# OPTIMIZING RESOURCE ALLOCATION IN COMPUTATIONAL SUSTAINABILITY: MODELS, ALGORITHMS AND TOOLS

A Dissertation
Presented to
The Academic Faculty

By

Amrita Gupta

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Computational Science and Engineering
College of Computing

Georgia Institute of Technology

May 2021

**OPTIMIZING RESOURCE ALLOCATION IN COMPUTATIONAL SUSTAINABILITY: MODELS, ALGORITHMS AND TOOLS**

Approved by:

Dr. Bistra Dilkina, Advisor
Department of Computer Science
*University of Southern California*
School of Computational Science & Engineering (Adjunct)
*Georgia Institute of Technology*

Dr. Duen Horng (Polo) Chau
School of Computational Science & Engineering
*Georgia Institute of Technology*

Dr. Ümit Çatalyürek
School of Computational Science & Engineering
*Georgia Institute of Technology*

Dr. Dan Morris
AI for Earth
*Microsoft Research*

Dr. Angela Fuller
Department of Natural Resources
*Cornell University*

Date approved: January 5, 2021

To my parents, the original Drs. Gupta.

## ACKNOWLEDGMENTS

Firstly, I am immensely grateful to my Ph.D. advisor Bistra Dilkina, for her support, enthusiasm, and encouragement, and for introducing me to the exciting realms of computational sustainability, algorithms and optimization. She has taught me by example to seek out and embrace challenges and opportunities, to strive for clarity of purpose and expression, and to keep improving always. Thanks for everything, Bistra! I also thank my other Ph.D. committee members: Polo Chau and Ümit Çatalyürek, for their sustained support for me as a remote student of the department; to Angela Fuller for her patience and invaluable guidance on bridging computational science and conservation science; and to Dan Morris, for giving me the opportunity to contribute to the Microsoft Camera Traps project and to work more closely than ever with the intended beneficiaries of my research.

My research was funded in part by the National Science Foundation under NSF grant CCF-1522054 (COMPSUSTNET:Expanding Horizons of Computational Sustainability), the U.S. Department of Homeland Security under grant 2015-ST-061-CIRC01, and the Microsoft AI for Earth initiative. I am grateful for the ongoing support these institutions provide for academic research.

My doctoral studies would not have been possible without the support and help of a large group of people at Georgia Tech. I'd like to thank the CSE administrative team, especially Deanna Richards, Nirvana Edwards, Della Phinisee and Arlene Washington-Capers, who never let me fall through the cracks and granted me their time and patience. I am also thankful to my friends and fellow graduate students for enriching my years at Tech. I am honored to have been part of the Fall 2015 CSE Ph.D. cohort–we made it! Special thanks are due to my labmates Elias Khalil, Caleb Robinson, and Aaron Ferber for their companionship and for the privilege of learning alongside them these past few years. I also have to thank Eisha Nathan, Rakshit Trivedi, Karl Gemayel, Mia Glaese, and Ari Afshar for making the days (and nights) spent in Klaus so much more enjoyable. Outside of CSE, the CPL

Drinks crowd helped me feel truly at home in Atlanta. Thanks to Brian Goldfain, Daniel Henderson, Nolan Wagener, Jason Fernando, Yannick Schroecker, Paul Drews, Shan Tie, and others for our weekly dinners around Midtown that kept me well-fed and reasonably well-socialized. Andrew Price has also been a great friend and inspiration, whose boundless energy and enthusiasm for work as well as off-campus adventures continues to inspire me to do as much as I can. I am also deeply thankful to Brian Goldfain for keeping me grounded and sane especially during the final years of my Ph.D. He has helped me celebrate wins, bounce back during difficult times, and (mostly) stick to a regular running schedule.

I have also had an amazing support network outside Georgia Tech. I was lucky to find a roommate like Kianna Phillip, who worked more hours in more jobs than anyone else I've ever met, yet was always keen to catch up when we were both home. The Goverdhan family gave me a sense of home too, welcoming me into their lives and always being ready to have me over for Thanksgiving, Christmas, Easter, or just a weekend away from campus. I am also grateful for my friendships with Saumya Gurbani, Nirav Shelat, and Greg Kheyfets, that have lasted despite geographic distance. In addition, I have had the unusual fortune of getting to know not one, but two university communities during my time as a doctoral student. Thank you to Tye Hines, Hailey Winetrobe-Nadel, Laksh Matai, Sina Aghaei, Taoan Huang and everyone else at USC CAIS for helping me find my bearings in a new city.

Most of all, I am thankful to my parents: their unwavering love and encouragement throughout this endeavor gave me the courage to keep going. Thank you for pushing me to pursue a good education and opportunities to grow, even if those took me far away from home.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF ACRONYMS

**Budget-PCSF** budget-constrained prize-collecting Steiner forest

**GMM** Gaussian mixture model

**ILPs** integer linear programs

**IUCN** International Union for Conservation of Nature

**MST** minimum spanning tree

**OD** origin-destination

**SCR** spatial capture-recapture

**SDGs** Sustainable Development Goals

## SUMMARY

The 17 Sustainable Development Goals laid out by the United Nations include numerous targets as well as indicators of progress towards sustainable development. Decision-makers tasked with meeting these targets must frequently propose upfront plans or policies made up of many discrete actions, such as choosing a subset of locations where management actions must be taken to maximize the utility of the actions. These types of resource allocation problems involve combinatorial choices and tradeoffs between multiple outcomes of interest, all in the context of complex, dynamic systems and environments. The computational requirements for solving these problems bring together elements of discrete optimization, large-scale spatiotemporal modeling and prediction, and stochastic models.

This dissertation leverages network models as a flexible family of computational tools for building prediction and optimization models in three sustainability-related domain areas: 1) minimizing stochastic network cascades in the context of invasive species management; 2) maximizing deterministic demand-weighted pairwise reachability in the context of flood resilient road infrastructure planning; and 3) maximizing vertex-weighted and edge-weighted connectivity in wildlife reserve design. We use spatially explicit network models to capture the underlying system dynamics of interest in each setting, and contribute discrete optimization problem formulations for maximizing sustainability objectives with finite resources. While there is a long history of research on optimizing flows, cascades and connectivity in networks, these decision problems in the emerging field of computational sustainability involve *novel objectives*, *new combinatorial structure*, or *new types of intervention actions*. In particular, we formulate a new type of discrete intervention in stochastic network cascades modeled with multivariate Hawkes processes. In conjunction, we derive an exact optimization approach for the proposed intervention based on closed-form expressions of the objective functions, which is applicable in a broad swath of domains beyond

invasive species, such as social networks and disease contagion. We also formulate a new variant of Steiner Forest network design, called the budget-constrained prize-collecting Steiner forest, and prove that this optimization problem possesses a specific combinatorial structure, restricted supermodularity, that allows us to design highly effective algorithms. In each of the domains, the optimization problem is defined over aspects that need to be predicted, hence we also demonstrate improved machine learning approaches for each.

# CHAPTER 1
# INTRODUCTION

## 1.1 Dissertation Overview

Sustainable development, a notion defined in 1987 by the World Commission on Environment and Development [1], aims to ”meet the needs and aspirations of the present without compromising the ability to meet those of the future”. In 2015 the United Nations put forward a comprehensive set of 17 Sustainable Development Goals (SDGs) and accompanying targets to serve as guidelines for securing socioeconomic prosperity and environmental protection. The SDGs also specify indicators for measuring progress towards each target that in many cases can serve directly as quantitative objectives for guiding policy and decision-making. For instance, SDG 11 (sustainable cities and communities) prescribes taking action to reduce ”disaster damage to critical infrastructure” in Target 11.5.2. SDG 15 (life on land) directs countries to increase the proportion of important biodiversity sites covered by protected areas (Targets 15.1.2 and 15.4.1) and to allocate adequate resources towards the control of invasive alien species (Target 15.8.1).

Designing effective policies or plans for achieving these sustainability objectives often translates to solving resource allocation problems in systems with complex underlying spatial or spatiotemporal dynamics. As a first step, we must acquire data or models that capture these dynamics well enough to inform the optimization step; then, we must design plans that are optimized against the available data or models [2]. Network models lend themselves to both of these steps. Networks or graphs are highly flexible structures made up of a discrete set of objects (vertices) in which some pairs of objects have connections (edges), and are a common choice for modeling geospatial phenomena in which vertices have locations and edges have lengths or weights. A wide variety of dynamics can be described

as deterministic flows or stochastic cascades between vertices along edges. Concurrently, there is a long history of research on optimizing flows, cascades and connectivity in networks, e.g. by selecting a subgraph over which the dynamics of interest are optimized; or by modifying vertex or edge properties or network topology [3, 4, 5] in order to manipulate the flows or cascades over the whole network. However, decision problems in the emerging field of computational sustainability involve *novel objectives*, *new combinatorial structure*, or *new types of intervention actions*.

This dissertation studies three sustainability-related application domains as sources of compelling new computational problems at the junction of **large-scale spatiotemporal modeling and prediction** and **discrete optimization** in networks: 1) limiting the spread of invasive alien species; 2) improving the climate resilience of critical infrastructure; and 3) protecting ecologically important sites in wildlife reserves. We use spatially explicit network models to capture the underlying system dynamics of interest in each setting, and contribute discrete optimization problem formulations for maximizing sustainability objectives with finite resources.

## 1.2 Research Contributions

### 1.2.1 Minimizing Stochastic Network Cascades

Chapter 2 of this dissertation studies minimizing stochastic network cascades with a focus on the application domain of invasive species management. The spread of invasive species to new areas disrupts ecosystem processes and causes crop loss, posing major threats to biodiversity and food security. According to the International Union for Conservation of Nature (IUCN), invasive species undermine progress towards 10 out of 17 of the the UN SDGs. Eradication efforts are typically costly and labor-intensive, so there is a great deal of interest in optimizing the allocation of control efforts in order to curb the invasion process as effectively as possible. Our work addresses the problem of strategically selecting land parcels in which to conduct invasive species eradications within a fixed budget, with the

objectives of minimizing either the population growth rate or the total population size of the invasive species at the end of a finite time horizon. Using our proposed modeling and optimization approaches, we are able to obtain eradication plans that may be nearly as effective as completely eradicating invasive plants from the landscape, but at only a fraction of the cost. This directly contributes towards addressing objectives in SDG 15 (life on land) as well as SDG 2 (zero hunger). Our work makes the following contributions in particular:

**Contributions to Network Cascade Modeling:** Multivariate Hawkes processes have been used to model self- and mutually-exciting activity in a wide variety of settings, including the occurrence of earthquakes and aftershocks and the spread of posts in social networks. We demonstrate for the first time that multivariate Hawkes processes can also describe the spatiotemporal spread dynamics of invasive plant species. We also propose a novel intervention for limiting stochastic event cascades in networks: history deletion. Prior literature on activity shaping in multivariate Hawkes processes has focused on increasing the continuous background rate of events at selected vertices. In contrast, our proposed discrete intervention modifies selected vertices by clearing past events at those vertices, thereby removing the self-exciting or mutually-exciting effect of those events to activity in the network. In the context of invasive species management, this intervention corresponds to eradicating invasive individuals in selected land parcels, but the proposed intervention is of high relevance in many other applications as well, such as containing disease contagion by treating patients in selected areas, or minimizing the spread of misinformation in social networks by hiding or making unshareable posts under select accounts that create misleading posts.

**Contributions to Cascade Optimization:** To address intervention optimization in the context of multivariate Hawkes processes, we derive novel analytical formulas for two control objectives of interest after history deletion at selected vertices: 1) the expected rate of activity in the network at the end of the planning horizon, and 2) the expected total ac-

tivity in the network up to the end of the planning horizon. These analytical results are an important contribution, as they unlock two key algorithmic benefits for optimizing diffusion in networks. The first advantage comes from the fact that our analytical expressions directly compute the control objectives *in expectation* over all possible realizations of the stochastic process. This avoids the need to estimate the objective functions empirically as is often necessary in stochastic optimization, for instance by simulating multiple realizations of random cascades and computing the sample average of objective function values resulting from a given strategy. Another key aspect of our analytical expressions is that each of the objective functions can be expressed as a linear function of binary decision variables encoding which vertices are selected for history deletion. Combined with linear budget constraints, this allows us to formulate our optimization problems as integer linear programs (ILPs) and correspondingly leverage powerful commercial solvers to obtain optimal solutions, or solutions with known optimality gaps.

**Contributions to Invasive Species Control:** We add multivariate Hawkes processes to the family of models applicable to invasive species. We demonstrate that, relative to other point process-based models of invasive species dynamics, the multivariate Hawkes process model provides a better fit to the data when evaluated on a real-world dataset recording the invasion of conifer trees into a meadow ecosystem. Furthermore, our Hawkes process optimization approach generates intervention plans that outperform several intervention decision practices currently used to identify land parcels to prioritize for invasive species eradication. Our intervention optimization approach has the potential to significantly improve the efficacy of invasive species removal efforts, especially in highly resource-constrained settings.

Most alternate models for invasion dynamics rely on agent-based simulations, making them challenging to integrate into intervention optimization schemes. In contrast, while the Hawkes process model is also able to produce simulated cascades based on the learned pa-

rameters, it facilitates the derivation of the aforementioned analytical expressions and subsequent optimization models for real-world management objectives. This offers a greater synergy between model-building and decision-making stages in data- and learning-driven decision-making.

### 1.2.2 Maximizing Deterministic Demand-Weighted Pairwise Reachability

Chapter 3 of this dissertation studies maximizing demand-weighted connectivity in networks in the context of transportation infrastructure resilience to flooding. Flood disasters are the most widespread, frequent, and expensive natural disasters affecting humankind today, and have the capacity to cause large-scale and long-term disruptions to road-based mobility. In order to minimize the overall negative impacts of these disasters, government agencies are interested in strategically upgrading key components of these critical infrastructures. Our work addresses the question of determining which set of roads should be fortified in order to protect the feasibility of as many trips as possible with a finite budget for fortifications. In contrast to our work on stochastic cascades over networks where the network topology was fixed and the goal was to minimize activity in the network as a whole by intervening at specific vertices; here, our goal is to choose a subgraph in our network by selecting edges within a fixed budget such that deterministic flows between connected pairs of vertices in the subgraph are maximized. By focusing on preserving the large-scale, normal functioning of road networks after flood disasters, we address an **understudied problem in climate adaptation** that also links to a relatively **understudied network design problem in computer science**. This work was awarded a prize by the United Nations Global Pulse and was presented at the 2017 United Nations Climate Change Conference (COP 23). In addition, our work furthers progress towards several SDGs, including SDG 13 (climate action), SDG 11 (sustainable cities and communities), and SDG 9 (industry, innovation and infrastructure). More specifically, we make contributions to the following areas:

**Contributions to Network Design:** We establish that our network design problem is in fact an instance of the budget-constrained prize-collecting Steiner forest (Budget-PCSF) problem, a previously understudied Steiner problem variant. We analyze the objective function of Budget-PCSF and discover a novel discrete structural property—restricted supermodularity. Seen as a set function over edges, the Budget-PCSF objective function exhibits *compounding gains* as edges are successively added to a set, as long as the subgraph induced by the edges in the set remains acyclic. In addition to modeling a real-world network design problem of high social impact, this characterization is also a novel category of optimization functions, as prior work has only considered purely supermodular functions or restricted *sub*modular functions. In order to maximize our restricted supermodular function over the graph matroid subject to our budget constraint, we extend recent work on using semigradient ascent to maximize supermodular functions subject to budget constraints. Our proposed suite of semigradient-based heuristic algorithms are shown to find solutions with dramatically better objective function value than a baseline greedy heuristic. It is worth highlighting that our algorithms can be applied directly to *any* problem involving the maximization of a restricted supermodular function over the graph matroid subject to a knapsack constraint.

