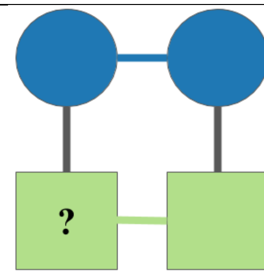
Conceptual Framework

In this research, we analyze two distinct, but highly interconnected networks. These include the collaborative network of Montana wetland management organizations and the wetland systems they manage. We refer to this two level network as a social-ecological network. Our framework for understanding this social-ecological network builds upon the established framework developed by Bodin & Tengö (2012). We first define network substructures, or building blocks, theorized to be important to our outcome of interest, effective resource management (Table 1). We then survey the social-ecological network for the occurrences of these building blocks, comparing them to expected occurrences given stochastic network formation or to each other. In our analyses, similar to the recent work by Barnes et al. (2019), we also draw inferences about the association between social-ecological network structure and resource health by incorporating a measure of wetland vegetation condition at the node level (Table 1).

We investigate our research questions by focusing on two key building blocks. Building block 1 represents the number of reported collaborations of each wetland managing organization and the reported environmental condition of their associated wetlands. This building block is imperative as a baseline for this study to understand how any collaboration, regardless of structure, is associated with wetland condition. The second building block we identified as critical for this study represents siloed (2a) or collaborative (2b) management of connected resources and the associated ecological condition of wetlands within each structure. Using building block 2, we are able to determine at what level of ecological connectivity between projects organizations are more likely to collaborate and the association between these collaborations and ecological condition. We also use building block 2 to explore the possible biases which can be introduced into social-ecological network studies by setting blanket connectivity thresholds.

Table 1.1: Social-ecological network building blocks modified from Guerrero et al. (2015) & Bodin et al. (2016)*.

Theory	**Building block**
1. *Degree of managing organization.* The number of collaborations, or degree, of an organization increases their access to relevant information and their influence within the network (Scott 2015). This is theorized to have an association with the ecological condition of the resources they manage.	
2. *Collaborative management of connected resources.* The position of an ecological node in either an open (a) or	

closed (b) square is measure of organizational collaboration (or lack thereof) on management of connected resources. This is theorized to be an indicator of social-ecological fit with implications for ecological condition (Bodin et al. 2016).	(2a)  (2b) 
* Social nodes are represented by blue circles and the connections between them by blue lines. Ecological nodes are represented by green squares and the connections between them by green lines. Resource management is represented by the grey lines between the social and ecological nodes. The “?” indicates that we are interested in node level characteristics of nodes in that specific position within the building blocks.	

Study Area & Scope

To answer our research questions, we chose to focus on wetlands and organizations involved in wetland management in the state of Montana, U.S. Wetlands systems are fitting for this research because individual wetlands are discrete in nature, but highly connected at the landscape scale (Calhoun et al. 2017). We concentrate on Montana because wetland restoration, mitigation, and preservation have emerged as a top priority for land management within the state (Montana Department of Environmental

Quality 2013). Montana has approximately 2.5 million acres of wetlands within the state, representing 2.6% of the land cover (Montana Wetland & Riparian Mapping Center 2019). These wetland areas are managed by over 150 different organizations, encompassing stakeholders at federal, state, and county scales, representing government, private, non-profit, and tribal interests. While we focused on capturing organizations who work within the state of Montana, some organizations included in the study are not physically located within the state, as they have jurisdictions that span multiple state lines. We treated these organizations no differently than those who have home offices within the state.

Data Collection

In this study, we aimed to identify and survey all organizations involved in wetland management in the state of Montana. To do this, we began with simple internet searches using key words such as: “Montana,” “wetlands,” “restoration,” “riparian,” “conservation,” etc. We then evaluated each resulting organization individually for relevance to this research. Once we believed we had a relatively representative sample of organizations, we used unstructured interviews with five key organizations to identify stakeholders we had missed through internet searches.

After our first round of identifying wetlands management organizations, we used Qualtrics (2017) survey software to design and distribute an online survey to all identified organizations (S1). This survey used a roster, or list, format to allow respondents to select other organizations with whom they collaborate on wetland management. In addition to the list of identified organizations, the survey also allowed organizations to self-identify any missing organizations who they collaborate with on wetland management. We then surveyed all relevant, newly identified organizations through snowball sampling. We also asked survey respondents to answer a variety of questions regarding the function of their organization in order to determine their relevance to this study and to classify each response as either federal, state, county, tribal, non-profit, or private.

To ensure that the ecological measures used in this study were in-line with wetland function, we first defined our environmental outcome of interest (reference quality of the wetland, i.e. vegetative makeup), and then defined reasonable connectivity thresholds based on this outcome. Wetland vegetation makeup is heavily influenced by nutrient flow from adjacent areas (<5km); this effect is diminished as distance increases (Houlahan, Keddy, Makkay & Findlay 2006). With this in mind, we constructed ecological networks using 1, 2, 5, 10, & 20km connectivity thresholds.

To gather relevant ecological data, we asked respondents to identify specific wetlands that have been a focus for their organization in the last year (name, lat/long) and estimate the ecological condition of these wetlands compared to a reference (pristine) wetland. Respondents reported ecological condition of their identified wetlands on a 4-factor Likert scale where the lowest score represents a highly degraded wetland and the highest represents a reference or pristine wetland (Table 2).

Table 1.2: Likert scale used to assess wetland vegetation condition

Score	Wetland Vegetation Condition
4	At a reference condition, i.e. pristine wetland with all native species
3	Level of disturbance indicates a slight departure from a reference condition
2	Level of disturbance indicates moderate departure from a reference condition
1	Level of disturbance indicates severe departure from a reference condition

To assemble the ecological networks, we created 1, 2, 5, 10, & 20km buffer areas around each identified wetland using ArcGis software (Fig. 1). We then created connectivity matrices for each threshold area, taking two wetlands as connected if the lat/long coordinate provided by the survey respondent of one wetland was within the buffer of the other.

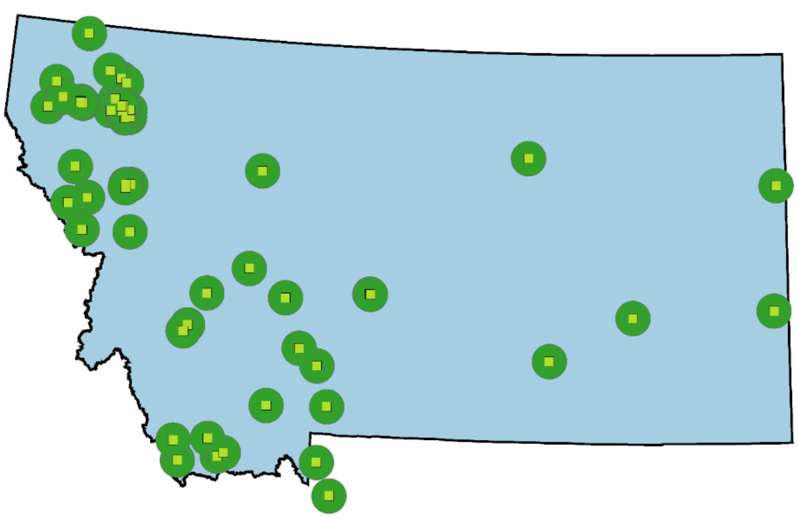


Figure 1.1 Simplified map of the wetlands and the ecological connectivity measure used in our study. The light green squares represent wetlands that were identified using the online survey. Dark green circles are a 20km threshold around each wetland.

Our sampling efforts in total produced data on the collaborative structure of 169 wetland management organizations and 55 managed wetlands. Using the inherent information on the management of these wetlands, we were able to link both networks into a complete social-ecological network for analysis.

Analyses

Social-Ecological Estimation

All two level (social-ecological) network analyses were completed using a combination of MPnet exponential random graph model simulation and estimation software for multilevel networks (Wang, Robins & Pattison 2009) and the 'R' coding language for statistical computing (2018). Using MPnet, we were able to estimate the prevalence of social collaboration within our network compared to what would be expected given stochastic network formation. This method is referred to as exponential random graph modeling (Frank & Strauss 1986; Wang, Robins, Pattison & Lazega 2013). Exponential random graph models compare observed network statistics to some number of randomly simulated networks of similar specifications (1,000 in this case). We use this method to calculate the number of ties (n) in building block 1 (Table 1) which would be expected given stochastic network formation and compare this to our observed network. This method was first proposed for use in social-ecological network analysis by Bodin & Tengö (2012).

Using MPnet, we were also able to count the occurrences of building blocks 2a and 2b (Table 1) and count the number of wetlands at or near a reference condition (reported condition of 3 or 4) in each configuration. These counts allowed us to make descriptive inferences about collaborative management of connected resources in this system, as well as to explore the implications and potential biases introduced by set connectivity thresholds in social-ecological network studies.

Social Network Exploration

To better understand the formation and implications of our observed social-ecological network, we further explored our study system using established social network metrics. All one level (social) network analyses were completed using ‘R.’ We intended to understand the overall structure of the collaborative network of Montana wetland management organizations by estimating the capacity for complex problem solving within our social network as a function of the observed social-ecological network.

We first assessed the modularity of the social network. To determine if the entire network is dominated by one cohesive core or multiple sub groups, we used the random walk method developed by Rosvall & Bergstrom (2008). This method, implemented in the ‘igraph’ (2006) package for ‘R’, maps the probability of information flows within a network to delineate the number and structure of distinct modules (Csardi & Nepusz 2006; Rosvall & Bergstrom 2008; Rosvall, Axelsson & Bergstrom 2009).

We further assessed the modularity of the social network by applying a k-core decomposition algorithm to identify the core organizations. This analysis was also done using ‘igraph’. The k-core algorithm defines a minimum set of ties k and recursively removes all nodes with fewer than k ties, maximizing k to produce the optimum core (Batagelj & Zaversnik 2002; Seidman 1983).

We then calculated the degree to which each management organization plays a bridging role, or contributes to the overall connectivity of the network. We estimated an organization’s role in bridging by calculating the betweenness centrality for each node. Betwennness centrality is a standard proxy for estimating an organization’s likelihood to fulfil a bridging role within a network (Berardo 2014; Geys & Murdoch 2010).

$$(V) = \sum_{s \neq v \neq t} \sigma_{st}(v)/\sigma_{st}$$

The betweenness centrality of any given node V is represented by the proportion of shortest paths σ between all combinations of nodes s & t which pass through node v. The betweenness centrality for any given organization is therefore representative of the number of times that the shortest path between any two organizations in the network goes through that specific organization.

Results

Social-Ecological Network Findings

Building Block 1

To estimate the association between an organization’s social connectivity and the ecological condition of the wetlands they manage, we ran a correlation test between the number of ties (degree) of each organization and the average ecological quality of the wetlands they reported managing. This yielded a very weak correlation of 0.17 (Fig. 2). This result is in-line with current literature which suggests that increased collaboration alone is not an adequate prescription for improving natural resource management.