**Contributions to Flood Resilient Road Infrastructure:** Our work details a framework for coupling travel demand and flood prediction models to quantify the impacts of floods in terms of mobility. We combine flood depth rasters with origin-destination travel volumes in order to estimate the mobility impacts of floods in terms of the satisfiability of typical travel demand patterns. Origin-destination (OD) travel volume data between pairs of zones is thus vital for transportation resilience planning, but in fact these flows are rarely measured. In the United States, OD travel volumes between census blocks due to *commutes to work* are collected by the Census Bureau as part of their Longitudinal Employer–Household Dynamics (LEHD) survey, but trips taken for other reasons are not captured. Overall travel

volume patterns can be mined from call detail records if, for instance, cell phone data for a sufficiently large randomized sample of the population in the study is available. This is the methodology we use in Section 3.2 to estimate OD flows in Senegal, although unfortunately, call detail record data are also not widely available for use.

There is a great deal of active research on predicting OD flows based on spatial, demographic, economic and geographic features, although most traditional and machine learning-based approaches for travel flow prediction train and test on data from the same study area (by splitting zones into disjoint train and test sets). In Section 3.3, we highlight recent results showing for the first time that mobility flow prediction models can generalize to unseen study areas. We study the decision error of these predicted flows in the context of informing road upgrades, by comparing the number of ground truth flows made feasible under the upgrade plans obtained using either predicted or ground truth travel flows as data inputs to the optimization procedure. We show that travel flow prediction models trained on data from other study areas capture generalizable mobility patterns that are sufficient to guide strategic road fortification efforts. This is promising as it enables the application of our mobility resilience analysis and optimization framework to be applied beyond only those areas with ground truth travel flow data.

### 1.2.3 Maximizing Vertex-Weighted and Edge-Weighted Connectivity

Chapter 4 of this dissertation studies maximizing demand-weighted connectivity in networks. This is a similar goal to the problem domain in Chapter 3: both aim to find a subgraph over which accessibility between pairs of vertices is maximized for as many individuals as possible. However, there are some important differences that distinguish the two network design problems. In Chapter 3 the solution subgraph is induced by selecting *edges*, whereas in Chapter 4 the solution subgraph is induced by selecting *vertices*. In Chapter 3, we consider flows only between vertex pairs that are reachable via selected edges. In contrast, in Chapter 4 we consider flows between pairs of selected vertices, where the connec-

tivity between them is a static function of the shortest path distance between them using any edges in the graph. We motivate this graph optimization problem through its applicability to designing effective protected areas for biodiversity. One of the Aichi biodiversity targets established in 2010 by the Convention on Biological Diversity was to conserve at least 17% of terrestrial natural areas by 2020 through well-connected reserve systems that should support the long-term persistence of species. Both population size and functional landscape connectivity are important factors contributing to persistence. We address the problem of strategically purchasing land parcels with a fixed budget to create protected areas or wildlife reserves, with the objectives of maximizing the number of target species individuals protected by the reserve as well as maximizing the landscape connectivity between protected land parcels. Our work contributes towards addressing objectives in SDG 15 (life on land). We make the following contributions:

**Contributions to Reserve Design Optimization:** Our work formally connects reserve design with landscape metrics derived from spatial capture-recapture (SCR) models. SCR models are a state-of-the-art hierarchical latent variable model gaining popularity in landscape ecology. The models are fit to animal observations from a set of spatially distributed detectors, such as motion- or thermally-activated cameras or camera traps. The SCR model provides spatially-explicit estimates of animal population density as well as landscape connectivity metrics that describe the capacity of individuals to move through the landscape, with implications for important ecological functions like foraging, dispersal and migration. Despite the utility of these density and connectivity estimates, there is still relatively limited work that directly integrates SCR model outputs into systematic conservation planning. We derive ILPs formulating budget-constrained reserve site selection with objectives maximizing estimated density within the reserve or estimated functional connectivity between reserve sites. We also provide an ILP formulation for maximizing a recently-proposed landscape metric—density-weighted connectivity—which combines population

density and landscape connectivity in an ecologically meaningful manner. This contributes to the growing body of work on reserve design using exact optimization methods via mixed integer linear programs. Furthermore, while most previous research on reserve design optimization focuses exclusively on either maximizing habitat or species density, or alternatively on maximizing reserve spatial contiguity, our hybrid density-weighted connectivity objective incorporates both properties without requiring multiobjective optimization techniques. Finally, to the best of our knowledge, ours is also the first piece of work that incorporates the resource requirements of individual animals into the reserve design process, which we achieve by adding home range constraints to the optimization problem.

**Contributions to Species Distribution Modeling:** Species distribution models in general and spatial capture-recapture models in particular rely on aggregating detections of species or individual animals during wildlife surveys. As mentioned before, these surveys are increasingly conducted by deploying camera traps at a set of locations to collect image data whenever animals encounter the traps. Typically, these images are manually reviewed by ecologists to identify species of interest. In order to accelerate this tedious and time-consuming process, machine learning and computer vision researchers have recently attempted to build deep neural network models for classifying species in camera trap images, with especial focus on strategies for learning with limited labeled data. In Chapter 5 of this dissertation, we explore semi-supervised learning as one such strategy, and propose a novel deep learning loss function based on Gaussian mixture model (GMM) likelihood. Coupled with a supervised loss function, the unsupervised GMM loss reduces the tendency of the model to overfit to the small labeled training points. In a different direction, we survey a wide range of active learning query strategies to gain insight into how to train a classifier with limited labels. Active learning can be considered as another example of resource constrained optimization, in which the limited resource is time or labeling effort from human experts, and the objective is to choose the subset of samples to label that will

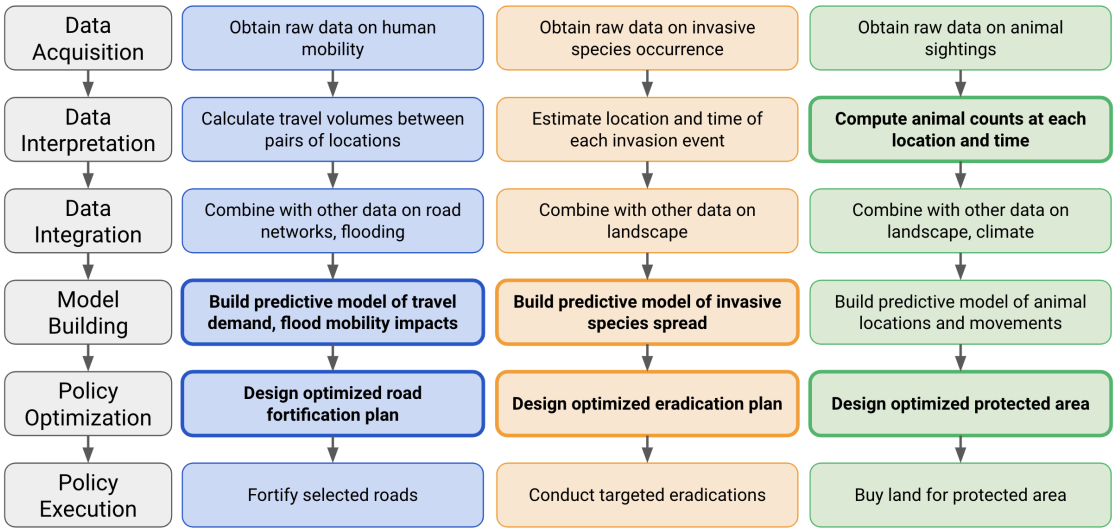



Figure 1.1: Framework for data- and learning-driven policy-making applied to three sustainability-related settings. While computing and artificial intelligence can be leveraged in all stages of this framework, this dissertation makes contributions to the components in bold.

most improve the classifier's performance.

### 1.2.4 Overall Research Contributions and Published Papers

Altogether, this dissertation presents several instances of novel modeling and decision problems in real-world settings. We present new scalable discrete optimization, network modeling and deep learning methodology to address these settings, thereby contributing to the overall workflow of implementing data-driven decision making in sustainable development Figure 1.1.

1. **Gupta, A.**, Farajtabar, M., Dilkina, B. and Zha, H., 2018, July. Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management. In *International Joint Conferences on Artificial Intelligence (IJCAI).*

2. **Gupta, A.***, Robinson, C.*, and Dilkina, B., 2018. Infrastructure Resilience for Climate Adaptation. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies.* (* denotes joint first-authorship)

3. **Gupta, A.** and Dilkina, B., 2019, November. Budget-Constrained Demand-Weighted Network Design for Resilient Infrastructure. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI).*

4. **Gupta, A.**, Dilkina, B., Morin, D.J., Fuller, A., Royle, J.A., Sutherland, C. and Gomes, C.P., 2019. Reserve Design to Optimize Functional Connectivity and Animal Density. In *Conservation Biology.*

5. Qiu, G., **Gupta, A.**, Robinson, C., Feng, S. and Dilkina, B., 2020, December. Learning-Based Travel Prediction in Urban Road Network Resilience Optimization. In *2021 Association for the Advancement of Artificial Intelligence Workshop on AI for Urban Mobility.*

# CHAPTER 2
# BUDGET-CONSTRAINED COMBINATORIAL OPTIMIZATION FOR NETWORK FLOWS: INVASIVE SPECIES MANAGEMENT

## 2.1 Introduction

Network diffusion models are a powerful tool for studying processes like the spread of influence and information through social networks [6, 7, 8], the dispersal of species through a landscape [4], and disease contagion in populations [9].The ability to model the dynamics of these diffusion processes enables the development of strategies for steering them towards desirable outcomes. For instance, one might selectively add nodes to an existing network to facilitate diffusion [4], or one can strategically block transmission along a set of links [3, 5] to limit it.

Two of the most studied network diffusion models are the independent cascade (IC) model and the linear threshold (LT) model [6]. In both, the spreading process is modeled as an activation of nodes over discrete time steps. Each node in the network is in a binary state (active or not), and nodes are activated by their active neighbors. In both the IC and LT models, once a node is active it remains so for the rest of the diffusion process, an assumption that is appropriate for modeling the spread of irreversible phenomena, e.g. the adoption of a product or infection by a disease that confers permanent immunity.

However, many network diffusion processes exhibit *non-progressive* cascades where an active node can become inactive probabilistically at each time step, so that the state of a node fluctuates over time. For example, in species dispersal, a previously occupied habitat patch may become unoccupied [4], or in infectious disease a patient may recover but be susceptible to reinfection. In this setting, repeated exposure to activation events plays an important role in continuing the diffusion process by reactivating nodes that have become

inactive. Sometimes, exposure to multiple activations can also cause a node to become "more" active, e.g. the posting frequency of an individual social media user can increase due to high activity in their network. In these cases, it is more fitting to model the state of a node as a time-varying, real- or continuous-valued function rather than binary states. Moreover, activation events typically arrive continuously rather than in discrete time steps, warranting the diffusion process to be modeled in continuous time. Other network models like Susceptible-Infected-Recovered (SIR) can encompass some of these characteristics and have been shown to have a close connection to point process models [10], which are more flexible and intuitive and easily allow the node activities to be modeled according to observed/assumed principles (e.g. mutual-excitation).

*Temporal point processes* offer a framework for modeling diffusion processes with both continuous activity states and continuous time. The activity of a node can be characterized by a parameter $\lambda$ representing the rate at which the node stochastically generates events. This $\lambda$ parameter itself can be responsive to activations arriving at the node, thereby capturing self-exciting behavior in the diffusion process. Temporal point processes have recently been applied to modeling several diffusion processes like the activity of Twitter users [11] and the spread of avian flu [12]. In the context of invasive species, [13] use a spatiotemporal point process model to characterize the spread of an invasive banana plant, although they do not consider any control.

In terms of controlling diffusion processes, a variety of intervention actions have been analyzed in the discrete-time, binary-state setting, such as selecting source nodes for initiating cascades [6] and modifying network connectivity to guide the diffusion by adding or removing nodes [4] and edges [3, 5] or modifying edge weights [14]. In contrast, there has been relatively little work on controlling dynamics in network temporal point processes. One possible control action is to manipulate the activity rate parameters $\lambda$ at specific nodes, e.g. by incentivizing social media users to post more frequently. Steering user activity in this manner was first considered in [11], and was used to develop a multistage strategy for

mitigating fake news [15]. Recent work has also applied methods from stochastic differential equations to find the best intensity for information guiding [16] and achieving highest visibility [17]. In our work, a *discrete* intervention for network point processes is considered for the first time that, unlike the above, modifies the activity rate parameter at select nodes by *deleting the history* of the point process.

Our work is motivated by the invasive species management problem in biodiversity conservation. The spread of non-native species to new areas is a cause of major concern, because they harm native species by disrupting food webs and ecosystem processes. These adverse effects have generated significant interest in limiting their spread. Management is often performed by eradicating invasive species individuals, but their removal can be prohibitively costly. In light of this, a common objective is to optimize the location of control efforts in order to maximize the efficacy of the intervention. We derive a novel approach for finding an optimal set of locations at which to remove species given a fixed budget. Although our work is motivated by a critical problem in environmental sustainability, the novel computational problem it poses appears in other domains that can be modeled using temporal point processes, such as mitigating the spread of infectious diseases using vaccination or active screening programs. The computational approach we develop here can be generalized to these broader applications.

### 2.2 Problem Statement

In the invasive species management problem, the goal is to identify locations at which to eradicate invasive individuals in order to minimize the spread of the species through the landscape. Let $L$ be a set of $n$ distinct land parcels corresponding to basic units of management. An invasive species is observed to be proliferating and dispersing through the landscape until a given time $\tau$, when an intervention is performed by eliminating all invasive individuals present before $\tau$ in a set of land units $U \subseteq L$. Each land unit $i \in L$ has an associated cost $c_i$ reflecting economic land management costs or effort needed to

eradicate the invasive individuals, and the total cost of the intervention cannot exceed a given budget $\mathcal{B}$. A feasible intervention plan is therefore a set of land units $U$ with total intervention cost within $\mathcal{B}$. After the intervention, the invasive species continues to spread until time $T > \tau$, but without the proliferative influence of the individuals eradicated at time $\tau$. Our goal is to find a feasible intervention plan that minimizes the degree to which the landscape is affected by the invasion.

### 2.3 Background: Multivariate Hawkes Processes

A multivariate Hawkes process can be thought of as a spatiotemporal point process, a random collection of points representing the time and location of events. An $n$-dimensional point process can be described by a counting process $\mathcal{N}(t) = (\mathcal{N}_1(t), \cdots, \mathcal{N}_n(t))^\top$ where $\mathcal{N}_i(t)$ is the number of events occurring at location $i$ before time $t$. The behavior of the process is characterized by the conditional intensity $\lambda(t)$. Given the history of the process up to time $t$, $\mathcal{H}_{t-}$, the expected number of events in a small time window $[t, t + dt)$ is given by $\mathbb{E}\left[d\mathcal{N}(t)|\mathcal{H}_{t-}\right] = \lambda(t)dt$.

Hawkes processes model self-exciting phenomena in which the occurrence of events causes additional events to be more likely, such as social media posts spurring reposts [11], earthquake aftershocks inducing further aftershocks [18],and in this work, an invasive species individual causing another individual to appear at the same or other nodes. This self-exciting behavior is modeled using a history-dependent intensity of the form:

$$\lambda_i(t) = \mu_i(t) + \sum_{e:t_e<t} \phi_{ij_e}(t, t_e) \tag{2.1}$$

$$= \mu_i(t) + \sum_{j=1}^{n} \int_0^t \phi_{ij}(t, s) d\mathcal{N}_j(s) \tag{2.2}$$

$\phi_{ij}(t, s)$ is called the impact function and captures the temporal influence of an event at location $j$ at time $s$ on the occurrence of events at location $i$ at time $t \geq s$. Here, the first

term $\mu_i(t)$ is the exogenous event intensity, from outside the network and independent of the history, and the second term $\sum_{e:t_e<t} \phi^{ij_e}(t, t_e)$ is the endogenous event intensity, modeling influence and interaction within the network. Defining $\Phi(t, s) = [\phi_{ij}(t, s)]_{i,j=1...n}$, $\lambda(t) = (\lambda_1(t), \ldots, \lambda_n(t))^\top$, and $\mu(t) = (\mu_1(t), \ldots, \mu_n(t))^\top$, we can compactly rewrite Eq (Equation 2.1) in matrix form:

$$\lambda(t) = \mu(t) + \int_0^t \Phi(t, s) d\mathcal{N}(s) \tag{2.3}$$

A common choice of impact function is the truncated exponential function $\Phi(t, s) = Ae^{-\omega(t-s)} \cdot \mathbf{1}_{\geq 0}(t - s)$ where $\phi_{ij}(t, s) = a_{ij}e^{-\omega(t-s)} \cdot \mathbf{1}_{\geq 0}(t - s)$. The coefficient $a_{ij}$ represents the strength of the influence of $j$ on $i$, and the influence of an event that occurs at time $s$ is 0 before $s$ and decays off after $s$ (e.g. a social media post becomes less relevant, an infected person becomes less contagious, or an invasive species becomes less likely to survive and reproduce).

### 2.4 Predicting Invasive Species Spread with Hawkes Processes

Traditionally, the spread of invasive species is modeled using a combination of differential equations describing the population dynamics and a dispersal kernel describing the displacement of individuals [19], or their stochastic counterparts [20]. More recently, studies have demonstrated the potential of point process models [13, 21] for characterizing the spatial distribution of invasive plant species.

We present a multivariate Hawkes process model for the dynamics of invasive species spread. To formulate it as a network diffusion process, we model the $n$ land units in landscape $L$ as nodes $V$ in a directed graph, with edges between nodes that are close enough for dispersal to occur between the corresponding land units. The appearance of a new invasive individual in node $i$ at time $s$ is denoted by the invasion event $(i, s)$. Indexing invasion events by $e$, the history of the network diffusion process up to immediately before some

time $t$ is the set of events $\mathcal{H}_{t-} := \{(i_e, s_e) | s_e < t\}$.

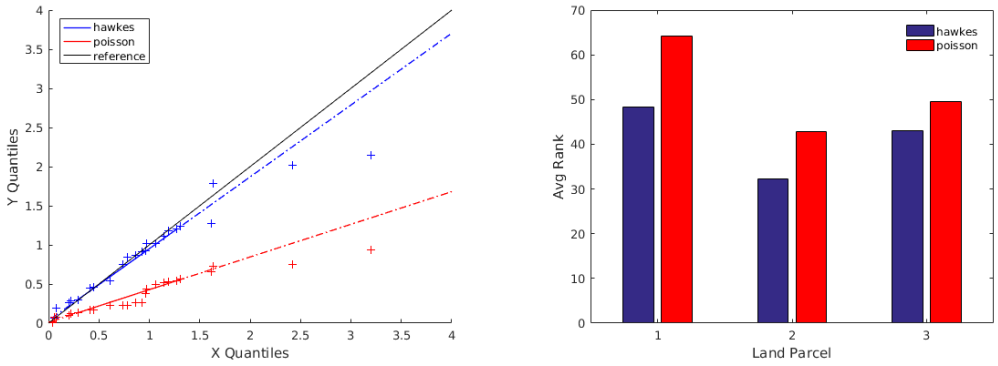
Invasive species can be introduced at any time by carriers like wind, animals or humans. These arrivals are called *exogenous* invasions, and their rate can vary spatially depending on landscape features or human activity. The instantaneous rate at which individuals are introduced to node $i$ at time $t$ is denoted by $\mu_i(t)$, and represents the probability of an exogenous invasion event in a small time window $[t, t + dt)$. Once an invasive individual has become established, it survives for an average lifetime $\beta$. Since many invasive species mature early and have short life expectancy [22], we assume an individual born at $s_e$ faces a constant risk of death $\omega = \frac{1}{\beta}$, so that the probability of the individual surviving until time $t$ is given by the survival function $e^{-\omega(t-s_e)}$. While the individual survives, it initiates *endogenous* invasions, e.g. by releasing offspring. The likelihood of an individual appearing at location $i$ due to the dispersal of the offspring of an individual at location $j$ depends on edge weight $a_{ij}$ between the two nodes, which can be, e.g., a decaying function of distance between them [23].

All these effects together influence the rate at which new individuals appear in a given node $i$ at time $t$, or the intensity $\lambda_i(t)$. This represents the conditional probability of observing an invasion event in a small time window $[t, t + dt)$ given the history $\mathcal{H}_{t-}$.

$$\lambda_i(t) = \mu_i(t) + \sum_{(j_e, s_e) \in \mathcal{H}_{t-}} a_{ij_e} \cdot e^{-w(t-s_e)} \tag{2.4}$$

The first term $\mu_i(t)$ is the rate of exogenous invasion events at node $i$, and the summation term captures the contribution of past invasion events $(j_e, s_e)$ in the network towards endogenous invasions in node $i$ at time $t$.

We have verified the applicability of this modeling framework using data about the encroachment of *A. grandis* trees into montane meadows at Bunchgrass Ridge in Oregon [24]. Notably, we find that a Hawkes model explains the observed data better than a Poisson process (Figure 2.1), suggesting that there are excitatory interactions in the invasion process.




(a) QQ-plot

(b) Prediction task results.

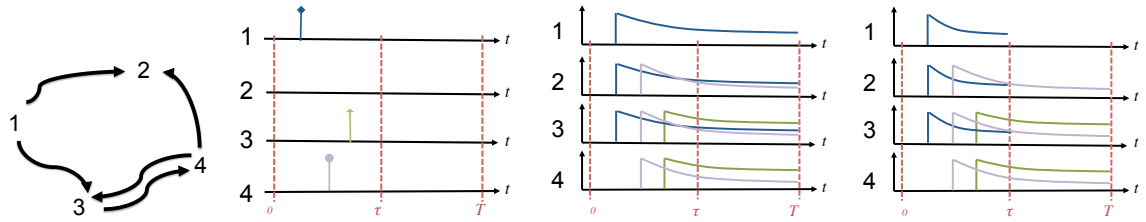
Figure 2.1: (a) Residual analysis for randomly chosen location in Bunchgrass Ridge. The events are better fit by a Hawkes process (slope closer to 1) than a Poisson process. (b) Hawkes process model has lower error when predicting the location of the next invasion.

### 2.4.1 Control Objectives

Given the graph representing our landscape and the invasion process dynamics described above, we can quantify the degree to which the landscape is affected by the invasive species spread at the end of our planning horizon $T$ in a number of ways. One reasonable goal is to minimize the rate of invasions at time $T$ subject to an intervention $u$ at time $\tau$, captured by $\lambda(t; u)$. Since $\lambda(t; u)$ depends on events that will stochastically occur between $\tau$ and $t$, it will vary across different realizations of the stochastic process, so instead we aim to minimize the total *expected* intensity at time $T$. Let $\eta_i(t; u) = \mathbb{E}\left[\lambda_i(t; u)\right]$, where the expectation is taken over all possible realizations of the process.

**Given:** A graph $G(V, E)$ representing landscape $L$, edge weights $A$ with $a_{ij} = 0$ for $(i, j) \notin E$ and $a_{ij} > 0$ for $(i, j) \in E$, intervention time $\tau$ and finite time horizon $T$, intervention costs $c_i$ for each node $i \in V$ and budget $\mathcal{B}$.

**Find:** A feasible intervention plan consisting of nodes $U \subseteq V$ such that $\sum_{i \in U} c_i \leq \mathcal{B}$, that minimizes $\sum_{i \in V} \eta_i(T; u)$.




(a) Sample network. (b) Events before $\tau$. (c) Intensity from events. (d) Effect of event removal at node 1.

Figure 2.2: A sample network and event history up to $\tau$. Each event contributes to the intensities at the event's node and its neighbors. If the event at node 1 is deleted at $\tau$, its contribution to intensities for $t > \tau$ disappears.

Another plausible goal is to minimize the *total expected number* of invasions that occur until time $T$, since the ecological damage resulting from invasions is often a function of the population size. We cannot affect the process until $\tau$, so this amounts to minimizing the number of invasions in the interval $[\tau, T)$. We store the number of invasion events at each node over time using an $n$-dimensional counting process where $\mathcal{N}_i(t; u)$ represents the number of invasive species individuals that have appeared in cell $i$ by time $t$. Then, given the same inputs as before,

**Find:** A feasible intervention plan consisting of nodes $U \subseteq V$ such that $\sum_{i \in U} c_i \leq \mathcal{B}$, that minimizes $\sum_{i \in V} \mathbb{E}\left[\mathcal{N}_i(T)\right]$.

## 2.5 Discrete Interventions in Hawkes Processes

Given a network diffusion process starting at time $t_0 = 0$, suppose we plan to perform a management action at time $\tau > t_0$ to steer the diffusion process over the network towards some objective at an arbitrary time $T > \tau$.

Since the intensity at any time $t'$ only depends on the history of events up to time $t'$, we can also define the *state* at any time $t'$ as $y(t') := \int_0^{t'} e^{-\omega(t'-s)}\, d\mathcal{N}(s)$, capturing the current effect of all events that have happened at each node up to time $t'$. Then considering the

time of intervention $\tau$,

$$\begin{aligned}\lambda(t) &= \mu(t) + Ay(t) = \mu(t) + \int_0^t Ae^{-\omega(t-s)} d\mathcal{N}(s) \\ &= \mu(t) + \underbrace{\int_0^\tau Ae^{-\omega(t-s)} d\mathcal{N}(s)}_{\text{events before } \tau} + \underbrace{\int_\tau^t Ae^{-\omega(t-s)} d\mathcal{N}(s)}_{\text{events after } \tau} \\ &= \mu(t) + Ae^{-\omega(t-\tau)} y(\tau) + \int_\tau^t Ae^{-\omega(t-s)} d\mathcal{N}(s)\end{aligned}$$

Our management action entails the deletion of all events at a given set of nodes $U$ (see Figure 2.2). This can alternatively be thought of as resetting the state of those locations to 0 at time $\tau$. Therefore, for $t > \tau$ we have the intervention-dependent intensity:

$$\begin{aligned}\lambda(t; u) = \mu(t) &+ Ae^{-\omega(t-\tau)}(u \circ y(\tau)) \\ &+ \int_\tau^t Ae^{-\omega(t-s)}\, d\mathcal{N}(s; u)\end{aligned} \tag{2.5}$$

where $\circ$ denotes element-wise product. Vector $u$ encodes our management action (intervention) where $u_i = 0$ indicates removing the history at node $i$ and $u_i = 1$ means not intervening at $i$.

### 2.5.1 Expected Behavior After Intervention

We now derive closed-form expressions for our control objectives in terms of the expected intervention-dependent intensity $\eta(t; u)$. The first objective of interest is to minimize the sum of expected rate of events at our target time: $\sum_i \eta_i(T; u)$.

By the superposition theorem of point processes, the process $\mathcal{N}(t; u)$ is decomposed into two independent processes:

$$\mathcal{N}(t; u) = \mathcal{N}_e(t; u) + \mathcal{N}_h(t; u)$$

$\mathcal{N}_e(t, u)$ is the counting process for events caused by the exogenous intensity from $\tau$ to $t$,

and $\mathcal{N}_h(t;u)$ comprises the events generated due to the effect of previous events (history) before $\tau$. Each of these processes have associated intensities $\lambda_e(t;u)$ and $\lambda_h(t,u)$:

$$\lambda_e(t;u) = \mu + \underbrace{\int_\tau^t Ae^{-\omega(t-s)}\, d\mathcal{N}_e(s;u)}_{\text{from new events generated by } \mu} \tag{2.6}$$

$$\lambda_h(t;u) = \underbrace{Ae^{-\omega(t-\tau)}(u \circ y)}_{\text{from events before } \tau} + \underbrace{\int_\tau^t Ae^{-\omega(t-s)}\, d\mathcal{N}_h(s;u)}_{\text{from new events generated by history}} \tag{2.7}$$

Correspondingly, we have their expected values $\eta_e(t;u) = \mathbb{E}[\lambda_e(t;u)]$ and $\eta_h(t;u) = \mathbb{E}[\lambda_h(t;u)]$. For $\eta_e(t;u)$, we can write:

$$\eta_e(t;u) = \mu + \mathbb{E}\left[\int_\tau^t Ae^{-\omega(t-s)}\, d\mathcal{N}_e(s;u)\right] \tag{2.8}$$

$$= \mu + \int_\tau^t Ae^{-\omega(t-s)}\, \eta_e(s;u) ds \tag{2.9}$$

Using Theorem 1 from [15], $\eta_e(t;u) = \Psi(t)\mu$ is a solution to Equation Equation 2.9 if and only if $\Psi(t) = I + \int_0^t Ae^{-\omega(t-s)}\Psi(s)ds$. For our choice of impact function:

$$\Psi(t) = I + A(A - \omega I)^{-1}(e^{(A-\omega I)t} - I) \tag{2.10}$$

Intuitively, $\Psi(t)$ is a matrix function indexed by $i, j$ which are nodes. $\Psi_{i,j}(t)$ can be interpreted as the total contribution of possible invasions at node i at time $t$ from events at j at any time before $t$ (directly and indirectly).

Additionally, according to Theorem 3 in [15], by using integration by parts and the Laplace transform of point processes from [11], we can show that $\eta_h(t;u) = \Xi(t-\tau)A(u \circ y)$ where $\Xi(t) = e^{(A-\omega I)t}$. Putting these two together we have the analytical form for our first objective:

$$\mathbb{E}[\lambda(t;u)] = \Psi(t)\mu + \Xi(t-\tau)A(u \circ y) \tag{2.11}$$

For the second objective we aim to minimize the total expected number of events in all nodes, $\sum_i \mathbb{E}[\mathcal{N}_i(T;u)]$:

$$\mathbb{E}[\mathcal{N}_i(T;u)] = \mathbb{E}[\int_0^T d\mathcal{N}_i(s;u)] = \int_0^T \eta(s;u)ds \tag{2.12}$$

Therefore, if we define $\Gamma(t) = \int_0^t \Psi(s)ds$ and $\Upsilon(t) = \int_0^t \Xi(s)ds$ we have:

$$\mathbb{E}[\mathcal{N}(t;u)] = \Gamma(t)\mu + \Upsilon(t-\tau)A(u \circ y). \tag{2.13}$$

Intuitively, $\Gamma_{i,j}(t)$ is the cumulative invasion from $i$ to $j$ up to time $t$.

In summary we have;

$$\mathbb{E}[\lambda(T;u)] = \Psi(T)\mu + \Xi(T-\tau)A(u \circ y)) \tag{2.14}$$

$$\mathbb{E}[\mathcal{N}(T;u)] = \Gamma(T)\mu + \Upsilon(T-\tau)A(u \circ y) \tag{2.15}$$

where

$$\Xi(t) = e^{(A-\omega I)t} \tag{2.16}$$

$$\Psi(t) = I + A(A-\omega I)^{-1}(e^{(A-\omega I)t} - I) \tag{2.17}$$

$$\Upsilon(t) = (A-\omega I)^{-1}(e^{(A-\omega I)t} - I) \tag{2.18}$$

$$\Gamma(t) = It + A(A-\omega I)^{-1}(\Upsilon(t) - It) \tag{2.19}$$

### 2.5.2 Optimization Formulations

Given the closed forms we have derived for the expected behavior of the network diffusion process after intervention, we can define our first optimization problem as:

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & \sum_i \Psi(T)\mu + \Xi(T-\tau)A(u \circ y) \\ \text{subject to:} \quad & \sum_i (1-u_i)c_i \leq \mathcal{B}, \\ & u_i = \{0,1\} \forall i \in \{1,2,\ldots,n\} \end{aligned} \tag{2.20}$$

where $c_i$ and $\mathcal{B}$ are defined as before.