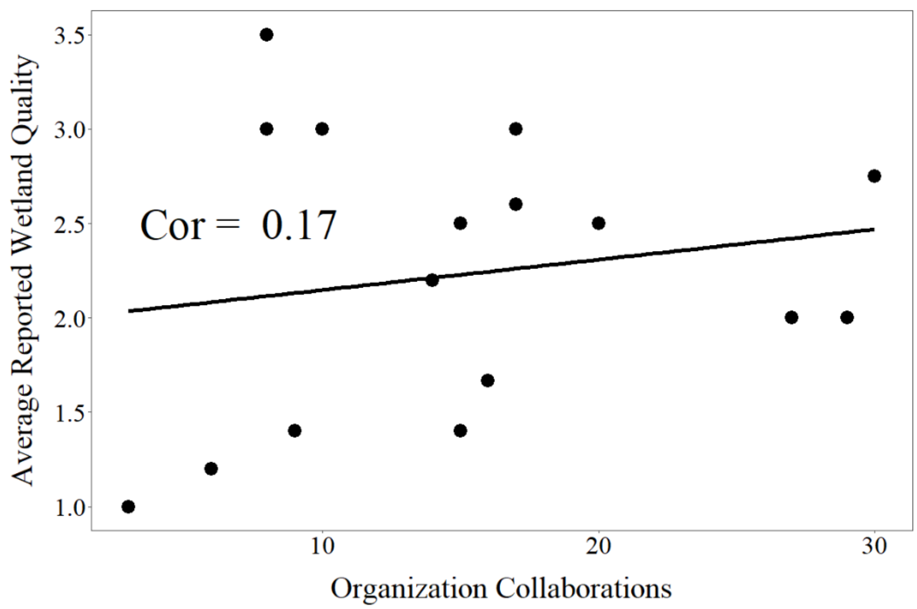


Figure 1.2. Correlation between the number of collaborations each organization reported (degree) and the average ecological condition of each organization's reported wetlands (quality). Wetland quality was reported on a factor scale from 1-4, where 1 represents a highly degraded wetland and 4 represents a pristine or reference condition wetland.

We assessed the degree to which wetland management organizations are collaborating on wetland projects compared to what would be expected under stochastic network formation. The resulting parameter estimate from our two level exponential random graph modeling was -0.49 with a standard error of 0.002. When an absolute value of an exponential random graph modeling estimate is more than 2x that of the standard error, the results are considered significant. This significant, negative output indicates that wetland management organizations collaborate significantly less (n) than we would expect given stochastic network formation.

Building Block 2

We counted the occurrences of both building blocks 2a and 2b, representing siloed and collaborative management of connected resources respectively. We counted these occurrences for our connectivity thresholds of 2, 5, 10, & 20km and counted the number of wetlands at or near a reference condition in each substructure (reported condition 3 or 4). Results from this descriptive analysis indicate that wetland management organizations tend to collaborate on connected wetland projects when the wetlands are further from a reference condition, i.e. more highly degraded. These results also suggest that this effect is exacerbated by increased proximity of the wetland projects (Fig. 3). This finding also demonstrates that results from social-ecological analyses can be variable depending on the defined threshold for ecological connectivity. In summary, this analysis shows that collaboration between wetland management organizations is associated with increasing project proximity and reduced ecological condition and that the ratio of observed substructures is variable based on the ecological connectivity threshold.

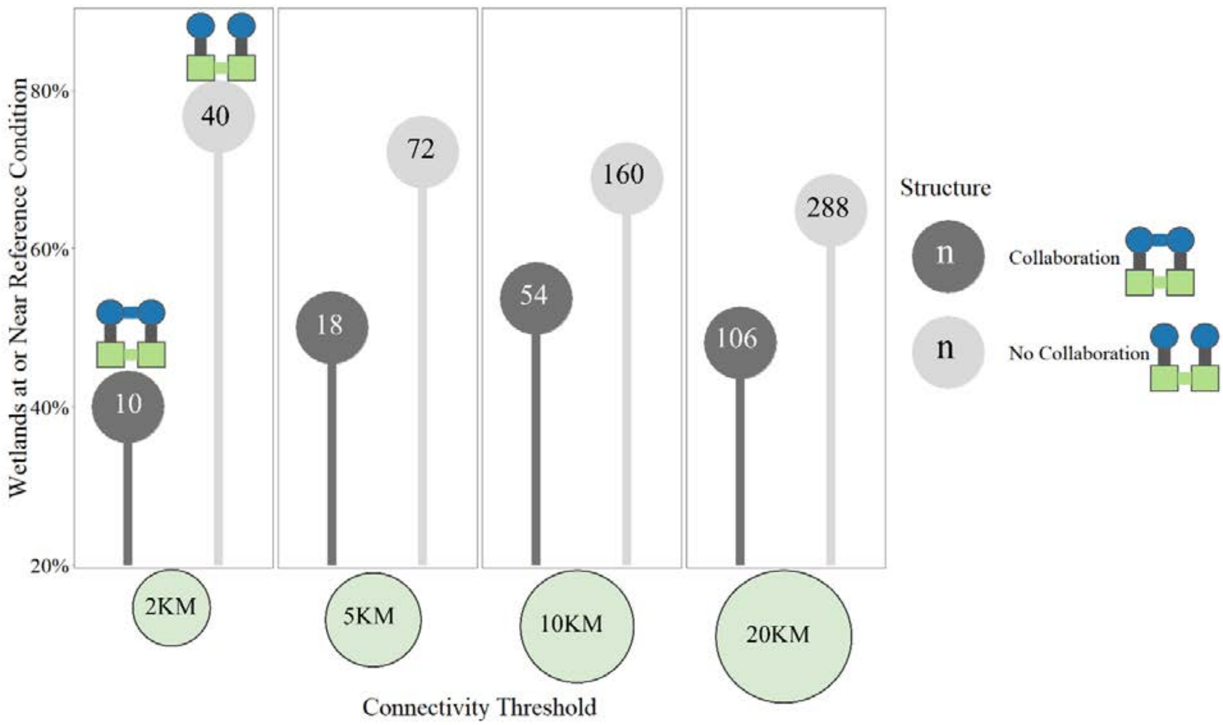


Figure 1.3. Change in the percentage of wetlands at or near a reference condition in substructures 2a and 2b at increasing connectivity thresholds. The numbers inside the grey circles show the number of substructures which occur at each given threshold.

<u>Social Network Findings</u>

Given that organizations collaborate largely on the basis of proximity, we would expect that the social network of wetland management organizations in the state would be highly modular based on region. We assessed the social network modularity as well as the role each node plays in overall network connectivity.

<u>Whole Network Findings</u>

The random walk algorithm showed that the social network is non-modular (i.e. resulting modularity estimate was 0). This result suggests that the peripheral organizations are all connected to one primary core of key organizations.

To further explore this result, we tested a k-core decomposition algorithm on the social network to identify if a core truly exists. The social network produced an optimal core with a k of 10 and 22 nodes, meaning that there are 22 interconnected core nodes with at least 10 connections to each other (Fig. 4). This result reinforces the conclusion that the social network has one cohesive core and is not modular. This is in contrast to what we would expect given the social-ecological network outputs.

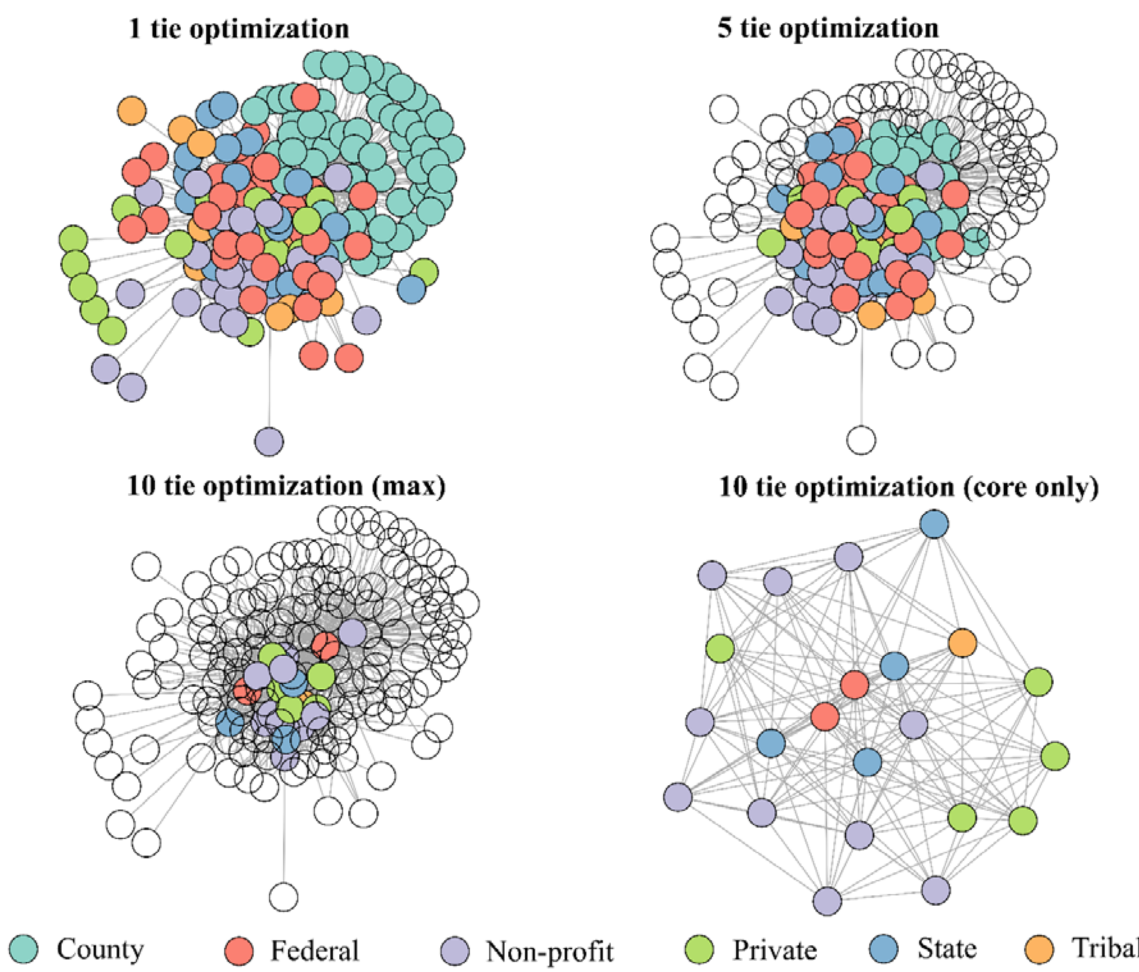


Figure 1.4. Results from the k-core decomposition algorithm in the social network of Montana wetland management organizations. In the first three panels, organizations become transparent when they are no longer have the required number of ties (1, 5, 10). The fourth panel shows just the optimal core with each organization optimized to have 10 ties.

Node Specific Findings

To understand how a non-modular, core periphery network can result from independent organizations primarily collaborating based on proximity, we assessed the bridging role of each individual organization (Fig. 5). To do this, we measured the betweenness centrality (number of times the shortest path between any given pair of organizations goes through that organization) of each organization in the sample. Results from this analysis showed that just two organizations are responsible for the cohesive and efficient structure of information sharing among wetland management organizations in Montana. The vast majority of wetland management organizations play little to no bridging role within the social network, i.e. they are never or only very rarely on the most direct path between any given pair of organizations in the network. The top two bridging organizations have a betweenness centrality of 2,465 & 4,935. Given that in this network there are 14,196 unique pairs of organizations, this means that ~35% & 17% of all possible communications go through the top two bridging organizations respectively. When we remove either of these organizations individually, and rerun the random walk algorithm testing for modularity, we continue to see a non-modular network (modularity of 0). In contrast, when we remove *both* of the top bridging nodes, our resulting modularity of information flow is 3. This suggests that the core of the wetland management network in Montana is resilient to removal of either of the two key collaborative organizations, but not both.