Similarly, our second objective is:

$$\begin{aligned} \underset{u}{\text{minimize}} \quad & \Gamma(T)\mu + \Upsilon(T-\tau)A(u \circ y) \\ \text{subject to:} \quad & \sum_i (1-u_i)c_i \leq \mathcal{B}, \\ & u_i = \{0,1\} \forall i \in \{1,2,\ldots,n\} \end{aligned} \tag{2.21}$$

The dependence on our control variable, $u$, is linear and we can incorporate effective binary optimization techniques to find the optimal intervention plan. We used the mixed integer linear programming solver offered through the `intlinprog` function in MATLAB 2016b.

### 2.5.3 Heuristic Interventions

Besides the optimized recommendations for intervention nodes, it is also possible to choose nodes on the basis of a number of heuristics that are very natural to multivariate Hawkes process models. In each case, nodes are considered in decreasing order of a heuristic criterion, and we greedily build a set of intervention nodes $U$ by adding each successive node as long as there are events to remove there and the cost of intervening at the node can be covered with our remaining budget. We consider the following heuristic criteria:

- **Exogenous intensity ($\mu_i$):** nodes with the highest rate of exogenous events.
- **Number of events until $\tau$ ($\mathcal{N}_i(\tau)$):** nodes with the highest number of events in the observation window.
- **Intensity due to global events at $\tau$ ($\lambda_i(\tau)$):** nodes with the highest intensity at the intervention time due to both local events and events in neighboring nodes.
- **Intensity due to local events (state) ($y_i(\tau)$):** nodes with the highest intensity due to only local events.

Interestingly, many of these criteria have analogs in invasion biology and control strategies based on them have been proposed. For example, exogenous intensity is comparable to "propagule pressure", which is believed to be an important determinant of whether non-native species successfully invade new habitats [25]. Density-based eradication (related to $\mathcal{N}_i(\tau)$) and strategies that balance the density and fecundity of the population (related to $y_i(\tau)$) have also been widely studied [26].

## 2.6 Experiments

We validate our discrete intervention methodology by testing the behavior of our analytical expressions on simulated ground-truth Hawkes process cascades, and analyze the runtime performance of our intervention optimization approach. We then apply our methodology to compare strategies for controlling the spread of a real invasive species through a landscape.

### 2.6.1 Validation of Analytical Expressions & Scalability

To verify our derived closed-form expressions against some ground-truth network diffusion processes and to test the scalability of our optimization approach, we generate synthetic networks with known parameters. The exogenous intensity at each node $\mu_i$ is constant over time and is a uniformly distributed random variable in $[0, \mu_{max}]$ with $\mu_{max} = 0.02$. A small

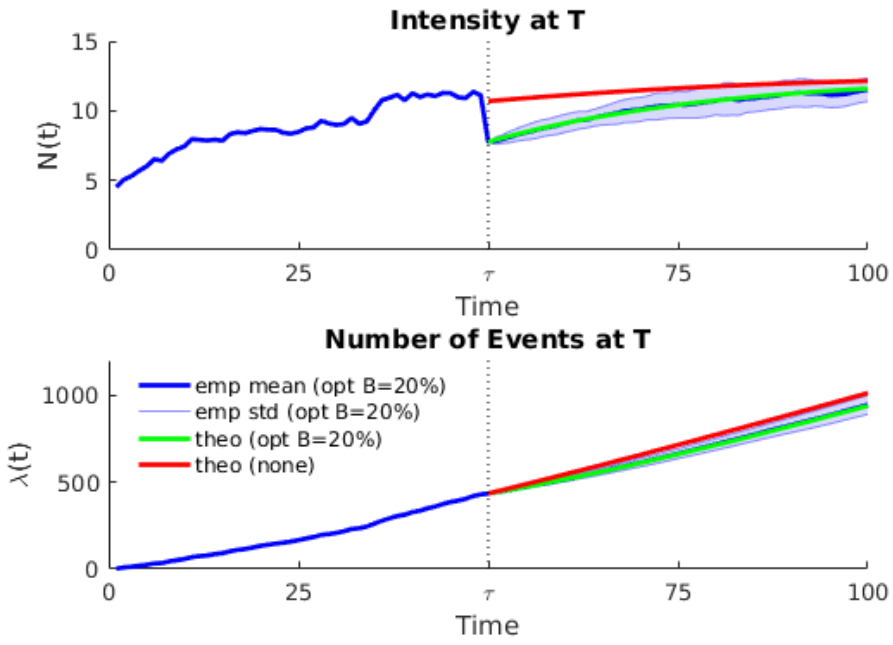



Figure 2.3: Simulated event cascades after an optimal intervention with $\mathcal{B} = 0.20 \cdot \mathcal{B}_{all}$ on a synthetic network compared to analytically computed intensity and number of events for the optimal plans.

number of nodes have a higher exogenous intensity and act as seed points in the diffusion process. We assign the mutually-exciting parameters $a_{ij}$ to be values in the range $[0, a_{max}]$, with $a_{max} = 0.05$. Finally, we set the influence decay rate $\omega = 0.15$ and a finite time horizon of $T = 100$.

First, we empirically evaluate the closed-form expressions for our intervention objectives $\mathbb{E}[\lambda(T)]$ and $\mathbb{E}[\mathcal{N}(T)]$. To do this, we simulate a single realization of an invasion cascade up to time $\tau = 50$, implement a fixed intervention $u$ and simulate many realizations of the subsequent cascade from time $\tau$ to $T$ with which we compute the empirical mean intensity and number of invasions at each time $\tau < t \leq T$. We compare these to the theoretical expected intensity and number of invasions computed using Equations Equation 2.14 and Equation 2.15, following the same intervention $u$. The results are shown in Figure 2.3. The theoretically computed values closely match the observed empirical mean values for both quantities.

Next, we test our discrete intervention optimization approach on networks of increasing

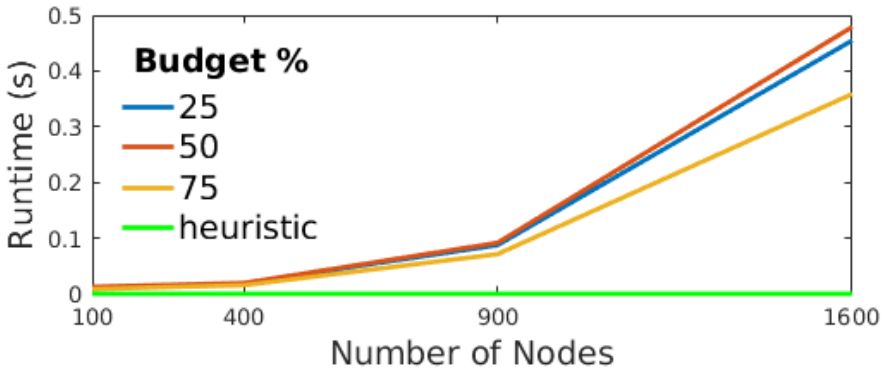



Figure 2.4: Runtime of ILP versus network size, compared to the time taken to compute intervention locations using a heuristic. The reported time is the average over 20 randomly simulated cascades.

size. The intervention cost $c_i$ at each node is set to a fixed unit cost plus a cost proportional to the number of events there at time $\tau$. We simulate multiple realizations of event cascades from $t = 0$ to $\tau$ and in each case compute $\mathcal{B}_{all}$ the cost of removing all events that have appeared in the network by $t = \tau$. We set the intervention budgets $\mathcal{B}$ as fixed percentages of $\mathcal{B}_{all}$ to allow comparisons between the different realizations. Due to the closed-form expressions for the expected effect of an intervention action, the optimization problems in Equations Equation 2.20 and Equation 2.21 take the form of knapsack problems, for which MIP solvers are highly effective. Figure 2.4 shows that these integer programs are surprisingly fast to solve and scale relatively well for the network sizes we consider here.

### 2.6.2 Invasives Management with a Limited Budget

We apply our discrete intervention optimization for multivariate Hawkes processes in a real-world setting to model the potential impact of different management strategies on *A. grandis* encroachment in Bunchgrass Ridge. We find that a multivariate Hawkes process is able to explain the invasive spread of this species better than a homogeneous Poisson process model. We use a Hawkes process model fit to these data and assign intervention costs for each land parcel in the same manner as in the synthetic networks to examine alternatives for budget-constrained invasive species control.

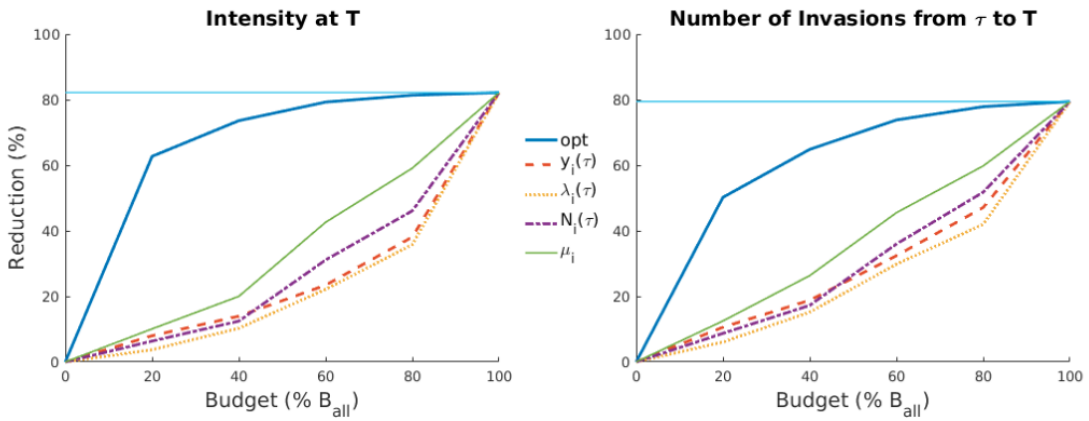



Figure 2.5: Reduction in invasion intensity and number of new invasions wrt to no intervention for Bunchgrass Ridge. The maximum possible reduction is obtained by intervening at all locations with invasions at time $\tau$.

In order to study the impact of budget restrictions on the effectiveness of invasive species management, we vary the intervention budget $\mathcal{B}$ available from 20% to 80% of $\mathcal{B}_{all}$. Results are shown in Figure 2.5. We observe that the optimized plans attain close to the same level of control as eradicating all individuals at time $\tau$, but at only 60-80% of the cost, indicating that our method for optimal invasive management has the potential to deliver significant cost savings.

Our analytical expressions allow us to compute the expected benefit of different hypothetical management actions, including the ones proposed by the heuristic strategies. Having our optimal solutions as a baseline, the results in Figure 2.5 indicate that the heuristic approaches considered here are highly sub-optimal, especially in low-budget settings. Figure 2.6 shows the % optimality gap of the heuristic management strategies relative to the optimal plan. The best performing heuristic approach for minimizing both the rate and the number of new invasions was the one based on the exogenous intensity $\mu_i$, suggesting that monitoring the processes by which *A. grandis* is introduced to the Bunchgrass Ridge region could be an effective approach to curb its invasive spread.

By visualizing the spatial distribution of intervention locations (omitted), we find that the optimal intervention plans for minimizing each objective are different from one another. In particular, it appears that minimizing intensity focuses intervention effort at relatively

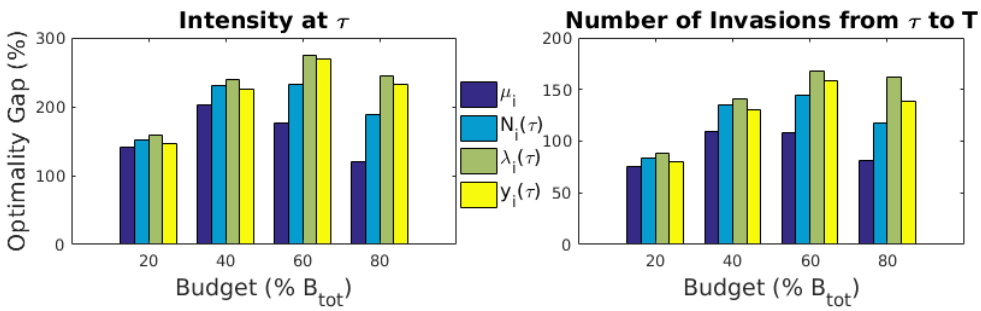



Figure 2.6: Mean optimality gap % for the heuristic strategies.

few core invasion hotspots whereas minimizing the total number of invasions targets more peripheral locations. This suggests there are possible trade-offs that may be of interest to conservation planners developing long-term strategies for invasive species management.

## 2.7 Conclusions

We develop a method for characterizing the effect of a novel discrete intervention in the context of spatiotemporal Hawkes processes. This intervention consists of removing the effect of selected events in the history of the Hawkes process. We derive closed-form expressions for the effect of such an intervention action on the expected intensity and expected number of events in the network at the end of a finite time horizon $T$. We build on this framework to obtain optimal intervention plans employing this type of intervention mechanism, and apply this approach to the problem of controlling the spread of invasive species through a landscape. Our results on a real-world tree invasion dataset suggest that optimized intervention plans obtained using our approach can achieve cost-effective control, and also provide a benchmark against which other intervention plans that are used in practice can be evaluated.

The proposed model showcases the potential of Hawkes process in managing invasive species propagation, and it can be extended in many ways to further cope with realistic settings. It can be easily generalized to handle multiple intervention points and combined with Markov Decision Processes for sequential decision making. Moreover, in practice all

species may not be successfully removed after an intervention, with success probability depending potentially on the location, the species itself, budget, or time. This will add a thinning process layer to the Hawkes process. In addition, the methodologies we have developed for the invasive species management problem can be applied to other problem domains, such as in limiting the spread of disease or misinformation in social networks.

# CHAPTER 3

# BUDGET-CONSTRAINED COMBINATORIAL OPTIMIZATION FOR NETWORK FLOWS: DISASTER-RESILIENT ROAD INFRASTRUCTURE

## 3.1 Introduction

Developing and maintaining resilient infrastructure systems is a key strategy for several UN sustainable development goals. Goal 13 (Climate Action) calls on countries to prepare for climate-related hazards through national strategies and planning. Even more explicitly, Goal 11 (Sustainable Cities and Communities) emphasizes that transportation and other basic services should be expanded following policies aimed at "resource efficiency, mitigation and adaptation to climate change" [27] in order to reduce deaths and economic losses caused by disasters, especially water-related disasters. These appeals are not without justification. Numerous recent studies have assessed how extreme weather, rising sea levels and altered temperature and precipitation regimes can damage essential infrastructures such as transportation [28], power [29], water [30] and sewage, or cause cascading failures across these systems [31]. In fact, the cost of damages to infrastructure due to disasters is expected to increase dramatically–e.g. tenfold in the European Union by the end of this century [32]–in part due to changing climate patterns but also largely driven by population and GDP growth [33].

Among all natural disasters, floods are the most frequent, widespread, and expensive. Direct flood damages in the US average $9bn each year [34], including critical public infrastructures such as highways, roads, bridges, and utilities. Floods affect road infrastructure by: inundating roads, causing traffic stagnation and stoppage; depositing silt or other material on roads that must be cleared; destabilizing road foundations, requiring repair; and in extreme cases completely destroying roads which must then be rebuilt [35]. These

impacts are of particular concern not only due to the key role mobility plays in evacuation, emergency response and disaster relief distribution, but also due to its importance to normal socioeconomic activity. For example, roads between the Mozambican capital Maputo and the rest of the country remained unusable for nearly a year after devastating floods in 2000, causing economic growth to come to a halt [28].

In order to avoid such severe outcomes, decision-makers must be able to first assess how flooding threatens road-based transportation. Spatial flood risk assessments conventionally take account of three components of risk [36]:

- **hazard**: the location, severity and frequency of flood events
- **exposure**: the populations and assets present where flooding occurs
- **vulnerability**: the degree to which losses are suffered as a result of exposure to the hazard

Several recent studies have examined how flood hazard and exposure interact to jeopardize road infrastructure by spatially intersecting flood hazard maps with road networks and reporting the total length of roadways within flood zones [37, 38, 39]. Although this gives us a measure of the potential for direct damages to roadways, this metric does not capture the degree of functional loss the road network may see in terms of reduced connectivity between locations [40]. The potential for functional losses partly depends on the network's topology: graph theoretic indicators such as edge or node centrality [41] can be used to identify key road segments or junctions, and metrics like graph clustering coefficients can quantify overall network connectivity [42]. However, the magnitude of impact on mobility also depends on the travel volume along the roads exposed to flooding. If data about the volume of trips taken between different origins and destinations is available, the impacts of road closures on mobility can be estimated, e.g. in terms of the percentage of attempted trips that can be successfully completed. In Section 3.2, we present an evaluation pipeline for quantifying the *mobility* impacts of flooding, using mobility data gleaned

from call detail records and from traffic simulations between traffic analysis zones (TAZs). Unfortunately in practice, comprehensive data on origin-destination flows are not widely available, calling for the use of predictive models for trip generation and distribution in disaster-resilient infrastructure planning. In Section 3.3, we briefly introduce some predictive models for origin-destination flow volumes that can be used when mobility data are unavailable.

Once decision-makers are able to anticipate which road network components may be inundated or damaged and how the unavailability of these road segments will impact mobility, they must then develop cost-effective strategies for allocating resources towards mitigating these threats. The associated network design, planning and scheduling problems have also inspired AI researchers to develop effective and scalable techniques that can be applied to these critical, real-world problems. There has been a great deal of research on variations of the pre-disaster transportation network preparation problem [43], e.g. to strategically upgrade roads such that evacuation paths are protected [44], or such that the average travel time of emergency response vehicles to service points is minimized [45]. There is also a growing body of work in post-disaster road network restoration, where the goal is to optimize the order in which to clear roads and the positioning of equipment to enable evacuations [46] or distribute emergency supplies [47]. However, relatively little work has tackled the goal of fortifying road networks such that *normal* travel flows remain feasible. In Section 3.4 address this gap in the transportation resilience literature and focus on the problem of finding a pre-disaster road fortification plan that ensures that a regional population's typical travel needs are still met as much as possible under likely disaster scenarios.

## 3.2 Data-Driven Vulnerability Analysis for Road Networks

Floods are the most frequent, widespread, and expensive natural disasters in the United States [48]. [33] estimate that $5.5 trillion of built assets in the contiguous USA currently

face exposure to a 1-in-100 year flood, and [38] identify cities like Key Biscayne, FL, Hoboken, NJ, Redwood City, CA and Cambridge, MA as likely to experience sudden increases in road flooding in the next few decades. In the following sections, we analyze potential flooding impacts on mobility in 3 major metropolitan areas in the United States that face significant flood risk: Washington D.C., Chicago, and Seattle.

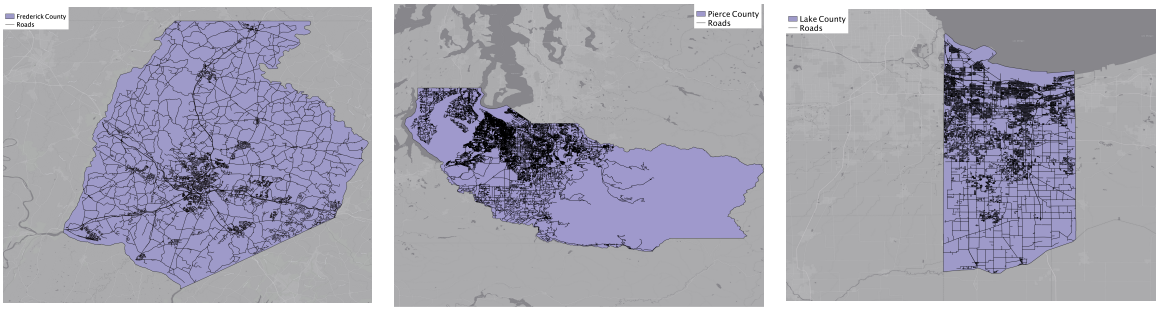
At the same time, flood damage to infrastructure is also a serious concern for developing countries, where investment in energy, water, communication and transport infrastructure is central to socioeconomic development. For instance, the estimated costs of road repair and maintenance across Africa under current climate change projections exceed $150 billion, which will significantly divert funding from initiatives for expansion and growth [49]. Thus, climate change will exacerbate existing socioeconomic vulnerabilities and threaten the success of crucial development schemes unless steps are taken to proactively mitigate these costs. In the following sections, we also assess flood mobility impacts in Senegal, where nearly every year major flood events result in the displacement of hundreds of thousands of people.

### 3.2.1 Data

The first step in a framework for making investment recommendations for climate-resilient road infrastructure is analyzing the exposure of roads to flooding, and the impacts flooding will have on population mobility. In order to do so, we require three main pieces of data:

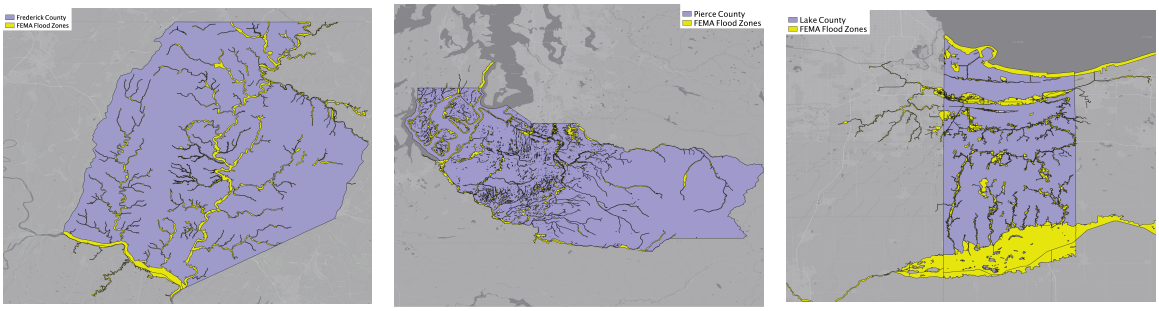
1. A GIS representation of the road network
2. A spatially-explicit model of flood risk
3. Mobility data describing travel demand between different regions

Fortuitously, geospatial data made available by projects like OpenStreetMap [50] can be freely used to obtain graph representations for road networks across the globe. It is



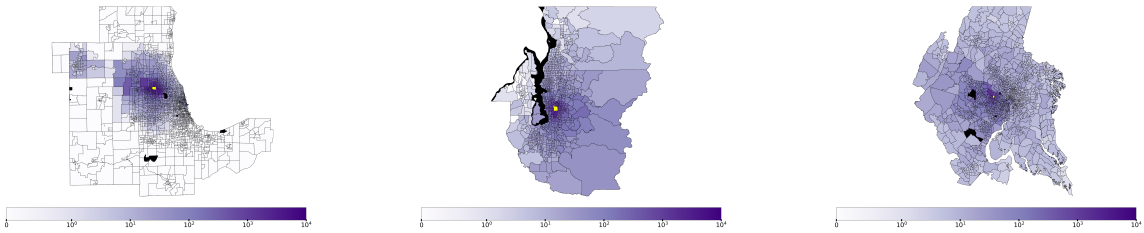

(a) Frederick County, MD. (b) Pierce County, WA. (c) Lake County, IN.

Figure 3.1: OpenStreetMap road networks for focal counties near Washington D.C., Seattle, and Chicago.




(a) Frederick County, MD. (b) Pierce County, WA. (c) Lake County, IN.

Figure 3.2: FEMA flood zones for for focal counties near Washington D.C., Seattle, and Chicago.



(a) Chicago (b) Seattle (c) Washington D.C.

Figure 3.3: Number of outgoing trips from a single origin census tract (yellow) to every other census tract in each study area; black census tracts have no population.

much more challenging to obtain publicly available flood maps. For locations within the United States of America, the Federal Emergency Management Agency provides shapefiles specifying flood zones and their types [51]. The Fathom Global dataset [52, 53] provides flooding data for the entire globe at the resolution of approximately $90\text{m}^2$. Specifically, this dataset provides flood depth rasters for floods of different severities characterized by return period $\lambda$, which is the estimated time interval between flooding events of a similar intensity. Intuitively, a 500-year flood is more severe and less likely to occur than a 100-year flood. Given a return period $\lambda$, the value of each cell of the flood depth raster is the maximum flood depth estimated by a hydrodynamic model for a flood of the specified severity. There are alternative freely-available data sources for flood mapping. For example, the NASA MODIS near real-time global flood mapping project [54] also provides flood data with global extent, although at lower spatial resolution compared to the Fathom Global dataset.

Travel demand data can also be challenging to obtain. For some geographies, mobility data are available in the form of origin-destination matrices storing the average number of trips between different traffic analysis zones (TAZs), which are special spatial units used by transportation officials to track traffic-related data. We obtain publicly available TAZ shapefiles for Chicago, Seattle, and Washington D.C. [55, 56, 57], as well as OD matrices containing the number of trips from each TAZ to every other TAZ in a given day. Geospatial data analyses more commonly use census tracts or blocks as the geographic units of analyses, so we resample the OD flow data between TAZs to estimate OD flows

between census tracts (Figure 3.3) using the following formula:

$$T_{ij} = \sum_a \sum_b T_{ab} p(a,i) p(b,j) \tag{3.1}$$

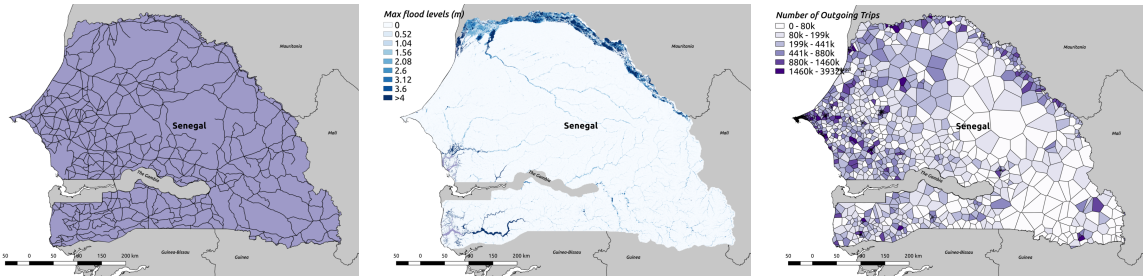
where $a$, $b$ are TAZs, $i$, $j$ are census tracts, and $p(a,i)$ denotes the proportion of area of TAZ $a$ that overlaps with census tract $i$, and $p(b,j)$ similarly. In other words, the trips to and from a TAZ are assumed to be evenly distributed across its area.

Call detail record (CDR) data, consisting of time-indexed sequences of cell towers used by anonymized users, is an excellent source of ground-truth human mobility data. We obtained CDR data for Senegal provided by Orange [58], the biggest mobile provider in Senegal with $N = 1,666$ cell towers across the country, through the UN Data for Climate Action Challenge [59]. Since CDR data record user locations only in terms of cell towers used, we use the approximate latitude-longitude locations of the $N$ cell towers to construct "cell tower zones", Voronoi regions containing all locations that are closer to a given cell tower than to any other tower (Figure 3.4(c)). Given the sequence of cell tower zones visited by each user in the dataset, e.g. $\{l_1, l_2, \cdots, l_K\}$ for a customer that moves through $K$ zones, we consider each consecutive pair of cell tower zones $(l_k, l_{k+1})$ to be a trip from zone $l_k$ to zone $l_{k+1}$. We obtain the inter-zone travel demand by constructing an $N \times N$ origin-destination (OD) zone trip matrix $\mathbf{T}$ whose $(i,j)$-th entry represents the total number of trips taken from cell tower zone $i$ to zone $j$ by all users in the dataset.

### 3.2.2 Vulnerability Analysis

We first construct an undirected graph representation of the road network, with edges $E$ representing road segments and vertices $V$ representing intersections or endpoints of these road segments. Each edge has a distance property representing the length of the corresponding road segment.

To estimate the exposure of each road segment to flooding in a given return period sce-




(a) Senegal roadway system. (b) 100-year flood depths. (c) Outbound trips per zone.

Figure 3.4: Data sources for estimating infrastructure exposure to flooding and potential impacts on mobility.

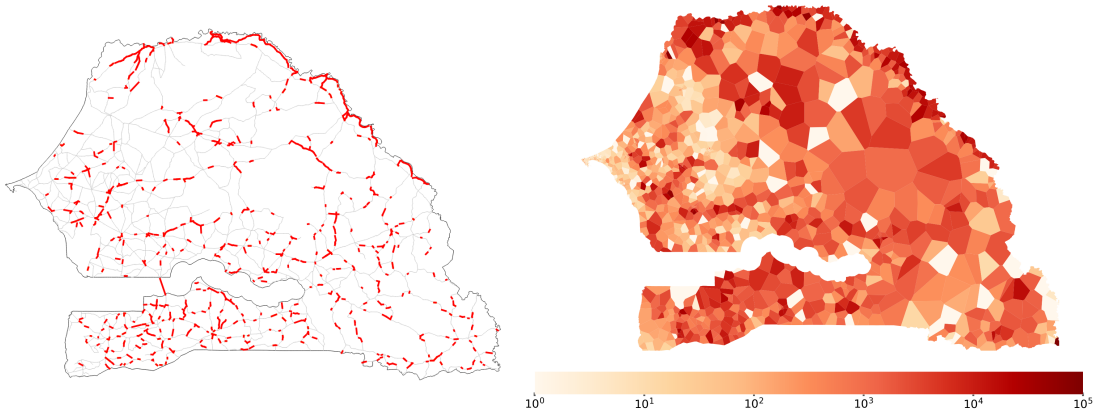



Figure 3.5: Impacts of the $\lambda = 100$ flooding scenario and threshold $\theta = 1$m. The **left** figure shows the road network of Senegal, with red links indicating roads that are impassable due to flooding. The **right** figure shows the estimated number of outgoing trips per zone that cannot be completed due to the damages in the road network.

nario, we first apply a threshold $\theta$ to the flood depth raster and then obtain the geometric intersection of roads with those raster cells with flood depth value $\geq \theta$. Edges corresponding to road segments that pass through any such cells form a subset $E_F^\lambda \subseteq E$ of flooded roads. For each edge in this set we compute the flooded distance to estimate the length of the road segment that would need to be upgraded to make the road segment traversable again. We also obtain an *unflooded* subgraph $G_U^\lambda = (V, E_U^\lambda)$ consisting of the original set of vertices $V$ and any edges corresponding to road segments not affected by flooding, representing the parts of the road network that are still traversable. Note, the two edge sets have no edges in common (i.e. $E_F^\lambda \cap E_U^\lambda = \emptyset$) and $E_F^\lambda \cup E_U^\lambda = E$.