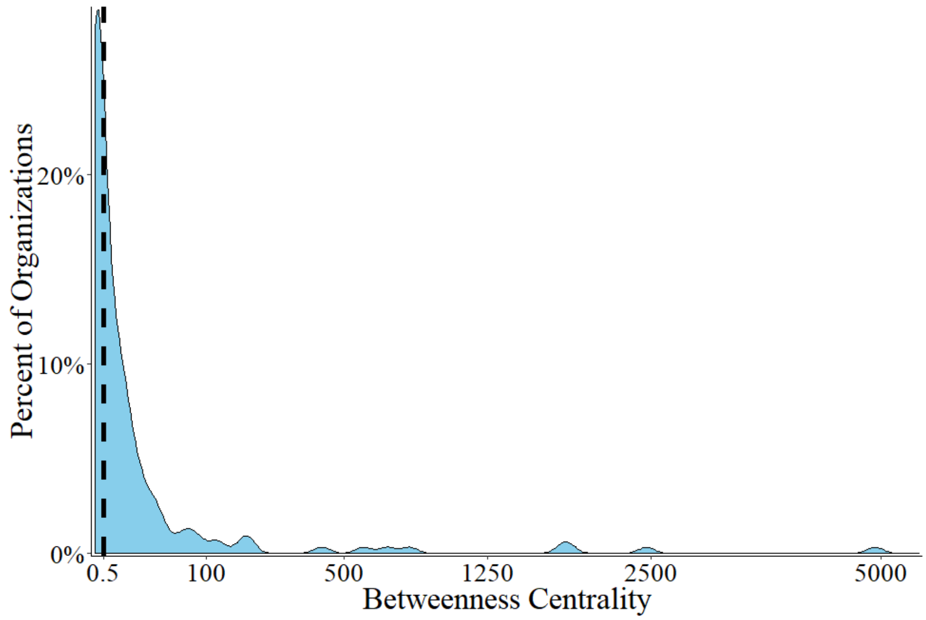


Figure 1.5. Density of betweenness centrality of the observed social network of wetland management organizations in Montana. The X axis is on the square root scale to maximize the amount of information displayed. The black dashed line represents the median betweenness centrality of observed social nodes.

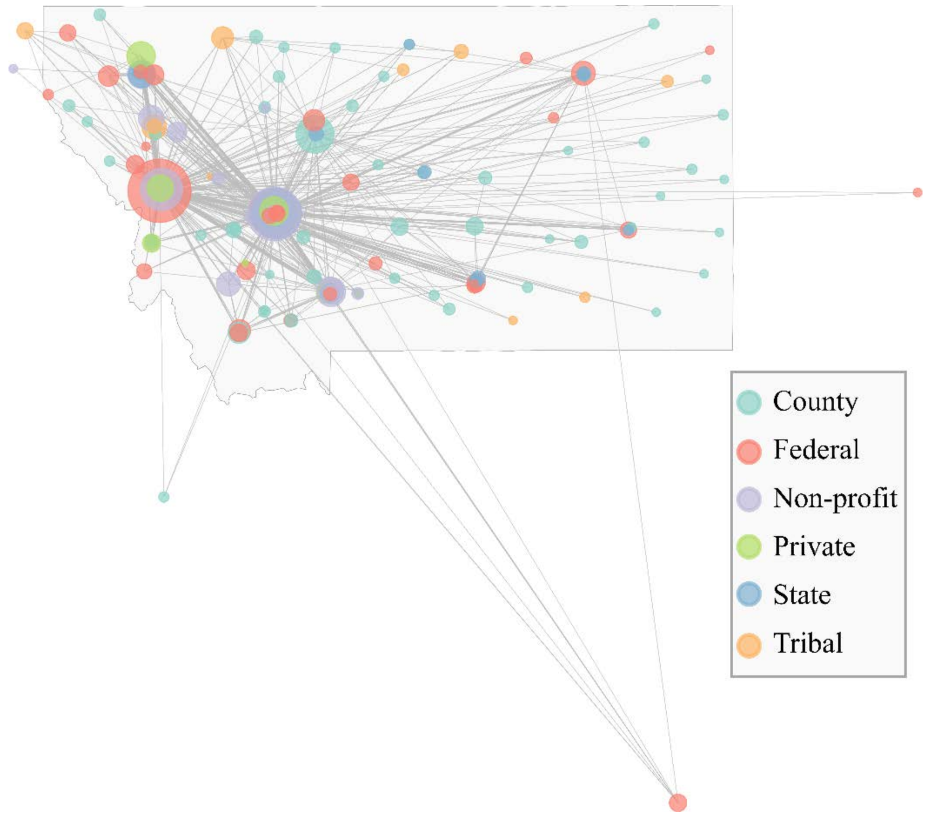


Figure 1.6. Observed social network of wetland management organizations in Montana. The node size is a function of the number of collaborations each organization has with others (degree). Exact office locations have been slightly adjusted to protect the identity of survey respondents.

Discussion

Our results from the social-ecological analyses for building blocks 1 & 2 show that wetland management organizations in Montana collaborate less readily than we would expect given stochastic network formation. Where collaborations are present, we illustrate that environmental variables (location & condition) are associated with, and to some extent likely dictate the structure of collaboration among managers. Given that proximity appears to be a strong indicator of collaboration (i.e. organizations tend to collaborate with other organizations who have projects close to theirs), we would expect the overall social network to be modular based on region. Highly modular networks are inefficient for complex problem solving and could result in less-than-optimal environmental outcomes. When we further examine the social network of wetland management organizations, we find a core periphery network structure. Core periphery, or non-modular networks, are associated with rapid diffusion of useful information and efficient complex problem solving (Mason & Watts 2012).

When we examine the role that individual organizations play in the overall collaborative network, we find that just two key (highest betweenness centrality) organizations are responsible for the coherence of the social network. We assume that cohesive management of ecological resources, notably highly connected resources such as wetlands, at the landscape scale should be a primary goal for all large scale resource management plans. This goal can be difficult to accomplish given the inconsistencies between management jurisdiction, the costs of collaboration, and varying management goals. Yet, with this in mind, we couple established methods and an emerging frontier in network science to show that just a small number of organizations willing to bear the burden of collaboration can facilitate cohesive management at a landscape scale.

This paper is not intended to make a strong statement specifically about wetland management in Montana or make prescriptions, calls to action etc. for wetland managers in the state. In this study, we aim to advance the burgeoning field of social-ecological network analysis by showing the utility of variable connectivity thresholds, incorporating node level measures of ecological condition, and demonstrating how measures of information diffusion and complex problem solving within the social nework can be used to further explore and substantiate findings from this emerging field. We also show that the ratio of network substructures, or building blocks is variable based on the defined ecological connectivity threshold. Because it is commonplace to set just one threshold in social-ecological network studies, this introduces a significant source of bias for this body of literature. We use this paper to caution against setting single ecological connectivity thresholds in future research and instead using variable or more advances measures of connectivity.

Constraints

A significant constraint in this study and with much survey-based research generally is the reliability of self-reported data. Self-reported survey data is known to have significant biases in terms of time, favoritism, self-image, etc. (Bound, Brown & Mathiowetz 2001). In addition to this limitation, we were also unable to survey the entire social network of wetland managers in Montana. While a strength of network science is the ability for each individual unit of analysis to be understood and influential, network studies are known to be highly influenced by incomplete sampling (Kossinets 2006). In this study, we show the influence that just a few nodes can have on network structure. For this reason, the incomplete sampling of the social network poses a significant limitation for the real-world implications of this research.

Future Research

We propose that future research into this specific study system would benefit from more robust measures of social connectivity and environmental condition. Leveraging data on collaborative interactions such as email correspondence or co-authorship on projects would provide a more empirical measure of collaboration compared to self-reporting. Researchers could also use a more robust measure of ecological condition such as floristic quality indexes or remotely sensed data.

We also urge the production of methods based research and tool development for multilevel network analysis and for estimating node characteristics as a function of network structure. One promising avenue for this is the advancement of auto-logistic actor attribute models (Lusher, Koskinen & Robins 2013). Increasing the usability of auto-logistic actor attribute models will allow future research to estimate the effect size of specific network building blocks on nodes within them; this method is similar to a linear modeling framework, while acknowledging the lack of independence in network data.

Conclusions

Social-ecological network analysis is a growing field with innumerable possible trajectories for future research. We build upon the current frameworks for operationalizing these networks to show that just two organizations willing to bear the burden of collaboration can facilitate cohesive management of connected resources at a state-wide scale. Alongside this empirical study, we explore a gradient of ecological connectivity thresholds to build a dynamic understanding of the role of connectivity in the two level system. We observed variable results based the gradient of connectivity thresholds, which leads us to warn against arbitrary thresholds of ecological connectivity in future social-ecological network studies as they may bias findings. Lastly, we employ traditional methods in social network analysis to further explore the social component of our two level network, showing the utility of these well-established methods to bolster social-ecological network findings. While the information presented in this study can surely be of use for informing wetland management practices in Montana, U.S., we want to make clear the constraints of this research due to data availability and emphasize the methodological advances made in this research for future social-ecological network studies and for natural resource management research broadly.

CHAPTER ONE REFERENCES

Andrews, R. N. L. (2006). Managing the Environment, Managing Ourselves: A History of American Environmental Policy, Second Edition. Yale University Press.

Baggio, J. A., & Hillis, V. (2018). Managing ecological disturbances: Learning and the structure of social-ecological networks. *Environmental Modelling & Software*, *109*, 32–40. https://doi.org/10.1016/j.envsoft.2018.08.002

Barnes, M. L., Bodin, Ö., McClanahan, T. R., Kittinger, J. N., Hoey, A. S., Gaoue, O. G., & Graham, N. A. J. (2019). Social-ecological alignment and ecological conditions in coral reefs. *Nature Communications*, *10*(1), 2039. https://doi.org/10.1038/s41467-019-09994-1

Batagelj, V., & Zaverˇsnik, M. (n.d.). An O(m) Algorithm for Cores Decomposition of Networks. 9.

Berardo, R. (2014). Bridging and Bonding Capital in Two-Mode Collaboration Networks. *Policy Studies Journal*, *42*(2), 197–225. https://doi.org/10.1111/psj.12056

Bodin, Ö. (2017). Collaborative environmental governance: Achieving collective action in social-ecological systems. *Science*, *357*(6352), eaan1114. https://doi.org/10.1126/science.aan1114

Bodin, Ö., Alexander, S. M., Baggio, J., Barnes, M. L., Berardo, R., Cumming, G. S., … Sayles, J. S. (2019). Improving network approaches to the study of complex social–ecological interdependencies. *Nature Sustainability*, 1. https://doi.org/10.1038/s41893-019-0308-0

Bodin, Ö., Crona, B., Thyresson, M., Golz, A.-L., & Tengö, M. (2014). Conservation Success as a Function of Good Alignment of Social and Ecological Structures and Processes. *Conservation Biology*, *28*(5), 1371–1379. https://doi.org/10.1111/cobi.12306

Bodin, Ö., Robins, G., McAllister, R., Guerrero, A., Crona, B., Tengö, M., & Lubell, M. (2016). Theorizing benefits and constraints in collaborative environmental governance: a transdisciplinary social-ecological network approach for empirical investigations. *Ecology and Society*, *21*(1). https://doi.org/10.5751/ES-08368-210140

Bodin, Ö., & Tengö, M. (2012). Disentangling intangible social–ecological systems. *Global Environmental Change*, *22*(2), 430–439. https://doi.org/10.1016/j.gloenvcha.2012.01.005

Bound, J., Brown, C., & Mathiowetz, N. (2001). Chapter 59 - Measurement Error in Survey Data. In J. J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3705–3843). https://doi.org/10.1016/S1573-4412(01)05012-7

Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: A conservation application of graph theory. *Journal of Environmental Management*, *59*(4), 265–278. https://doi.org/10.1006/jema.2000.0373