In order to connect the flooding effects on the road network to the zone-to-zone demand data, we assume that trips between pairs of zones occur between random origin nodes and destination nodes within each zone. Specifically, for each zone $i$, we calculate $Z_i$, the set of vertices from the road network that are within its boundaries. Then, we assume that a trip leaving zone $i$ and arriving at zone $j$ is equally likely to start from any vertex within $Z_i$, and similarly equally likely to end at any vertex in zone $Z_j$. For a given road network $G^\lambda = (V, E_U^\lambda)$, we compute two $N \times N$ matrices $\mathbf{C}^0(G^\lambda)$ and $\mathbf{C}^1(G^\lambda)$. In $\mathbf{C}^1(G^\lambda)$ entry $(i, j)$ stores the number of pairs of vertices $(u, v)$ with $u \in Z_i$ and $v \in Z_j$ between which a path exists in graph $G^\lambda$. Similarly, in $\mathbf{C}^0(G^\lambda)$ the $(i, j)$-th entry stores the number of pairs of vertices $(u, v)$ between which there is no path in graph $G^\lambda$. Then, based on our assumptions, the fraction of unsuccessful trips from zone $i$ to $j$ is given by $\mathbf{C}^0_{ij}(G^\lambda) / \left(\mathbf{C}^0_{ij}(G^\lambda) + \mathbf{C}^1_{ij}(G^\lambda)\right)$. Then, we let

$$I(G^\lambda) = \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{T_{ij} \mathbf{C}^0_{ij}(G^\lambda)}{\mathbf{C}^0_{ij}(G^\lambda) + \mathbf{C}^1_{ij}(G^\lambda)}$$

be the number of infeasible trips. In other words, the number of infeasible trips between an origin $i$ and destination $j$ is counted as the total number of trips from $i$ to $j$, multiplied the fraction of infeasible paths between the two zones.

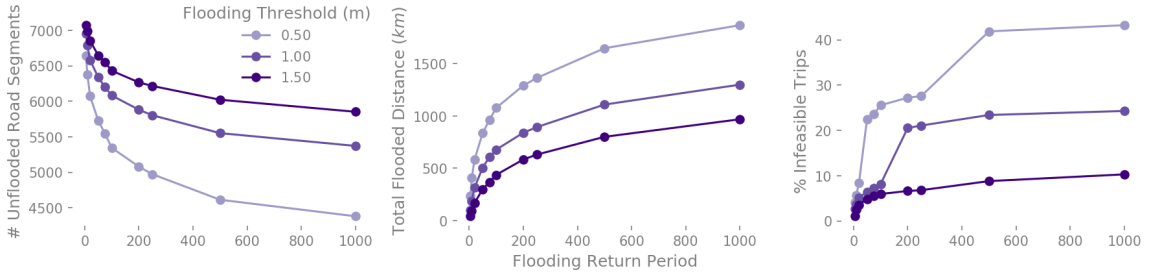



Figure 3.6: Graphs showing the effect of floods of increasing severity on the number of unflooded road segments (**left**), total length of flooded roads (**center**), and estimated percent of infeasible trips over the road network (**right**). Three flooding thresholds, 0.5, 1m, and 1.5m are shown.

## 3.3 Predicting Mobility for Vulnerability Analysis in Road Networks

In the absence of TAZ or CDR data, analytical or predictive models of human mobility can be used to approximate zone-to-zone trips. There has been a great deal of research on travel flow prediction, or estimating the number of trips taken between pairs of locations [60] given demographic, socioeconomic and/or geographic information about each location. Models developed for this purpose range from simple traditional ones with few parameters to complex ones capable of learning complex interactions from a large set of variables in order to more accurately capture the structure of mobility patterns [61, 62, 60]. The estimated origin-destination (OD) flows produced from such travel demand models can help decision-makers understand the use of road infrastructure and plan for its future.

Recent work by [63] evaluates several travel demand prediction models in terms of their capacity to capture census-tract-level urban travel demand patterns in the context of their suitability for informing road network resilience planning. Modeling these flows is often split into two parts: trip generation or production, which estimates the total number of trips leaving from or arriving at a given zone; and trip distribution, which characterizes what proportion of the trips generated for a given zone go to or come from each other zone. In other words, trip distribution models characterize the conditional probability $P(j|i)$ that a trip starting in zone $i$ ends in zone $j$, based on features of the origin and destination zones and various assumptions about what other factors impact human mobility [64]. An estimate of the number of trips from zone $i$ to zone $j$, $T_{ij}$, is given by

$$\tilde{T}_{ij} = T_i P(j|i), \tag{3.2}$$

where $T_i$ is the number of trips leaving zone $i$. $T_i$ is often estimated by a production function $\tilde{T}_i = \lambda m_i$ where $\lambda$ is a parameter that can be fitted. Historically, dozens models have been proposed for this task, but many of these fall into two major categories of approaches.

### 3.3.1 Gravity Models

Gravity models [65, 66, 67] assume the probability $P_{ij}$ that a trip begins in zone $i$ and ends in zone $j$ is proportional to the product of populations of the two zones, and inversely proportional to an exponential or power function of the distance $d_{ij}$ between the zones, where $d_{ij}$ can be the great-circle distance, Euclidean distance or travel distance between two zones.

$$P_{ij} \propto \frac{m_i m_j}{e^{\beta d_{ij}}} \tag{3.3}$$

$$P_{ij} \propto \frac{m_i m_j}{d_{ij}^{\beta}} \tag{3.4}$$

where $\beta$ is a parameter that can be adjusted, and $P_{ij}$ is normalized so that $\sum_i \sum_j P_{ij} = 1$. We adopt the common practice of ignoring trips within a zone, i.e., for any $i$, $P_{ii} = 0$.

### 3.3.2 Intervening Opportunities Models

The second main family of human mobility models encompasses different variants of intervening opportunities models [68]. The number of intervening opportunities between zone $i$ and zone $j$, $s_{ij}$, refers to the total number of jobs located closer to zone $i$ than zone $j$ is. We approximate this quantity with total number of jobs in all zones that are closer to zone $i$ than zone $j$:

$$s_{ij} = \sum_{k:d_{ik}<d_{ij}} s_k \tag{3.5}$$

where $s_k$ is the number of jobs in zone $k$. The number of jobs in a zone is sometimes estimated by the population of the zone, a convention we also adopt in this work.

Intervening opportunities trip distribution models approximate $P(j|i)$, the conditional probability that a trip leaving from zone $i$ will go to zone $j$. Note that $P(j|i) = \frac{P_{ij}}{\sum_j P_{ij}}$. Different intervening opportunities models include:

- **Schneider's intervening opportunities model** [69]:

$$P(j|i) = e^{-\gamma s_{ij}} - e^{-\gamma(s_{ij}+m_j)}, \tag{3.6}$$

- **Radiation model** [70]:

$$P(j|i) = \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{3.7}$$

- **Extended Radiation model** [71]:

$$P(j|i) = \frac{[(m_i + m_j + s_{ij})^\alpha - (m_i + s_{ij})^\alpha](m_i^\alpha + 1)}{[(m_i + s_{ij})^\alpha + 1][(m_i + m_j + s_{ij})^\alpha + 1]} \tag{3.8}$$

### 3.3.3 Learning-Based Travel Demand Models

One drawback of the traditional models is that the $\alpha$, $\beta$, $\gamma$ and $\lambda$ parameters that are fit to reflect travel demand patterns in one region do not generalize well to other regions due to the rigid functional forms of the models and reliance on a small set of features. Here, we present two models that incorporate a wider set of features and capture more varied functional relationships between those features and the travel demand. Specifically, given data consisting of a set of census tracts (zones), features $F$ for each zone, joint features $J$ between pairs of zones, and ground truth pairwise OD flows between zones over some time horizon, our goal is to learn a function $f(F_i, F_j, J_{ij}) = \tilde{T}_{ij}$ for predicting the OD flows for new areas in which we do not know the ground truth.

- **Extended Gravity Model**: This model generalizes the gravity model to incorporate a much wider set of features[72]. This extended gravity model adds power laws of additional features to the original gravity model, expressed as

$$\tilde{T}_{ij} = \beta \prod_l \phi_l^{\alpha_l}(i) \prod_m \phi_m^{\alpha_m}(j) \prod_n \phi_n^{\alpha_n}(i,j) f(d_{ij}) \tag{3.9}$$

where $\phi_l$ are features of the origin tract, $\phi_m$ are features of the destination tract, and $\phi_n$ are features related to both the origin and the destination (except the distance, which is included in the decay function $f(d)$). $f(d)$ can have either the power form or the exponential form. The bias $\beta$, and $\alpha$ coefficients are variables to be fit. A simpler form of the extended gravity model was proposed by Lowry [73, 74], which also considers employment status and income level as features which can be proportional or inversely proportional to travel demand. Earlier work by Alonso [75] also established a power-decay gravity model where the population values are replaced by power laws of population values.

- **Random Forest**: We also use a random forest model, which is one of the most accurate predictors used in a recent study [61] on predicting travel demand between cities based on socioeconomic features. Compared to the aforementioned models, random forest is a black-box model that is computationally more complex and more challenging to interpret.

### 3.3.4 Experiments

[63] evaluate the proposed travel demand prediction models in Washington D.C., Chicago and Seattle. They use geographic and socioeconomic features at the census-tract level to fit our travel demand models (Table 3.1). The census tract extents and indices are those used in the 2010 U.S. census, clipped to the TAZ-based study area for each city. Numerical features pertaining to population, land cover, employment, per capita income etc. for each census tract are obtained from the American Community Survey, the 2011 Environmental Summaries, and the 2015 Longitudinal Employer-Household Dynamics datasets on Social Explorer. For each pair of origin census tract $i$ and destination census tract $j$, we have 13 zone-based features $F_i$ relating to the origin, 13 zone-based features $F_j$ relating to the destination, and 5 joint features $J_{ij}$.

For each of the learning-based models, the OD flows from 2 of the 3 study areas are

Table 3.1: Feature set for the machine learning methods.

| Feature Category | Zone-based Feature | Origin-Destination Pair Feature |
|---|---|---|
| Population | Population | Intervening population |
| | Population density | |
| Geographic | Area | Euclidean distance |
| | Open space area | Shortest-path travel distance |
| | Low intensity development area | |
| | Medium intensity development area | |
| | High intensity development area | |
| | Forested area | |
| Work | Employed population | Intervening jobs |
| | Unemployed population | Intervening income |
| | Average commute time | |
| | Number of jobs | |
| | Per capita income | |

used as a training set and the trained models are tested on the OD flows of the third held-out study area. This process is repeated in 3 folds, corresponding to all 3 combinations of pairs of study areas. Only inter-zone trips are predicted: for any $i$, $\tilde{T}_{ii} = 0$. 4 evaluation metrics are considered to measure the prediction performance of the travel demand models in terms of the agreement between the off-diagonal entries of the ground truth and predicted origin-destination travel demand matrices (Table 3.2). We report the normalized root mean square error ($NRMSE$) and coefficient of determination ($r^2$) which are commonly used evaluation metrics for regression models. In addition, we use two variants of the common part of commuters metric ($CPC$) [76, 62] widely used in travel prediction:

$$CPC(T, \tilde{T}) = \frac{2\sum_{i,j=1}^{n} \min(T_{ij}, \tilde{T}_{ij})}{\sum_{i,j=1}^{n} T_{ij} + \sum_{i,j=1}^{n} \tilde{T}_{ij}} \tag{3.10}$$

The greater the agreement between the predicted travel flow volumes and the ground truth values, the closer the $CPC$ is to 1. [64] recently proposed the common part of commuters according to distance $CPC_d$, which measures how well a model predicts the distribution of travel distance, disregarding specific origins and destinations. If $N_k$ is the number of trips

Table 3.2: Average prediction performance. For the traditional trip distribution models, evaluation metrics are computed for a model fit to each of the 3 study areas and then averaged. For the learning-based models, the average is taken over the 3-fold tests. Higher is better for all metrics except for $NRMSE$.

| Category | Method | CPC | $\mathrm{CPC_d}$ | NRMSE | $\mathrm{r^2}$ |
|---|---|---|---|---|---|
| Traditional models | Gravity, exponential decay, euclidean distance | 0.583±0.032 | 0.860±0.063 | 6.099±2.854 | 0.338±0.118 |
| | Gravity, exponential decay, travel distance | 0.590±0.039 | 0.865±0.061 | 6.197±2.881 | 0.316±0.116 |
| | Gravity, power decay, euclidean distance | 0.552±0.070 | 0.790±0.081 | 6.993±3.711 | 0.155±0.257 |
| | Gravity, power decay, travel distance | 0.552±0.078 | 0.781±0.077 | 7.427±4.039 | 0.051±0.318 |
| | Schneider's model | 0.533±0.021 | 0.841±0.032 | 6.036±2.553 | 0.344±0.057 |
| | Radiation model | 0.297±0.046 | 0.430±0.036 | 17.660±7.619 | -4.612±0.545 |
| | Extended radiation model | 0.553±0.064 | 0.799±0.081 | 6.692±3.239 | 0.211±0.160 |
| Learning based models | Random Forest | 0.654±0.068 | **0.907±0.084** | **5.287±2.529** | **0.506±0.095** |
| | Extended gravity, exponential decay | **0.658±0.067** | 0.879±0.134 | 5.597±3.260 | 0.462±0.221 |
| | Extended gravity, power decay | 0.629±0.060 | 0.826±0.160 | 6.216±3.606 | 0.340±0.269 |

with distance between $2k-2$ and $2k$ km, and $\tilde{N}_k$ is the corresponding prediction:

$$CPC_d(T,\tilde{T}) = \frac{\sum_{k=1}^{\infty} \min(N_k, \tilde{N}_k)}{\sum_{i,j=1}^{n} T_{ij}} \tag{3.11}$$

which equals to 1 if, for every distance bin, the ground truth and the prediction have the same number of trips within the range; it equals to 0 if every trip from the ground truth data is within a different distance bin than all predicted trips.

The intervening opportunities models generally had poorer performance for predicting census-tract-level travel flows around a city. Among the gravity models, those using an exponentially decaying function of distance outperformed those using a power decay function. The learning-based travel flow prediction models outperformed all of the traditional models in all evaluation metrics. Random forest was the best performing model overall in terms of $CPC_d$, $NRMSE$ and $r^2$, while the extended gravity model with exponential decay was the best according to the $CPC$ metrics. Note that the extended gravity model generally is much less time- and space-demanding than random forest.

## 3.4 Optimizing Satisfiable Travel Demand on Road Networks

Now, we are in a position to address the problem of deciding which roads to fortify in order to ensure that the number of trips that can be completed in a given flood scenario

is maximized. We model this problem as an instance of *Budget*-Prize-Collecting Steiner Forest (Budget-PCSF): given a graph, pairs of vertices that need to be connected and travel demand between them, and edge repair costs, we select a subset of edges to repair in order to maximize the satisfied travel demand while respecting a budget constraint[1]. In the regular Prize-Collecting Steiner Forest (PCSF) problem [77], the goal is to minimize the combination of edge purchasing costs and incurred penalties for failing to connect designated pairs of vertices. However, in the disaster preparedness setting, it may be undesirable to combine the financial costs of upgrading roads and the socioeconomic costs of failing to connect certain location pairs into a single optimization objective. Moreover, government agencies and development initiatives typically must operate within strict budget plans, and hence it is necessary to include a hard budget constraint.