Cadenasso, M. L., Pickett, S. T. A., Weathers, K. C., & Jones, C. G. (2003). A Framework for a Theory of Ecological Boundaries. *BioScience*, *53*(8), 750–758. https://doi.org/10.1641/0006-3568(2003)053[0750:AFFATO]2.0.CO;2

Calhoun, A. J. K., Mushet, D. M., Alexander, L. C., DeKeyser, E. S., Fowler, L., Lane, C. R., … Walls, S. C. (2017). The Significant Surface-Water Connectivity of “Geographically Isolated Wetlands.” *Wetlands*, *37*(4), 801–806. https://doi.org/10.1007/s13157-017-0887-3

Cook, B. J., & Hauer, F. R. (2007). Effects of hydrologic connectivity on water chemistry, soils, and vegetation structure and function in an intermontane depressional wetland landscape. *Wetlands*, *27*(3), 719–738. https://doi.org/10.1672/0277-5212(2007)27[719:EOHCOW]2.0.CO;2

Cowling, R. M., Egoh, B., Knight, A. T., O'Farrell, P. J., Reyers, B., Rouget, M., … Wilhelm-Rechman, A. (2008). An operational model for mainstreaming ecosystem services for implementation. *Proceedings of the National Academy of Sciences*, *105*(28), 9483–9488. https://doi.org/10.1073/pnas.0706559105

Crona, B., & Hubacek, K. (2010). The Right Connections: How do Social Networks Lubricate the Machinery of Natural Resource Governance? *Ecology and Society*, *15*(4). https://doi.org/10.5751/ES-03731-150418

Csardi G & Nepusz T: The igraph software package for complex network research, InterJournal, Complex Systems 1695. 2006. http://igraph.org

Dallimer, M., & Strange, N. (2015). Why socio-political borders and boundaries matter in conservation. *Trends in Ecology & Evolution*, *30*(3), 132–139. https://doi.org/10.1016/j.tree.2014.12.004

de Groot, R. S., Alkemade, R., Braat, L., Hein, L., & Willemen, L. (2010). Challenges in integrating the concept of ecosystem services and values in landscape planning, management and decision making. *Ecological Complexity*, *7*(3), 260–272. https://doi.org/10.1016/j.ecocom.2009.10.006

Frank, O., & Strauss, D. (1986). Markov Graphs. *Journal of the American Statistical Association*, *81*(395), 832–842. https://doi.org/10.1080/01621459.1986.10478342

Geys, B., & Murdoch, Z. (2010). Measuring the 'Bridging' versus 'Bonding' Nature of Social Networks: A Proposal for Integrating Existing Measures. *Sociology*, *44*(3), 523–540. https://doi.org/10.1177/0038038510362474

Guerrero, A., Bodin, Ö., McAllister, R., & Wilson, K. (2015). Achieving social-ecological fit through bottom-up collaborative governance: an empirical investigation. *Ecology and Society*, *20*(4). https://doi.org/10.5751/ES-08035-200441

Hamilton, M., Fischer, A. P., & Ager, A. (2019). A social-ecological network approach for understanding wildfire risk governance. *Global Environmental Change*, *54*, 113–123. https://doi.org/10.1016/j.gloenvcha.2018.11.007

Hein, L., van Koppen, K., de Groot, R. S., & van Ierland, E. C. (2006). Spatial scales, stakeholders and the valuation of ecosystem services. *Ecological Economics*, *57*(2), 209–228. https://doi.org/10.1016/j.ecolecon.2005.04.005

Houlahan, J. E., Keddy, P. A., Makkay, K., & Findlay, C. S. (2006). The effects of adjacent land use on wetland species richness and community composition. *Wetlands*, *26*(1), 79–96. https://doi.org/10.1672/0277-5212(2006)26[79:TEOALU]2.0.CO;2

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Kareiva, P., & Wennergren, U. (1995). Connecting landscape patterns to ecosystem and population processes. *Nature*, *373*(6512), 299. https://doi.org/10.1038/373299a0

Kauffman, G. J. (2002). What if… the United States of America were based on watersheds? *Water Policy*, *4*(1), 57–68. https://doi.org/10.1016/S1366-7017(02)00019-3

Kininmonth, S., Bergsten, A., & Bodin, Ö. (2015). Closing the collaborative gap: Aligning social and ecological connectivity for better management of interconnected wetlands. *Ambio*, *44*(Suppl 1), 138–148. https://doi.org/10.1007/s13280-014-0605-9

Koontz, T. M., & Thomas, C. W. (2006). What do we know and need to know about the environmental outcomes of collaborative management?. *Public administration review*, *66*, 111-121.

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks*, *28*(3), 247–268. https://doi.org/10.1016/j.socnet.2005.07.002

Laita, A., Kotiaho, J. S., & Mönkkönen, M. (2011). Graph-theoretic connectivity measures: what do they tell us about connectivity? *Landscape Ecology*, *26*(7), 951–967. https://doi.org/10.1007/s10980-011-9620-4

Leibowitz, S. G., Wigington, P. J., Rains, M. C., & Downing, D. M. (2008). Non-navigable streams and adjacent wetlands: addressing science needs following the Supreme Court's Rapanos decision. *Frontiers in Ecology and the Environment*, *6*(7), 364–371. https://doi.org/10.1890/070068

Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2013). Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press.

March, J. G. (1991). Exploration and Exploitation in Organizational Learning. *Organization Science*, *2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

Mason, W., & Watts, D. J. (2012). Collaborative learning in networks. *Proceedings of the National Academy of Sciences*, *109*(3), 764–769. https://doi.org/10.1073/pnas.1110069108

McRae, B. H., Hall, S. A., Beier, P., & Theobald, D. M. (2012). Where to Restore Ecological Connectivity? Detecting Barriers and Quantifying Restoration Benefits. *PLOS ONE*, *7*(12), e52604. https://doi.org/10.1371/journal.pone.0052604

Miller, K. D., Zhao, M., & Calantone, R. J. (2006). Adding Interpersonal Learning and Tacit Knowledge to March's Exploration-Exploitation Model. *Academy of Management Journal*, *49*(4), 709–722. https://doi.org/10.5465/amj.2006.22083027

Montana Department of Environmental Quality (2013). Priceless Resources: A Strategic Framework for Wetland and Riparian Area Conservation and Restoration in Montana, 2013 - 2017. Helena, Montana. https://deq.mt.gov/Portals/112/Water/WPB/Wetlands/StategicFramework2013-2017.pdf

Montana Wetland and Riparian Mapping Center (2019). Montana Natural Heritage Program, Montana State Library. Helena, Montana. http://mtnhp.org/default.asp

Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

Nkhata, A., Breen, C., & Freimund, W. (2008). Resilient Social Relationships and Collaboration in the Management of Social–Ecological Systems. *Ecology and Society*, *13*(1). https://doi.org/10.5751/ES-02164-130102

Qualtrics (2017). Provo, UT, USA. https://www.qualtrics.com

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ricketts, T. H. (2001). The Matrix Matters: Effective Isolation in Fragmented Landscapes. *The American Naturalist*, *158*(1), 87–99. https://doi.org/10.1086/320863

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, *178*(1), 13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, *105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Sayles, J. S., & Baggio, J. A. (2017). Social–ecological network analysis of scale mismatches in estuary watershed restoration. *Proceedings of the National Academy of Sciences*, *114*(10), E1776–E1785. https://doi.org/10.1073/pnas.1604405114

Scott, T. (2015). Does Collaboration Make Any Difference? Linking Collaborative Governance to Environmental Outcomes. *Journal of Policy Analysis and Management*, *34*(3), 537–566. https://doi.org/10.1002/pam.21836

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks*, *5*(3), 269–287. https://doi.org/10.1016/0378-8733(83)90028-X

Tassier, T. (2013). *The Economics of Epidemiology*. Springer Science & Business Media.

Treml, E. A., Fidelman, P. I. J., Kininmonth, S., Ekstrom, J. A., & Bodin, Ö. (2015). Analyzing the (mis)fit between the institutional and ecological networks of the Indo-West Pacific. *Global Environmental Change*, *31*, 263–271. https://doi.org/10.1016/j.gloenvcha.2015.01.012

Wang, P., Robins, G., Pattison, P. (2009) PNet: program for the simulation and estimation of exponential random graph models. Melbourne School of Psychological Sciences, The University of Melbourne.

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks*, *35*(1), 96–115.

Wolf, K. L., Noe, G. B., & Ahn, C. (2013). Hydrologic Connectivity to Streams Increases Nitrogen and Phosphorus Inputs and Cycling in Soils of Created and Natural Floodplain Wetlands. *Journal of Environmental Quality*, *42*(4), 1245–1255. https://doi.org/10.2134/jeq2012.0466

Yarrow, M. M., & Marín, V. H. (2007). Toward Conceptual Cohesiveness: a Historical Analysis of the Theory and Utility of Ecological Boundaries and Transition Zones. *Ecosystems*, *10*(3), 462–476. https://doi.org/10.1007/s10021-007-9036-9

CHAPTER TWO: BRINGING FORECASTING INTO THE FUTURE: USING GOOGLE TO PREDICT VISITATION TO U.S. NATIONAL PARKS

Abstract

In recent years, visitation to U.S. National Parks has been increasing, with the majority of this increase occurring in a subset of parks. As a result, managers in these parks must respond quickly to increasing visitor-related challenges. Improved visitation forecasting would allow managers to more proactively plan for such increases. In this study, we leverage internet search data that is freely available through Google Trends to create a forecasting model. We compare this Google Trends model to a traditional autoregressive forecasting model. Overall, our Google Trends model accurately predicted 97% of the total visitation variation to all parks one year in advance from 2013-2017 and outperformed the autoregressive model by all metrics. While our Google Trends model performs better overall, this was not the case for each park unit individually; the accuracy of this model varied significantly from park to park. We hypothesized that park attributes related to trip planning would correlate with the accuracy of our Google Trends model, but none of the variables tested produced overly compelling results. Future research can continue exploring the utility of Google Trends to forecast visitor use in protected areas, or use methods demonstrated in this paper to explore alternative data sources to improve visitation forecasting in U.S. National Parks.

Introduction

Visitation to parks and protected areas benefits human health, local and national economies, and promotes pro-conservation behavior (Cullinane Thomas, Koontz, & Cornachione, 2018; Halpenny, 2010; Maller, Townsend, Pryor, Brown, & St Leger, 2006; Maples, Sharp, Clark, Gerlaugh, & Gillespie, 2017). In 2017, the United States National Park Service (NPS) broadly contributed an estimated 306,000 jobs and $35.8 billion in direct economic output; visitor spending specifically contributed to an estimated 188,600 jobs and $14.4 billion in economic output, and visitors spent an estimated $18.2 billion in local gateway regions (Cullinane Thomas et al., 2018). But while park visitation leads to positive outcomes for humans and economies, some argue that too many people are “loving parks to death” (e.g., Daysog, 2018; Duncan, 2016; Simmonds et al., 2018). Large numbers of visitors can stress natural, cultural, and human resources, and lead to a decrease in the quality of visitor experiences (Graefe, Vaske, & Kuss, 1984; Hallo & Manning, 2010; Marion, Leung, Eagleston, & Burroughs, 2016). Additionally, legal standards may be violated under rapid visitation growth scenarios. The NPS is required to identify the maximum number of visitors an area can hold without causing resource damage, and to manage visitation at or below this capacity (Cahill, Collins, McPartland, Pitt, & Verbos, 2018), but unpredictable increases in visitation may limit mangers’ ability to adhere to these standards under changing conditions. One notable example of rapid visitation increase can be seen in Joshua Tree National Park (Fig. 2.1) starting in 2013. In 2017, 61 of 417 areas managed by the NPS set a new record for visitation. Forty-two of these areas broke a record high set in just 2016, and between 2012 and 2017 visitation to the NPS overall grew by 17% (National Park Service, 2018c;

Ziesler & Singh, 2018). Throughout the paper we refer to all areas managed by the National Park Service (national parks, national battlefields, national memorials, etc.) as NPS units. Without forewarning and sufficient time to prepare, a dramatic increase in visitation at an individual national park unit may necessitate that staff address only the most pressing needs, at the expense of long-term planning.

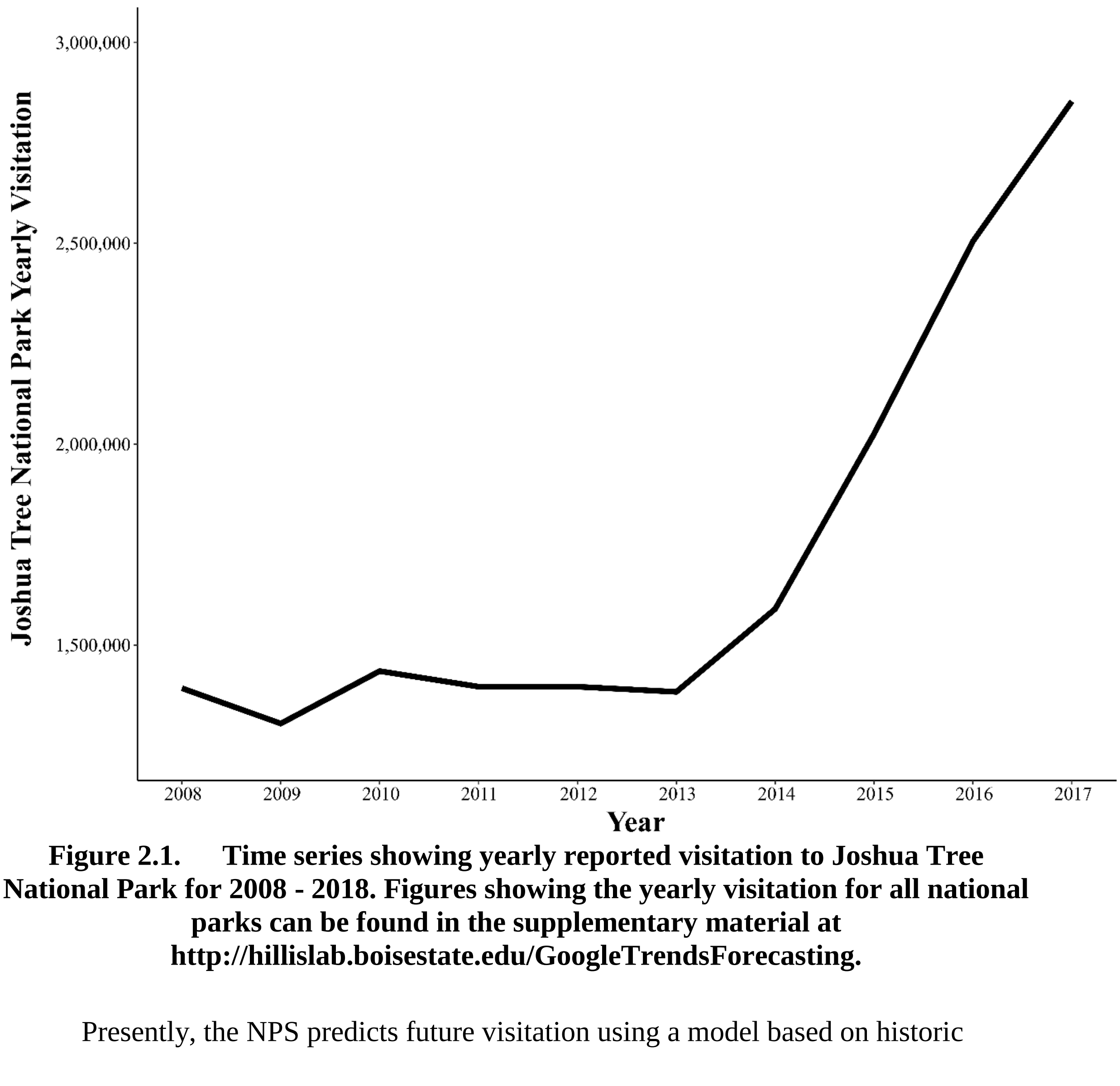


Figure 2.1. Time series showing yearly reported visitation to Joshua Tree National Park for 2008 - 2018. Figures showing the yearly visitation for all national parks can be found in the supplementary material at http://hillislab.boisestate.edu/GoogleTrendsForecasting.

Presently, the NPS predicts future visitation using a model based on historic visitation from the previous five years (Ziesler, 2016). While past visitation may be a reasonably accurate predictor of future visitation, these models, often referred to as autoregressive, do not account for outside factors, such as the overall state of the economy or news & social media attention (Wilmot & McIntosh, 2014). Additionally, events such as hurricanes and eclipses influence visitation and are not correlated with the previous year's visitation (Ziesler & Singh, 2018). Managers would benefit from having a more accurate method for predicting future visitation quickly and comprehensively. Improved forecasting ability could help managers better understand trends in future visitation. For example, managers could assess whether a recent spike in visitation is a new baseline, a unique anomaly, or whether visitation will continue to increase. Finally, predicting visitation can help determine which management actions park officials should consider and implement.

While improved forecasting ability would enable managers to mitigate impacts of rapidly increasing visitation, it is important to recognize that limited financial or staff capacity could inhibit managers' access to collecting new data. Therefore, there is a need to explore how existing data sources can be utilized, especially those that are cheap, relatively easy to analyze, and can be collected at any time. Open-source digital data, such as those reported through Google Trends, are relatively effortless to collect and represent an opportunity for park managers to make use of search engine data. Mining digital data can be especially useful because, by analyzing the records that visitors leave behind online, it may be possible to predict changes in rates of visitation that are not captured by the current autoregressive model.

Overall, the goal of this research is not to identify the absolute best forecasting model for each and every national park unit, but rather to explore the use of easily accessible search engine data and test an alternative forecasting model which can be applied to all parks and protected areas in general. To do this, we analyzed Google Trends data for its predictive ability across U.S. National Parks; we did not include other units managed by the National Park Service such as national monuments, historic sites, etc. The specific objectives of this study are to: (1) investigate whether Google Trends is useful for predicting future visitation to U.S. National Parks as compared to an autoregressive model, and (2) explore explanations for the discrepancy in model efficacy between parks. We hypothesized that the utility of Google Trends as a predictor would not be uniform across all parks. Specifically, we speculated that our ability to use Google Trends to forecast park visitation may be affected by the proportion of people who plan their visits to each park well in advance (e.g., the previous year), operationalized as the population surrounding each park and park popularity.

Literature Review

A majority of Americans (86%) use general search engines such as Google to plan travel (Fesenmaier, Xiang, Pan, & Law, 2011). Additionally, 65% said that general search engines were very useful or essential for planning a trip (Fesenmaier et al., 2011). Given that such a high percentage of people use general search engines to plan travel, researchers have started exploring the feasibility of using search engine data to forecast tourism arrivals (e.g. Bangwayo-Skeete & Skeete, 2015; Dergiades, Mavragani, & Pan, 2018; Yang, Pan, Evans, & Lv, 2015). However, no previous study has explored using Google Trends to predict visitation to parks or protected areas. Other sources of publically available online data, such as social media, have been useful for exploring visitation to public lands (Sessions, Wood, Rabotyagov, & Fisher, 2016; Tenkanen et al., 2017; Wood, Guerry, Silver, & Lacayo, 2013). However, obtaining data from social media sites can be time-intensive and currently requires knowledge of how to interact with application programming interfaces (APIs). Additionally, many social media sites are now restricting access to their data. Since many public lands managers may not have time, knowledge, or access to gather this data, we explore the usability of Google Trends, which is easy and free for anyone to download.

Previous studies have explored the utility of using Google Trends to forecast a range of social phenomena, including flu-related emergency room visits, cinema admissions, private consumption, and tourist demand (Araz, Bentley, & Muelleman, 2014; Hand & Judge, 2012; Önder & Gunter, 2016; Vosen & Schmidt, 2011). Search engine data has numerous advantages, including the ability to track preferences in real time and providing a high frequency of data (Yang et al., 2015). In one of the earliest studies investigating the utility of Google Trends, Choi and Varian (2012) found that Google Trends was useful for predicting present conditions in a variety of contexts, such as sales of motor vehicles and parts, claims for unemployment, and predicting visitors to Hong Kong. However, the authors state that more research is needed to explore whether this data would be useful for making future projections (Choi & Varian, 2012).

After Choi and Varian's initial finding that Google Trends may be useful for tourism, more researchers started to explore ways to use this data. Bangwayo-Skeete and Skeete (2015) tested whether Google search data can predict visitor arrivals at popular tourist destinations in the Caribbean Islands, and found that Google search data significantly improved the ability of models to forecast future visitation. Additionally, Li, Pan, Law, and Huang (2017) found that using a search index to forecast future tourism demand in Beijing was more accurate than traditional models using past visitation alone.

Park, Lee, and Song (2017) also found that models using Google Trends to forecast short-term tourism inflows to South Korea performed better than traditional time-series models. However, Dergiades et al. (2018) noted that using search engine data to forecast tourism is often filled with language and platform bias, particularly for destinations that have many international visitors. Not all visitors use the same search engines or search for things in the same languages.

This body of literature shows that search engine data can be highly useful for forecasting tourism demand. However, it is uncertain how well this data can predict visitation to parks and protected areas specifically. These visitors may have different search habits than visitors to big cities or hotels. Google Trends data has the potential to improve current visitation forecasting methods by capturing trends in social media, news media, and other cultural or social shifts that influence public desire to plan and subsequently visit any given park unit. Google Trends therefore may represent the culmination of these various social phenomena, but further research is necessary to better understand the utility of this emerging tool.

Methodology

Study Sites

The U.S. National Park Service (NPS) has 60 units designated as National Parks. Two of these sites were not included in this study because of their recent designations (Pinnacles and Gateway Arch, which were designated in 2013 and 2018 respectively). The relatively new designations did not allow enough historical data for modeling. One site, National Park of American Samoa, does not have visitation data for 2008 – 2010, and was therefore also not included in this study. The 57 parks studied collectively had

85.2 million visits in 2017 (National Park Service, 2018b). National Parks were chosen as opposed to other units managed by the National Park Service because they have the most reliable visitation data, the highest numbers of visitors, the highest economic and cultural impact, and have seen unprecedented visitation changes in recent years (Ziesler & Singh, 2018).

Data Collection

All data used in this paper is readily available through an open source application found here: http://hillislab.boisestate.edu/GoogleTrendsForecasting/. This application was created using the ‘shiny’ package for the ‘R’ statistical platform (Chang, Cheng, Allaire, Xie, & McPherson, 2018).

Park Visitation

We retrieved data on historic park visitation from the National Park Visitor Use Statistics Portal (National Park Service, 2018c). Methods for collecting these data generally include the use of car counters, concessioner reports, and permit information, but are specific to each NPS unit. Unit-specific protocols can be found on the NPS Visitor Use Statistics website (https://irma.nps.gov/Stats/) (Ziesler & Singh, 2018).We downloaded monthly visitation data for each of the 57 U.S. National Parks from 2006 – 2017; we then summed all months into yearly counts to avoid confounding seasonal variation and increase the interpretability of this research. Although we believe some reported visitation counts may be erroneous (e.g. “0”), we took all data as is.