### 3.4.1 Budget Prize-Collecting Steiner Forest

We are given an undirected, uncapacitated graph $G = (V, E)$, where edges represent road segments, and vertices represent junctions or endpoints of the road segments. We also have an OD matrix whose entries $(i, j)$ contain the expected number of trips from vertex $i$ to vertex $j$ over the road network, which we refer to as the travel demand from $i$ to $j$. Travel demands need not be symmetric, and we assume that as long as a path exists in the network between vertices $i$ and $j$, the travel demand between them in both directions is satisfied. This constitutes a profit function $p(u, v) : V \times V \to \mathbb{R}_+$ for connecting pairs of vertices. We emphasize that in our formulation of Budget-PCSF, we use these demands to set pairwise profits on vertices to be maximized, rather than setting penalties to be minimized as is typical in the PCSF problem. We are also given a cost function $c : E \to \mathbb{R}_+$ on the edges, and a fixed budget $B$. These edge costs reflect the projected cost of satisfactorily upgrading a given road segment to ensure it withstands a given flood scenario. The planner's task is to decide which road segments to upgrade through the allocation of the budget $B$, such that

[1] https://github.com/amritagupta/budget-pcsf-semigradient-ascent.

the maximum travel demand is satisfied. This leads to the following problem:

**Given:** Graph $G = (V, E)$, a non-negative edge cost function $c : E \to \mathbb{R}_+$, a budget $B$, pairs of vertices $\mathcal{P} = \{(u_1, v_1), (u_2, v_2), \ldots, (u_k, v_k)\}$ to be connected, and a non-negative profit function $p : \mathcal{P} \to \mathbb{R}_+$ for successfully connecting vertex pairs in $\mathcal{P}$

**Find:** A forest $F \subseteq E$ such that $\sum_{e \in F} c(e) \leq B$ and $\sum_{(u,v) \in \mathcal{Q}} p(u, v)$ is maximized, where $\mathcal{Q} \subseteq \mathcal{P}$ is the set of vertex pairs connected by edges in $F$.

We specify that the selected road segments should form an acyclic subgraph (a forest). Although a cyclic subgraph would be a feasible solution with respect to our goal of maximizing satisfied travel demand, we note that the presence of cycles in the solution means there are multiple edge-disjoint paths between the same pairs of vertices, providing no benefit in terms of additional connectivity but consuming more of the budget than needed. Hence, we can restrict our search to solutions that form a forest in $G$.

### 3.4.2 Related Work

Closely related to our graph optimization problem, the prize-collecting Steiner forest (PCSF) problem asks the following question: given an undirected graph $G = (V, E)$, a non-negative edge cost function $c : E \to \mathbb{R}_+$, pairs of vertices $\mathcal{P} = \{(u_1, v_1), (u_2, v_2), \ldots, (u_k, v_k)\}$, and a non-negative *penalty* function $\pi : \mathcal{P} \to \mathbb{R}_+$, which subgraph $F$ of $G$ minimizes the cost of edges in $F$ plus the sum of penalties for pairs in $\mathcal{P}$ that are not connected by $F$? This problem has been studied extensively and applied in domains as diverse as molecular biology for discovering signaling pathways in a cellular interactome [78], and backbone discovery in transportation networks [79]. The PCSF problem was shown to be APX-hard [80], and has a 3-approximation algorithm based on the primal-dual method [77]. [81] devised an $O(|V|^{2/3} \cdot 2 \cdot log|V|)$-approximation algorithm for the *Quota*-PCSF with uniform profits, where the goal is to find a minimum cost forest such that the satisfied demand

$$S = \{\emptyset\}$$
$$T = \{e_{AB}\}$$
$$U = \{e_{AB}, e_{BC}\}$$

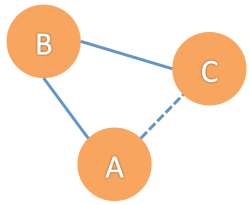



Figure 3.7: Restricted supermodularity. $f(S \cup \{e_{BC}\}) - f(S) \leq f(T \cup \{e_{BC}\}) - f(T)$, showing compounding gains. However, $f(T \cup \{e_{AC}\}) - f(T) \not\leq f(U \cup \{e_{AC}\}) - f(U) = \mathbf{0}$, showing restricted supermodularity.

is at least $Q$. However, despite the practical relevance of the budget-constrained variant of PCSF, it has received relatively little attention, and there is no known approximation algorithm for Budget-PCSF known to the authors.

Existing approaches for solving Budget-PCSF largely rely on computationally heavy approaches. [82, 83] propose mixed integer programming formulations for this problem and demonstrate results on a graph with 23 vertices and 34 edges representing the Ohio interstate system. [84] propose a bi-objective integer programming model to solve a closely related problem to Budget-PCSF, in which the goal is to maximize the satisfied demand (where demand at an origin is satisfied if a path exists to *at least one* of a set of predefined destinations) while minimizing travel time between origin-destination pairs as a second objective with application to retrofitting bridges along critical routes for earthquake response. In contrast, we consider the setting in which the demand is specified between each pair of vertices. In general, flow-based integer programming formulations are known to encounter challenges when scaling to larger instances of these families of network design problems and can be particularly sensitive to the number of OD-pair flow variables. Instead, in this work we focus on developing scalable heuristic approaches. For example, in [39] the authors employed a simple greedy algorithm to solve Budget-PCSF.

### 3.4.3 Restricted Supermodularity

In recent work, results about the modularity of objective functions have played an instrumental role in designing effective algorithmic approaches to solving network design prob-

lems in domains such as robotic motion control [85] and sensor placement [86]. We now turn our attention to an analysis of the objective function in Budget-PCSF, and specifically prove that it is *restricted supermodular* over subsets of edges that form forests. A set function $f(S) : 2^E \rightarrow \mathbb{R}$ is supermodular if it exhibits the property of compounding gains, or formally if $\forall S \subset T \subset E$ and $\forall e \in E \setminus T$:

$$f(S \cup \{e\}) - f(S) \leq f(T \cup \{e\}) - f(T) \tag{3.12}$$

However, when the above property holds only over a collection of subsets of $E$, $f$ is referred to as a *restricted* supermodular function. The analogous property of restricted submodularity for functions with diminishing returns was first described by [87] who used it to analyze a greedy algorithm for the Steiner tree problem.

Let $S \subseteq 2^E$ denote a set of edges selected from graph $G$. We aim to maximize the following objective function subject to a knapsack (budget) constraint:

$$f(S) = \sum_{(u,v) \in \mathcal{Q}} p(u, v) \tag{3.13}$$

where $\mathcal{Q} \subseteq \mathcal{P}$ is the set of vertex pairs connected by edges in $G(S)$, the graph induced by edges in the set $S$, and $p(u, v)$ is the profit function described earlier.

**Proposition 1:** *$f(S)$ is monotone non-decreasing.*

This is trivial since augmenting the set $S$ with another element (edge) can never reduce the number of pairs of vertices connected in $G$; and the profit function $p$ is non-negative.

**Proposition 2:** *$f(S)$ is supermodular when restricted to the set of forests $F$ on $G$ (the graph matroid).*

A first intuition regarding the relationship between $f$ and the structure of subsets on which it is defined, is that there must exist a set $S$ that maximizes $f$ and contains no cycles– i.e. $G(S)$ is a forest. As described earlier, any solution containing a cycle would contain

multiple edge-disjoint paths between the same pair of vertices, providing no additional connectivity but consuming more of the budget than strictly necessary. Therefore, we can restrict our search to forests.

A second intuitive notion is that connectivity builds upon itself. See, for instance, the graph in Figure 3.7. If edge A–B is restored, then demand between A and B can be satisfied. The same holds for edge B–C. However, when *both* A–B and B–C are restored, in addition to restoring connectivity between A and B and between B and C, we get the *extra* benefit of connecting A and C. This phenomenon of *compounding gains* or *increasing differences* is the characteristic feature of supermodular functions, and has been observed and leveraged in various other settings, such as influence maximization under the linear threshold model with edge addition [5]. Figure 3.7 also illustrates the concept of restricted supermodularity. When attempting to augment set $U$ by adding edge $e_{AC}$ that forms a cycle with respect to the other edges already in $U$, we have $f(U \cup \{e_{AC}\} - f(U) = 0$, although adding $e_{AC}$ to a subset of $U$ could have strictly increased $f$; in this case the gain associated with $e_{AC}$ was diminished, violating the requirement for general supermodularity. A formal proof that $f$ is restricted supermodular function follows.

*Proof.* For notational convenience, let Let $G(u; S)$ be the connected component in $G(S)$, the subgraph induced by edges in the set $S$, that contains vertex $u$. Let $f_G(u; S)$ be the total profit between pairs of vertices in the connected component $G(u; S)$. Consider a subset of edges $S \subset E$ and two different edges $e_1 = (u_{e_1}, v_{e_1})$ and $e_2 = (u_{e_2}, v_{e_2})$ in $E \backslash S$. Let $\Delta_e f(S) := f(S \cup \{e\}) - f(S)$. We want to show that $\Delta_{e_2} f(S \cup \{e_1\}) \geq \Delta_{e_2} f(S)$ whenever $G(S \cup \{e_1, e_2\})$ contains no cycles. (We already provided a counterexample in Figure 3.7 showing that this is not true when $S$ contains a cycle.) There are 3 possible cases:

**1. $\mathbf{e_2}$ has no endpoints in $\mathbf{G(S \cup \{e_1\})}$.** Then $e_2$ also has no endpoints in $G(S)$, so

$$\begin{aligned}\Delta_{e_2} f(S) &= f(S \cup \{e_2\}) - f(S) \\ &= f_G(u_{e_2}; S \cup \{e_2\}) - f_G(u_{e_2}; S) \\ &= f_G(u_{e_2}; S \cup \{e_2\}) - 0 = f_G(u_{e_2}; \{e_2\})\end{aligned}$$

$$\begin{aligned}\Delta_{e_2} f(S \cup \{e_1\}) &= f(S \cup \{e_1, e_2\}) - f(S \cup \{e_1\}) \\ &= f_G(u_{e_2}; S \cup \{e_1, e_2\}) - f_G(u_{e_2}; S \cup \{e_1\}) \\ &= f_G(u_{e_2}; S \cup \{e_1, e_2\}) - 0 \\ &= f_G(u_{e_2}; \{e_2\}) = \Delta_{e_2} f(S)\end{aligned}$$

**2. $\mathbf{e_2}$ has exactly 1 endpoint in $\mathbf{G(S \cup \{e_1\})}$.** Let $u_{e_2}$ be the endpoint of $e_2$ in $G(S \cup \{e_1\})$. Either $u_{e_2}$ is in the same connected component as the endpoints of $e_1$, or it is in a different connected component of $G(S \cup \{e_1\})$. If $u_{e_2} \in G(u_{e_1}; S \cup \{e_1\})$, there are two possibilities–either $u_{e_2}$ is a vertex in $G(S)$ or $u_{e_2} = v_{e_1}$. In the former case:

$$\begin{aligned}\Delta_{e_2} f(S) &= f(S \cup \{e_2\}) - f(S) \\ &= f(u_{e_2}; S \cup \{e_2\}) - f(u_{e_2}; S) \\ &= \{\text{flow } G(S) \leftrightarrow v_{e_2}\}\end{aligned}$$

$$\begin{aligned}\Delta_{e_2} f(S \cup \{e_1\}) &= f(S \cup \{e_1, e_2\}) - f(S \cup \{e_1\}) \\ &= f_G(u_{e_2}; S \cup \{e_1, e_2\}) - f_G(u_{e_2}; S \cup \{e_1\}) \\ &= \{\text{flow } G(S) \leftrightarrow v_{e_2}\} + \{\text{flow } v_{e_1} \leftrightarrow v_{e_2}\} \\ &\geq \{\text{flow } G(S) \leftrightarrow v_{e_2}\} = \Delta_{e_2} f(S)\end{aligned}$$

In the latter case:

$$\Delta_{e_2} f(S) = f_G(u_{e_2}; S \cup \{e_2\}) - f_G(u_{e_2}; S)$$
$$= \{\text{flow } u_{e_2} \leftrightarrow v_{e_2}\}$$

$$\Delta_{e_2} f(S \cup \{e_1\}) = f_G(u_{e_2}; S \cup \{e_1, e_2\}) - f_G(u_{e_2}; S \cup \{e_1\})$$
$$= \{\text{flow } G(u_{e_1}; S \cup \{e_1\}) \leftrightarrow v_{e_2}\}$$
$$\geq \{\text{flow } u_{e_2} \leftrightarrow v_{e_2}\} = \Delta_{e_2} f(S)$$

If, on the other hand, $u_{e_2} \notin G(u_{e_1}; S \cup \{e_1\})$ then:

$$\Delta_{e_2} f(S) = f(S \cup \{e_2\}) - f(S)$$
$$= f_G(u_{e_2}; S \cup \{e_2\}) - f_G(u_{e_2}; S)$$

$$\Delta_{e_2} f(S \cup \{e_1\}) = f(S \cup \{e_1, e_2\}) - f(S \cup \{e_1\})$$
$$= f_G(u_{e_2}; S \cup \{e_1, e_2\}) - f_G(u_{e_2}; S \cup \{e_1\})$$
$$= f_G(u_{e_2}; S \cup \{e_2\}) - f_G(u_{e_2}; S) = \Delta_{e_2} f(S)$$

**3.$\mathbf{e_2}$ has both endpoints in $\mathbf{G(S \cup \{e_1\})}$.** Since we restrict ourselves to subsets of $E$ that do not form cycles, the endpoints of $e_2$ must be in two separate connected components both in $G(S)$ and $G(S \cup \{e_1\})$. Either $e_2$ links two connected components that do not contain the endpoints of $e_1$—in this case $\Delta_{e_2} f(S \cup \{e_1\}) = \Delta_{e_2} f(S)$. Otherwise, $e_2$ links one connected component containing the endpoints of $e_1$ with another connected component– in this case $\Delta_{e_2} f(S \cup \{e_1\}) \geq \Delta_{e_2} f(S)$. □

### 3.4.4 Budget-Constrained Restricted Supermodular Maximization

With the results of the previous secion, we can approach Budget-PCSF as maximizing a monotone non-decreasing, restricted supermodular function over the independent sets of a graph matroid, subject to a knapsack constraint. There is very limited work to date on maximizing restricted submodular or supermodular functions subject to additional constraints. In [87], the authors studied the implications of restricted submodularity over forests for a greedy algorithm for the minimum Steiner tree problem, in which there are no cardinality or knapsack constraints for acquiring edges. In [88], the authors studied maximizing a function that is submodular when restricted to the set of solutions satisfying a cardinality constraint. In contrast to these works, in Budget-PCSF the supermodularity of the objective function holds over the independent sets of a matroid, while there is an additional, separate knapsack constraint on edge selection.

In this work, we treat both the graph matroid structural restrictions and the knapsack constraint as hard constraints. Even though a solution violating the matroid restriction is not strictly infeasible, the violations degrade the supermodularity property that serves as the basis of our algorithm design. Therefore, we aim to maximize a supermodular function subject to a knapsack and a matroid constraint. [89] provide bounds for maximizing a supermodular function subject to a *cardinality* constraint (equivalent to knapsack constraints with unit or uniform edge costs) **or** a matroid constraint, but not both. They showed that a simple greedy algorithm that adds items in order of maximum benefit with respect to the current set achieves a $(1 - \kappa^f)$ approximation bound, where $\kappa^f$ is the supermodular curvature of function $f$. It is unclear whether the same bound holds for maximizing supermodular functions subject to a knapsack constraint, rather than a cardinality constraint; or for the combination of a knapsack and a matroid constraint, as we have here. [39] employed a similar greedy algorithm to [89] to solve Budget-PCSF, in which items are selected in order of benefit-cost ratio. However, as we will show, this greedy algorithm scales poorly with problem size.