Google Trends

We downloaded search history data for each national park individually from 2007 – 2017 using the Google Trends interface, which can be accessed at https://trends.google.com/trends/. These data are reported and were downloaded at the monthly scale for each park. For most search terms, data is available from 2004 – present. In order to complete the search instantly, Google analyzes a sample of the total volume of searches and the data is then indexed from 0 to 100, where 100 is the highest volume of searches for the selected range. A value of 50 indicates there are half as many searches for the term that month compared to the month indexed at 100. In summary, the indexed Google Trends data represents the total number of people searching for the specified term, compared to the total volume of searches in the selected area, scaled such that the highest value in the selected time frame is set to 100.

Google Trends provides the option to track either search terms or topics. While search terms represent only those who type in the exact phrase in a specified language, topics represent anyone searching for the specified concept, in any language. We therefore used topics rather than search terms due to the ability to capture a broader array of searches in other languages and reduce bias. We also set Google Trends to provide data based on worldwide searches, since many U.S. National Parks host international visitors.

Spatial Data

We downloaded two sets of spatial data for this study to explore our second research question. The first dataset included shapefiles of the locations of each national park in the U.S., which we downloaded from the NPS (National Park Service, 2018a).

We also downloaded 2010 U.S. census block data from ESRI Data & Maps (ESRI, 2018).

Data Analysis

Modeling

In this study, we created an autoregressive model to compare against our predictions using Google Trends values alone. We created our own autoregressive model, rather than comparing our projections to those of the National Park Service, to establish that the variation in model accuracies are a result of the predictive variable (Google Trends vs. past visitation), rather than statistical methods. By creating our own autoregressive model, we can ensure that we are comparing parallel methodologies and achieving the greatest level of interpretability and contrast between the two models. Our autoregressive model predicts the expected visitation for each specific park for a given year (yi) based on the visitation to that specific park from the five previous years:

$$X_{\text{Vis t-1}}, X_{\text{Vis t-2}}, X_{\text{Vis t-3}}, X_{\text{Vis t-4}}, X_{\text{Vis t-5}}$$

We chose a 5-year autoregressive interval because this is the interval used by the National Park Service for forecasting, although they use a simple trend line extension based on the last 5 years of visitation (Ziesler, 2016). We used a hierarchical model structure to allow each park to retain its own intercept in the equation ($\beta 0_{\text{Park[i]}}$). We fit this model to a negative binomial distribution in a Bayesian framework. We chose a negative binomial distribution as opposed to a Poisson distribution for these models because the negative binomial distribution includes a term (ϕ) to account for overdispersion, or high amounts of variability between parks (Gardner, Mulvey, & Shaw, 1995). We constructed these models with the ‘rstanarm’ package in the R statistical programming language (Goodrich, Gabry, Ali, & Brilleman, 2018). A Bayesian model is preferred to a frequentist model in this situation because it offers greater flexibility when assessing predictor and outcome variables which are on considerably different scales (e.g. Google Trends values and park visitation) (Clark, 2005).

$$yi \sim NB(\mu i, \phi)$$

$$\log(\mu i) = \beta 0 + \beta 0_{Park[i]} + \beta 1 * X_{Vis\ t-1} + \beta 2 * X_{Vis\ t-2} + \beta 3 * X_{Vis\ t-3} + \beta 4 * X_{Vis\ t-4} + \beta 5 * X_{Vis\ t-5}$$

Our Google Trends model has a similar overall structure, although it uses a specific Google Trends parameter, or slope estimate for each park ($\beta 1_{Park[i]}$) to predict visitation, and is informed by the sum of the Google Trends values for each park one year previous to the year being predicted (X_{Google}), rather than by previous visitation.

$$yi \sim NB(\mu i, \phi)$$

$$\log(\mu i) = \beta 0 + \beta 0_{Park[i]} + \beta 1 * X_{Google} + \beta 1_{Park[i]} * X_{Google}$$

Both the autoregressive and Google Trends models predict park visitation on the annual scale, one year in advance. For example, when we are predicting visitation for 2015, we are only using visitation through 2014 and Google Trends values through 2014 for the autoregressive and Google Trends models respectively.

For both models, we used the default weakly informative prior distributions in the ‘rstanarm’ package (Goodrich et al., 2018). The default priors for both the intercept and all coefficients, are normally centered at 0, with a standard deviation of 10 and 2.5 for the intercept and coefficients respectively. The default weakly informative error standard deviation or “sigma” is exponential. These prior distributions were chosen because they are extremely conservative. The package automatically rescales these priors if necessary to match the order of magnitude of the data. Our autoregressive model did not require any rescaling, so the default priors were kept. The Google Trends model rescaled the standard deviation of our Google Trends coefficient only; the rescaled standard deviation was 0.017. Both models showed adequate mixing and Markov Chain convergence.

Validation

To assess the out-of-sample predictive ability of both models, we blocked all data from 2013 - 2017 by year so that each block contains the data for all parks for that year. We then used all data prior to that year to inform or "train" predictions for that block. As we progressed through the blocks, we included blocks prior to the year being predicted or "tested." (Fig. 2.2). This procedure is often called cross-validation on a rolling basis. We chose to validate our models in this way because it allowed us to make use out of all available data, while not informing any predictions based on present or future data (Bergmeir & Benítez, 2012). It is in this same vein that we blocked our data by entire years, as opposed to by both park and year. This prevented the models from using any present or future data, even those from other parks.

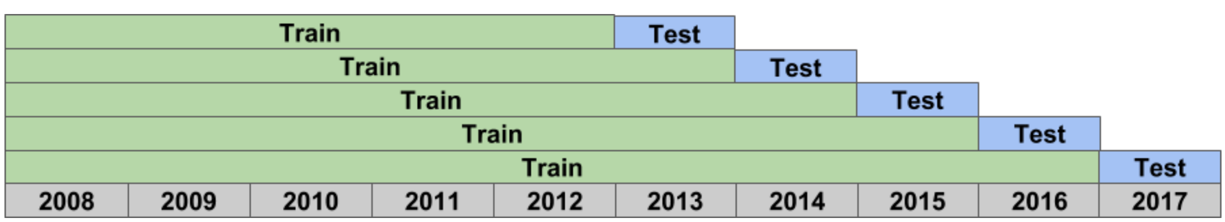


Figure 2.2. Our implementation of cross-validation on a rolling basis.

Error

We specified our models to yield 2,000 visitation predictions for each park, for each year. We took the median of these predictions as our projected visitation forecast. All error metrics were calculated based on these median predictions compared to the observed visitation for each park. We chose to use three different metrics to test the accuracy of our median predictions. These included R^2, sometimes referred to as the coefficient of determination, the mean absolute error (MAE), and mean percent variation from the observed visitation, or mean percent error. The first two metrics were used to compare the overall accuracy of our predictions (median prediction) for all parks, and the latter two were used to test the accuracy of our median predictions for each park individually. R^2 is a useful measure for comparing overall model accuracy (Fig. 3), but is unreliable for small sample sizes (e.g. park specific error). R^2 also assumes a normal distribution for all data, which is not met for the park specific data, further highlighting the limitation of this metric for park specific error estimation (The Pennsylvania State University, 2018). To compare the error for specific parks, we use the other two metrics. For transparency, the R^2 for specific parks is provided on the error metrics page of the supplementary online application, but we do not recommend using this as an accuracy metric for the reasons stated above. We do not use mean percent error to measure overall model error because summing total visitation and total model predictions to calculate this would result in information on small parks being dominated by larger parks.

Exploratory Analysis

With model results in hand, we explored under what conditions Google Trends accurately forecasted national park visitation. We hypothesized model accuracy would be influenced by both the population surrounding each park and park popularity; we used average visitation as an analog for park popularity. We found the population within 50 miles (80.5 km) of each park by creating a 50-mile buffer around each park area using

ArcGIS and summing the populations of all 2010 census blocks for which the centroid was located inside the buffer area.

To explore these hypotheses, we ran correlation tests, looking at the association between both the mean park visitation (Fig. 5A) and the total population within 50 miles (80.5 km) of each park (Fig. 5B), and the mean percent error between our median visitation prediction and the observed visitation for each park.

Results

Overall Model Accuracy

We calculated the mean absolute error (MAE), and R^2 between the observed visitation and the median prediction for all parks, for all years (2013 – 2017) for both models. Our Google Trends model outperformed our autoregressive model by both metrics (Table 2.1).

Table 2.1: Overall error metrics for autoregressive and Google Trends median model predictions

Model	**MAE**	**R^2**
Google Trends	202,080	0.977
Autoregressive	230,547	0.867

Overall, our Google Trends model explains 97.7% of all variation in National Park visitation (Fig. 2.3A). Compared to our autoregressive model, which explains 86.7 % of all variation (Fig. 2.3B), the Google Trends model is much more consistent; especially when predicting high visitation numbers.

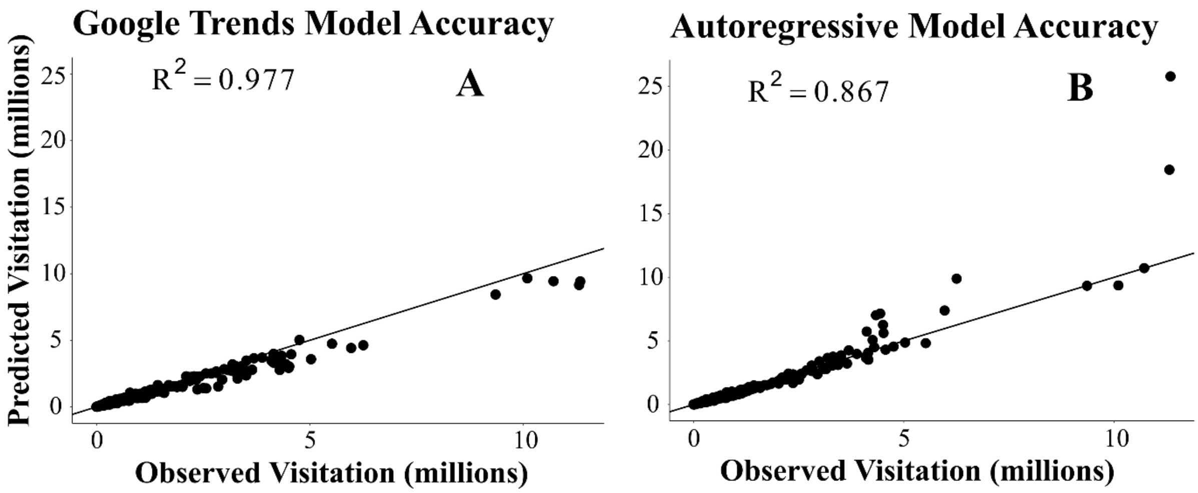


Figure 2.3. Scatterplots showing observed vs predicted visitation using the Google Trends model (Fig. A) and autoregressive model (Fig. B). The lines represent a 1:1 line of perfect fit. An interactive version of these plots (showing the year and park for each data point) is available at http://hillislab.boisestate.edu/GoogleTrendsForecasting.

Park-Specific Accuracy

We calculated the MAE and mean percent error (Fig. 2.4) between the observed visitation and the median prediction for each park, for all years (2013 – 2017) for both models (S2). At the park level, both the Google Trends and autoregressive models showed considerable variation in accuracy. Our autoregressive model produced a mean percent error that ranged from 4.37% to 39.61% for individual parks. For our Google Trends model, the low and high of this metric were 3.51% and 26.31% respectively. These values can be interpreted as follows: on the scale of the observed visitation, on average for all modeled years, how much higher or lower were the model projections for that specific park from the real visitation.

We also show the MAE for each specific park. Because MAE is highly correlated with the scale of the data (Willmott & Matsuura, 2005), we suggest that MAE should be used only to compare between models for individual parks, rather than between parks (i.e. larger parks will tend to naturally have larger MAE). For this reason, we compare predictions between parks using the mean percent error (Fig. 2.4).

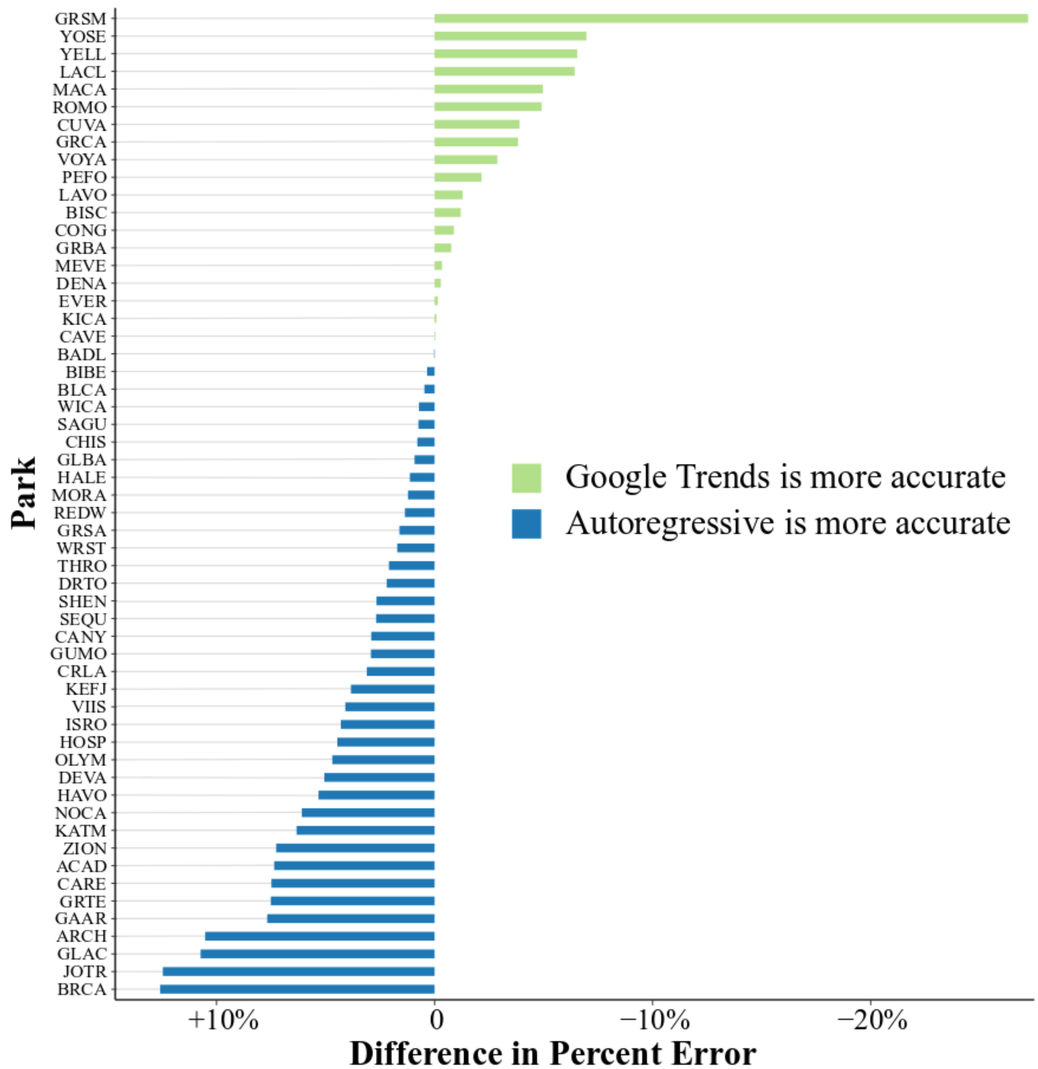


Figure 2.4. Difference in mean percent error between the Google Trends and autoregressive models, by national park. The full park name associated with each 4-letter code can be found on the online application (http://hillislab.boisestate.edu/GoogleTrendsForecasting/) under the tab "Unit code key & population data."

For the majority of national parks individually, our autoregressive model outperformed our Google Trends model. In these cases, where the autoregressive model is preferred, it is from 0.34% to 12.6% more accurate than the Google Trends model. In cases where the Google Trends model outperforms the autoregressive model, it is 0.03% to 27.2% more accurate.

Exploratory Results

Exploratory analyses examining which factors might influence the accuracy of Google Trends model predictions were largely insignificant. The mean yearly visitation to each park yielded an insignificant correlation of -0.07 with the mean percent error of each park (Fig. 2.5A). When we calculated the same metric for population within 50 miles of each park, we produced a weak correlation of -0.31 (Fig. 2.5B).

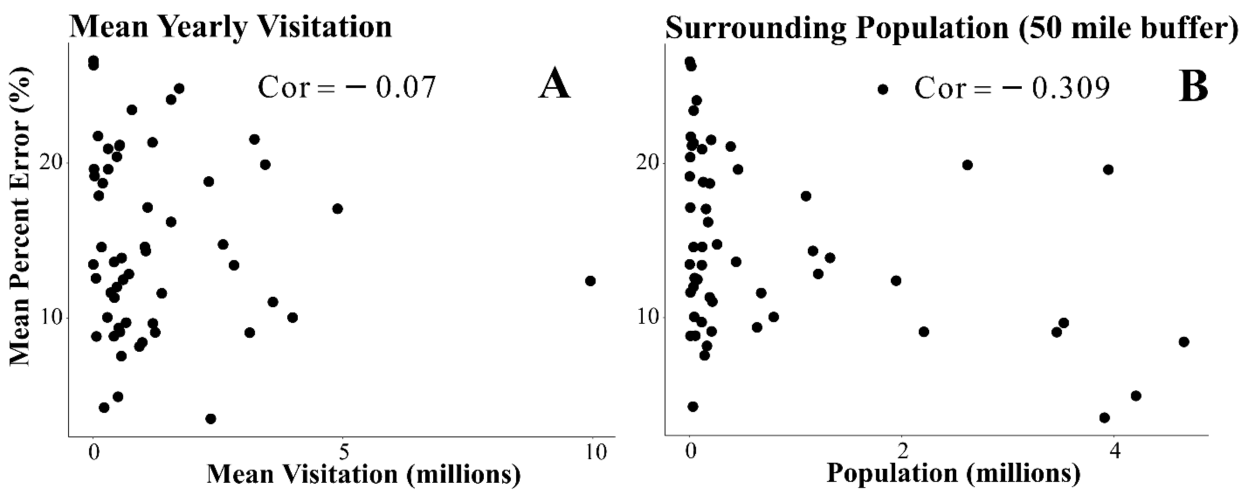


Figure 2.5. Correlations between the mean percent error of the Google Trends model and mean park visitation (Fig. A) and population within 50 miles of the park (Fig B). Each point represents one national park.

Discussion

Our study found that Google Trends is a useful tool for forecasting future visitation at U.S. National Parks. As with previous studies, which demonstrate that search engine volume is a useful indicator of future tourism arrivals (Bangwayo-Skeete & Skeete, 2015; Dergiades, Mavragani, & Pan, 2018; Yang, Pan, Evans, & Lv, 2015), we show that Google Trends can perform well in the context of U.S. National Parks. This is true despite the factors that make park visitation different from general tourism arrivals, such as limited cellular or internet service, or differences in planning behaviors. However, this study does not suggest that Google Trends is always a better tool than previously established models; rather, we encourage consideration of these data as a supplemental resource where appropriate. We speculate that Google data is most useful when park visitation is measured consistently, and given Google's status as a leading search engine. Futher, we aimed to demonstrate a method for testing the usefulness of mining search engine data for park settings, and suggest that future research continue exploring how and when these data sources can augment or update present visitation forecasting efforts.

While our Google Trends model performed better than our autoregressive model overall, the autoregressive model performed better for a higher number of individual parks. To explain these differences, we predicted that factors related to pre-trip planning (i.e. nearby population) and popularity of parks (i.e. number of visitors) would correlate with the accuracy of the Google Trends model; we expected that parks with smaller proximate populations and higher visitation would be searched more often in the pre-planning phase, and thus the Google Trends model would perform better for those parks. However, only one of these factors (nearby population) correlated loosely (cor = -0.31) with forecasting accuracy, and the relationship was the opposite of what we hypothesized (Fig. 5B). This correlation indicates that Google Trends was a slightly better predictor in parks that had larger nearby populations compared to parks with smaller nearby populations. Our hypothesis that the magnitude of visitation would impact the efficacy of our Google Trends model resulted with an insignificant correlation of -0.07. This suggests that the utility of Google Trends as a predictor is unaffected by the number of visitors a park receives. We found no minimum visitation threshold for this model to be useful.

It also appears that previous growth rate contributes to the discrepancy in model performance. The autoregressive model, although extremely accurate for the majority of parks, shows a tendency to predict unrealistically high levels of visitation (e.g. >12 million visitors) for years following visitation spikes in large parks. This tendency appears to explain the majority of the error in the autoregressive model.

Limitations and Future Research

A significant limitation when considering Google Trends data, especially from the practitioner perspective, results from how Google reports the data. Google Trends does not report raw numbers, but rather rescales values between 1 and 100, where 100 is always the highest volume of searches for the selected time range. This means that every time there is a new high in Google search interest included in a user's search parameters, the data will rescale. In other words, the values Google reports may vary based on the time range selected. It is therefore not possible to create a permanent database of trend numbers, nor is it possible to make an assessment about visitation based on a single number. Any given value on Google Trends lacks meaning alone, but rather needs to be interpreted in the context of trends over time. Additionally, values cannot be compared across search topics or time frames and it cannot be assumed that a certain value means the same thing each time Google Trends data are viewed. Alternatively, access to the algorithm, or collaboration with Google, may allow researchers to use the raw search data and yield numbers that can be used by practitioners.

Additionally, the accuracy of visitation data reported by the National Park Service (NPS) may affect the predictive ability of these models. For example, Kobuk Valley National Park reports zero visitors in 2014 and 2015. Because we used a hierarchical approach where all park predictions borrow strength from each other, the impact of a few inaccurate parks may impact the model's ability to predict for other parks (Steenbergen & Jones, 2002). Future research could couple the visitation data reported by the NPS with other sources, such as interviews with NPS staff, to build more accurate estimates of yearly park visitation.

Another limitation of using Google Trends is that countries which do not use Google would not be accounted for in a Google Trends model. While the use of Google "topics" rather than search terms accounts for language differences, visitors from those nations where use of Google is restricted or uncommon would not be included in forecasting calculations. Future research can delve into the applicability of Google Trends for specific types of cases by applying U.S. only searches, rather than international searchers, for parks that see low international visitation.

Future research into Google Trends can also experiment with smaller temporal scales, such as weekly or monthly data, or spatial scales, such as sites within parks or larger geographic regions. Smaller time scales may also allow researchers to test the hypothesis that Google Trends can be used to predict visitation changes as a direct result of acute events (e.g. superblooms, wildfire, or news & social media attention). Researchers could also explore what lag times exist between Google searching and visitation; for example, they could use questionnaires to determine how far in advance people begin researching their destination park via Google, perhaps exploring whether visitors to certain parks begin trip planning sooner. Since this study used search data from the current year to predict visitation the following year, we assumed some visitors would be searching for information about a park the year prior to visiting. Finally, future research may test alternative hypotheses as to when and why Google Trends models perform better or worse than autoregressive models.

Management Implications

Due to the limitations outlined above, we do not recommend managers substitute current autoregressive forecasting with Google Trends modeling. However, managers may consider Google Trends, or similar search volume data, as part of a mosaic of data informing expectations of future conditions. Additionally, parks and protected area managers who do not have access to forecasting tools due to time or monetary constraints, can monitor Google Trends to gain an idea of future visitation volume, particularly as it relates to past trends.

Conclusions

While the Google Trends model constructed for this study performed better than our autoregressive model overall, it does not necessarily follow that Google Trends is a superior tool for modeling individual U.S. National Parks. Instead, we suggest that Google Trends, or other search engine volume metrics, be considered when modeling future visitation, and utilized in part or in full when appropriate. Further research is needed to further explore this tool, as well as address limitations. Finally, future research may employ the methods presented in this paper to test new and emerging data sources related to visitor volume, density, spatiotemporal distribution, and more.

CHAPTER TWO REFERENCES

Araz, O. M., Bentley, D., & Muelleman, R. L. (2014). Using Google Flu Trends data in forecasting influenza-like–illness related ED visits in Omaha, Nebraska. *The American Journal of Emergency Medicine, 32*(9), 1016-1023.

Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46,* 454-464.

Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192-213.

Cahill, K., Collins, R., McPartland, S., Pitt, A., & Verbos, R. (2018). Overview of the Interagency Visitor Use Management Framework and the Uses of Social Science in its implementation in the National Park Service. *The George Wright Society Forum*, *35*(1), 32-41.

Chang, W., Cheng, J., Allaire, J. J., Xie, Y., & McPherson, J. (2018). Shiny: web application framework for R. https://CRAN.R-project.org/package=shiny

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, *88*, 2-9.

Clark, J. S. (2005). Why environmental scientists are becoming Bayesians. *Ecology letters*, *8*(1), 2-14.

Cullinane Thomas, C. L., Koontz, L., & Cornachione, E. (2018). 2017 national park visitor spending effects: Economic contributions to local communities, states, and the nation. Retrieved from Fort Collins, CO:

Daysog, R. (2018). Are people loving Hanauma Bay to death? A new study is trying to answer that question. Retrieved from http://www.hawaiinewsnow.com/2018/11/14/are-people-loving-hanauma-bay-death-new-study-is-trying-answer-that-question/

Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management, 66,* 108-120.

Duncan, D. (2016). Are we loving our National Parks to death? Retrieved from https://www.nytimes.com/2016/08/07/opinion/sunday/are-we-loving-our-national-parks-to-death.html

ESRI. (2018). USA Census Block Group Boundaries. Retrieved from http://www.arcgis.com/home/item.html?id=1c924a53319a491ab43d5cb1d55d8561

Fesenmaier, D. R., Xiang, Z., Pan, B., & Law, R. (2011). A framework of search engine use for travel planning. *Journal of Travel Research, 50*(6), 587-601.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological bulletin, 118*(3), 392.

Goodrich, B., Gabry, J., Ali, I., & Brilleman, S. (2018). rstanarm: Bayesian applied regression modeling via Stan. http://mc-stan.org/

Graefe, A. R., Vaske, J. J., & Kuss, F. R. (1984). Social carrying capacity: An integration and synthesis of twenty years of research. *Leisure Sciences*, *6*(4), 395-431.

Hallo, J. C., & Manning, R. E. (2010). Analysis of the social carrying capacity of a national park scenic road. *International Journal of Sustainable Transportation, 4*(2), 75-94.

Halpenny, E. A. (2010). Pro-environmental behaviours and park visitors: The effect of place attachment. *Journal of Environmental Psychology, 30*(4), 409-421. doi:10.1016/j.jenvp.2010.04.006

Hand, C., & Judge, G. (2012). Searching for the picture: forecasting UK cinema admissions using Google Trends data. *Applied Economics Letters, 19*(11), 1051-1055.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57-66.

Maller, C., Townsend, M., Pryor, A., Brown, P., & St Leger, L. (2006). Healthy nature healthy people:'contact with nature'as an upstream health promotion intervention for populations. *Health promotion international*, *21*(1), 45-54.

Maples, J. N., Sharp, R. L., Clark, B. G., Gerlaugh, K., & Gillespie, B. (2017). Climbing out of Poverty: The Economic Impact of Rock Climbing in and around Eastern Kentucky's Red River Gorge. *Journal of Appalachian Studies*, *23*(1), 53-71.

Marion, J. L., Leung, Y.-F., Eagleston, H., & Burroughs, K. (2016). A review and synthesis of recreation ecology research findings on visitor impacts to wilderness and protected natural areas. *Journal of Forestry, 114*(3), 352-362.

National Park Service. (2018a). Administrative Boundaries of National Park System Units 9/30/2018. Retrieved from https://irma.nps.gov/DataStore/Reference/Profile/2224545?lnv=True

National Park Service. (2018b). Annual Visitation Summary Report for 2017. Retrieved from https://irma.nps.gov/Stats/SSRSReports/National Reports/Annual Visitation Summary Report (1979 - Last Calendar Year)

National Park Service. (2018c). Welcome to Visitor Use Statistics. Retrieved from https://irma.nps.gov/Stats/

Önder, I., & Gunter, U. (2016). Forecasting tourism demand with Google trends for a major European city destination. *Tourism Analysis, 21*(2-3), 203-220.

Park, S., Lee, J., & Song, W. (2017). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing, 34*(3), 357-368.

Sessions, C., Wood, S. A., Rabotyagov, S., & Fisher, D. M. (2016). Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management, 183*, 703-711. doi:10.1016/j.jenvman.2016.09.018

Simmonds, C., Annette, M., Reilly, P., Maffly, B., Wilkinson, T., Canon, G., . . . Whaley, M. (2018). Crisis in our national parks: how tourists are loving nature to death. Retrieved from https://www.theguardian.com/environment/2018/nov/20/national-parks-america-overcrowding-crisis-tourism-visitation-solutions

Steenbergen, M. R., & Jones, B. S. (2002). Modeling multilevel data structures. American Journal of Political Science, 46, 218-237. doi:10.2307/3088424

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports, 7*(1), 17615. doi:10.1038/s41598-017-18007-4

The Pennsylvania State University. (2018). Lesson 1.5: The Coefficient of Determination, r-squared. In *STAT 501: Regression Methods*. Retrieved from https://onlinecourses.science.psu.edu/stat501/

Vosen, S., & Schmidt, T. (2011). Forecasting private consumption: survey-based indicators vs. Google trends. *Journal of Forecasting, 30*(6), 565-578.

Wilmot, N. A., & McIntosh, C. R. (2014). Forecasting recreational visitation at US National Parks. *Tourism Analysis, 19*(2), 129-137.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30(1), 79-82.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific reports, 3*, 2976. doi:10.1038/srep02976

Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management*, *46*, 386-397.

Ziesler, P. S. (2016). *Statistical Abstract: 2015*. Retrieved from Fort Collins, CO: https://irma.nps.gov/DataStore/DownloadFile/548275

Ziesler, P. S., & Singh, P. (2018). *Statistical Abstract: 2017*. Retrieved from Fort Collins, CO: https://irma.nps.gov/DataStore/DownloadFile/600257

APPENDIX A

S1. Chapter 1 Supplemental Information

Survey tool used for data collection in chapter 1

Default Question Block

Welcome and thank you for participating in this survey!

Your participation is helping to inform research on efficient management of Montana's wetlands!

Principle investigator: Matt Clark
Matthewclark969@boisestate.edu

**Please make sure that the person responding to this survey has a working knowledge of your organization's wetlands projects. This information is all collected, stored, and published anonymously.

Q1. What organization do you work for? (If you work for multiple organizations, indicate the organization that fills the more of your time.)

	Organization	Location/Field office

Q2. Over the past two years, has your organization been involved in wetland conservation in Montana?

- ◯ No
- ◯ Yes

S2. Chapter 2 Supplemental Information

Table S1. Park specific error metrics for autoregressive (AR) and Google Trends (GT) model predictions.

Park	GT MAE	GT mean percent error (%)	AR MAE	AR mean percent error (%)
Acadia	488,431	14.75	217,869	7.39
Arches	309,460	21.33	153,804	10.81
Bad Lands	81,481	8.17	80,212	8.14
Big Bend	45,484	11.64	43,998	11.30
Biscayne	23,656	4.93	29,966	6.13
Black Canyon of the Gunnison	48,663	18.70	46,745	18.24
Bryce Canyon	539,741	24.09	240,666	11.50
Canyonlands	149,904	21.16	125,657	18.26
Capitol Reef	243,327	23.43	164,637	15.95
Carlsbad Caverns	42,936	8.83	42,953	8.87
Channel Islands	58,537	19.60	60,121	18.81
Congaree	22,580	17.89	23,847	18.77
Crater Lake	143,822	21.10	120,365	17.99
Cuyahoga Valley	78,037	3.51	172,225	7.41
Denali	118,870	20.41	118,892	20.70
Death Valley	179,446	14.59	114,889	9.53
Dry Tortugas	8,351	12.57	6,778	10.37
Everglades	88,890	8.43	89,947	8.57
Gates of the Arctic	1,472	13.46	663	5.79
Glacier	542,936	18.80	218,064	8.07
Glacier Bay	64,049	12.01	58,709	11.09
Great Basin	31,801	21.74	31,626	22.50
Grand Canyon	968,392	17.05	1,242,266	20.88

Great Sand Dunes	82,670	20.92	76,583	19.31
Great Smokey Mountains	1,340,246	12.40	4,469,575	39.61
Grand Teton	422,420	13.41	190,025	5.90
Guadalupe Mountains	26,590	14.58	22,051	11.66
Haleakala	176,253	17.14	172,697	16.01
Hawaii Volcanos	299,417	16.20	191,847	10.87
Hot Springs	172,655	11.60	105,476	7.14
Isle Royale	5,769	26.31	5,149	22.01
Joshua Tree	605,830	24.81	273,236	12.35
Katmai	5,743	19.16	4,013	12.83
Kanai Fjords	30,799	10.05	20,170	6.21
Kings Canyon	78,167	13.88	80,267	13.96

Table S1 cont.

Park	GT MAE	GT mean percent error (%)	AR MAE	AR mean percent error (%)
Kobuk Valley	8,117	NA	8,247	NA
Lake Clark	5,309	26.61	6,270	33.04
Lassen Volcanic	67,888	13.62	72,385	14.91
Mammoth Cave	53,842	9.37	80,152	14.34
Mesa Verde	49,219	9.11	51,552	9.45
Mount Rainier	129,197	9.66	110,668	8.44
Northern Cascades	4,909	19.61	3,603	13.52
Olympic	297,997	9.06	140,088	4.37
Petrified Forest	76,577	9.71	89,325	11.86
Redwood	58,523	11.32	51,400	9.96
Rocky Mountain	834,281	19.90	1,037,235	24.80

Saguaro	111,526	12.84	102,733	12.10
Sequoia	169,893	14.33	132,021	11.65
Shenandoah	128,865	9.08	87,602	6.42
Theodore Roosevelt	86,008	12.48	70,686	10.39
Virgin Islands	46,978	13.46	30,911	9.36
Voyageurs	10,086	4.23	16,909	7.11
Wind Cave	44,646	7.55	41,348	6.83
Wrangell-St. Elias	6,427	8.82	5,444	7.10
Yellowstone	458,662	11.04	695,581	17.58
Yosemite	467,972	10.04	718,864	17.00
Zion	874,822	21.53	572,177	14.26