Instead of relying on a simple cost-benefit greedy criterion, we implement an iterative heuristic based on *semigradient ascent* [90], which is based on computing modular lower bounds (MLB) for the objective function $f$ and then efficiently maximizing the MLB. This theoretical framework was recently adapted by [5] to supermodular maximization with a cardinality constraint for influence maximization in graphs using edge addition.

Supermodular functions have discrete subdifferentials that can be used to construct tight MLBs [89]. Given a set function $f$ and a set $S$, the subdifferentials of $f$ at $S$ are all vectors $y$ such that $f(S) - y(S) + y(S') \leq f(S')\ \forall S' \in E$; a subgradient is one such vector $y$. Therefore, a subgradient essentially provides a lower bound for the function $f$ evaluated on set $S'$. In [89], the following two *discrete* subgradients for supermodular functions are proposed:

$$\check{y}(j) = \begin{cases} f(j|S \setminus \{j\}), & \text{if } j \in S \\ f(j|\emptyset), & \text{otherwise} \end{cases} \tag{3.14}$$

and,

$$\hat{y}(j) = \begin{cases} f(j|E \setminus \{j\}), & \text{if } j \in S \\ f(j|S), & \text{otherwise} \end{cases} \tag{3.15}$$

These lead to the following two modular lower bounds for the function $f$ at a new solution $S'$ using current solution $S$:

$$\check{m}_S(S') = f(S) - \sum_{j \in S \setminus S'} f(j|S \setminus j) + \sum_{j \in S' \setminus S} f(j|\emptyset) \leq f(S') \tag{3.16}$$

$$\hat{m}_S(S') = f(S) - \sum_{j \in S \setminus S'} f(j|E \setminus j) + \sum_{j \in S' \setminus S} f(j|S) \leq f(S') \tag{3.17}$$

The terms $f(j|S \setminus j)$ and $f(j|S)$ depend on the current solution $S$ and thus need to be computed on the fly through calls to a function evaluator for $f$. However, $f(j|\emptyset)$ in

Eq. Equation 3.16 is simply the travel demand between the endpoints of edge $j$, so this term can be looked up from the demand matrix.

For general supermodular functions, the $f(j|E \setminus j)$ terms in Eq. Equation 3.17 can be precomputed for each $j \in E$. These terms quantify the largest reduction in objective value that could occur as a result of removing element $j$ from the current solution. However, for our restricted supermodular function, we need to find $\max_{S \in \mathcal{C}} f(j|S \setminus j)$, where $\mathcal{C}$ is the collection of sets over which the supermodularity property holds, or the graph matroid in our case. The largest objective value reduction for removing a given edge $j$ occurs when $S$ is a specific spanning tree on $G$ such that $j$ is the cut-set for a weighted max-cut on the spanning tree. We adopt a sampling approach to evaluate the maximum possible impact of each edge $j$, sampling a set $\mathcal{T}_j$ of $N$ random spanning trees of $G$ in which $j$ is a member and setting $f(j|E \setminus j) \approx \max_{\mathcal{T} \in \mathcal{T}_j} \{f(j|\mathcal{T} \setminus j)\}$.

Another simple modular lower bound that was successfully used for constrained submodular maximization [5] is the following:

$$\sum_{j \in S'} f(j|\emptyset) \leq f(S') \tag{3.18}$$

where $f(j|\emptyset)$ is again the demand between the endpoints of edge $j$. Eq. Equation 3.18 does not require a current solution or any expensive function evaluations to compute, and is in fact equivalent to Eq. Equation 3.16 or Eq. Equation 3.17 when the initial solution $S$ is the empty set.

*Semigradient Ascent*

The modular lower bounds in Eqs. Equation 3.16 and Equation 3.17 are computed with respect to a solution $S$, and can be used to iteratively find solutions of increasing objective value using Algorithm Algorithm 1 proposed in [90] and adapted to our restricted supermodularity setting. A current solution to Budget-PCSF ($E_{add}$) is used to compute the

coefficients for one of the subgradient-based modular lower bounds (lines 9 and 12). Then, we find a new solution that maximizes this modular lower bound subject to budget and matroid constraints (lines 10 and 13). We alternate between the two bounds, terminating when no further changes to the solution are made under either of them.

**Algorithm 1** SEMIGRADIENT-ASCENT

1: **function** SEMIGRAD($E, c, B, p, \epsilon$)
2: `current_mlb` $\leftarrow \check{m}$
3: $\check{m}$`_converged`$\leftarrow$`False`
4: $\hat{m}$`_converged`$\leftarrow$`False`
5: `converged`$\leftarrow$`False`
6: $E_{\text{add}} \leftarrow \emptyset$
7: **while not** `converged` **do**
8: **if** `current_mlb` $= \check{m}$ **then**
9: $\check{m}_{E_{\text{add}}}(\cdot) \leftarrow$GETMLBCOEFFS($E_{\text{add}}$)
10: $E'_{\text{add}} \leftarrow \arg\max_{S \in \mathcal{C}} \check{m}_{E_{\text{add}}}(S)$
11: **else**
12: $\hat{m}_{E_{\text{add}}}(\cdot) \leftarrow$GETMLBCOEFFS($E_{\text{add}}$)
13: $E'_{\text{add}} \leftarrow \arg\max_{S \in \mathcal{C}} \hat{m}_{E_{\text{add}}}(S)$
14: **if** $E'_{\text{add}} = E_{\text{add}}$ **then**
15: **if** `current_mlb`$=\check{m}$ **then**
16: $\check{m}$`_converged`$\leftarrow$`True`
17: `current_mlb` $\leftarrow \hat{m}$
18: $E_{\text{add}} \leftarrow E'_{\text{add}}$
19: **else**
20: $\hat{m}$`_converged`$\leftarrow$`True`
21: `current_mlb` $\leftarrow \check{m}$
22: $E_{\text{add}} \leftarrow E'_{\text{add}}$
23: **if** $\check{m}$`_converged` **and** $\hat{m}$`_converged` **then**
24: `converged`$\leftarrow$`True`
25: **else**
26: $E_{\text{add}} \leftarrow E'_{\text{add}}$
27: $\check{m}$`_not_converged`$\leftarrow$`True`
28: $\hat{m}$`_not_converged`$\leftarrow$`True`
29: **return** $E_{\text{add}}$

*Maximizing Modular Lower Bounds*

Maximizing any of the above modular lower bounds corresponds to maximizing a modular function subject to *both a knapsack and a matroid constraint*. There is relatively little work on solving the 0-1 knapsack problem (whose objective is the modular sum of the item values), subject to additional constraints beyond the budget constraint. However, maximizing *submodular* functions subject to combinations of different numbers and types of constraints is an active area of research. We leverage these approaches based on the fact that modular functions are in fact submodular (as well as supermodular), with the inequality satisfied with equality.

Maximizing a monotone submodular function subject to a knapsack *and* a matroid constraint is NP-hard to approximate to within a factor better than $1-\frac{1}{e}$ [91], and only a handful of relatively recent works have proposed algorithmic techniques for addressing this problem. One of the first approaches for submodular maximization suitable for a combination of matroid and knapsack constraints was the randomized swap rounding algorithm proposed by [92]. More recently, [91] presented a simpler $\frac{1-e^{-2}}{2}$-approximation algorithm that also handles the simultaneous application of a matroid constraint and $k$ knapsack constraints. The algorithm proceeds by greedily building the solution set by considering local search moves (either adding a single element or by swapping a previously added element for a new one) that respect the matroid constraints. The best move is chosen greedily according to benefit-cost ratio. Despite the attractive guarantees regarding the quality of this solution set, the algorithm is not guaranteed to terminate in polynomial time.

We adopt simpler heuristic strategies for maximizing the modular lower bounds subject to the knapsack and matroid constraints. Our first approach, `GreedyMLB` is similar to [91] in that edges are chosen in order of benefit-cost ratio, but swaps are not allowed. Each edge is added to the solution only if it can be purchased with the remaining budget and if it does not introduce a cycle into the solution. We also propose a second algorithm, `Knapsack-Repair`, shown in Algorithm Algorithm 2. This algorithm alternates

**Algorithm 2** KNAPSACK-REPAIR

**function** KR($E, c, B, p, \epsilon$)
    `add_edges`←`True`
    $E_{\text{add}} \leftarrow \emptyset$
    **while** `add_edges` **do**
        $E_{\text{K}} \leftarrow$KNAPSACKILP($E, c, B, p, \epsilon$)
        $E_{\text{add}} \leftarrow E_{\text{add}} \cup E_{\text{K}}$
        $E_{\text{prune}} \leftarrow$REPAIR($E_{\text{add}}$, $c$)
        $E_{\text{add}} \leftarrow E_{\text{add}} \backslash E_{\text{prune}}$
        $B \leftarrow B - \sum_{e \in E_{\text{add}}} p(e_u, e_v)$
        $E \leftarrow E \backslash (E_{\text{add}} \cup E_{\text{prune}})$
        **if** $B < \min \{c(e)\}_{e \in E}$ OR $E = \emptyset$ **then**
            `add_edges`←`False`
    **return** $E_{\text{add}}$
**function** REPAIR($E_{\text{add}}, c$)
    `cycles_fixed`←`False`
    $E_{\text{prune}} \leftarrow \emptyset$
    $E_{\text{cycles}} \leftarrow$ FINDEDGESINCYCES($E_{\text{add}}$)
    **if** $E_{\text{cycles}} = \emptyset$ **then**
        `cycles_fixed`←`True`
    **while not** `cycles_fixed` **do**
        $e_{\text{expensive}} \leftarrow \arg\max_{e \in E_{\text{cycles}}} c(e)$
        $E_{\text{prune}} \leftarrow E_{\text{prune}} \cup \{e_{\text{expensive}}\}$
        $E_{\text{cycles}} \leftarrow$FINDEDGESINCYCES($E_{\text{add}}$)
        **if** $E_{\text{cycles}} = \emptyset$ **then**
            `cycles_fixed`←`True`
    **return** $E_{\text{prune}}$

between phases of edge addition, in which an integer linear program for the 0-1 knapsack problem is solved to allocate the available budget towards edges, and repair, in which matroid constraint violations are corrected by greedily removing the most expensive edge participating in a cycle in the solution, recovering the cost of the edge, and continuing until all cycles are repaired. The algorithm alternates between the knapsack and repair phases until either the budget is exhausted or all edges have been either added or discarded.

In both `GreedyMLB` and `Knapsack-Repair`, each edge is added to the solution at most once (rather than allowing edges to be swapped out and then potentially swapped back in), and so the algorithms terminate quickly and are good candidates for solving the subproblems in lines 10 and 13 of Algorithm Algorithm 1.

### 3.4.5 Experiments and Results

We compare the performance of four algorithms: 1) a baseline `Greedy` algorithm that iteratively constructs a solution by adding the edge with the best benefit to cost ratio without violating the knapsack and matroid constraints ([39], implemented with lazy evaluation); 2) `KR` maximizing $f$ by applying `Knapsack-Repair` (Algorithm Algorithm 2) to the modular lower bound in Eq. Equation 3.18; 3) `Semigrad-GreedyMLB` (Algorithm Algorithm 1) where at each iteration the modular lower bound is optimized greedily; and 4) `Semigrad-KR` (Algorithm Algorithm 1) where at each iteration the modular lower bound is optimized by calling `Knapsack-Repair`. We report results on a real road network from Senegal, a country where flood resilience is of particular concern, and on synthetic instances to more fully characterize the algorithms' performance. For each problem instance, we compute the cost of the minimum spanning tree (MST), which is the minimum cost at which all the available profit can be obtained. We then vary the budget allocated for Budget-PCSF as a fraction of the MST cost. The knapsack ILPs within `KR` were solved using Gurobi v8.0. All experiments were run on a cluster of five 32-core machines with 2.10GHz processors and 256GB of RAM.

*Synthetic Instances*

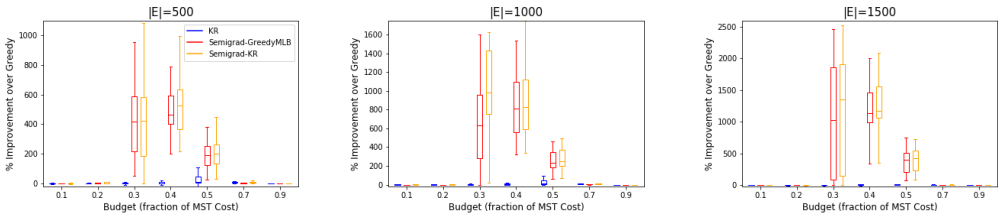



Figure 3.8: Solution quality relative to `Greedy` versus budget level on random planar graphs of different sizes; `KR` is on average as good as `Greedy`, while `Semigrad-GreedyMLB` and `Semigrad-KR` find solutions 3-15 times better than `Greedy`.

We synthetically generate instances of Budget-PCSF on which to evaluate our proposed methods. Specifically, we generate random planar graphs with mean degree close to 3.8, the mean degree of the Senegal road network retrofitting instance under the 50-year flood scenario. We vary the size of the instances parameterized by number of edges, generating 20 random instances of each graph size. We also created a similar dataset with random Erdos-Renyi graphs, for which results can be found on our GitHub page. We generated costs for edges and demand between pairs of vertices following a random uniform distribution.

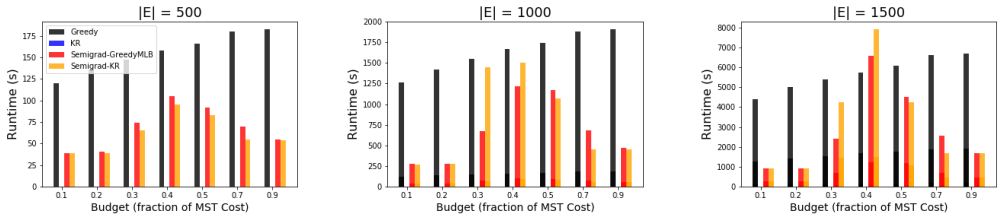



Figure 3.9: Mean runtime versus budget level for `Greedy`, `KR`, `Semigrad-GreedyMLB` and `Semigrad-KR` on random planar graphs of 3 sizes; `Knapsack-Repair` (KR) is extremely fast across all instance sizes and budget levels.

Figure 3.8 shows the relative quality of solutions found by `KR`, `Semigrad-GreedyMLB` and `Semigrad-KR` compared to the `Greedy` method at different budget levels for instances varying from 500 to 1500 edges. Solution quality is reported as % improvement in objective value compared to the `Greedy` baseline. The results show that for budget frac-

tions 0.1-0.2 and 0.7-0.9, all 3 algorithms achieve results on par with the `Greedy` baseline. *At intermediate budget levels, the two iterative semigradient ascent-based methods deliver dramatic improvements over the `Greedy` solutions.*

Figure 3.9 shows the mean runtime for each algorithm for Budget-PCSF at different budget levels and instance sizes. `KR` is the fastest method by far, with runtimes under 1 second in nearly all cases. This means that *`KR` can provide solutions with quality on par with `Greedy` with a speed-up of 100 to 5000 times*. The runtimes for the two semigradient ascent-based methods also include the time taken to estimate the maximum value of each edge in the graph instance, as described in subsubsection 3.4.4. Semigradient-based methods are typically faster than `Greedy` across different budgets; hence, at intermediate budget levels, semigradient-based methods beat `Greedy` in both solution quality and time.

*Road Upgrades for Flood Resilience in Senegal*

Road infrastructure plays a key role in socioeconomic development, and consequently infrastructure expansion initiatives are a core focus of many countries' economic plans. However, extreme weather events cause damage to essential infrastructure, costing billions of dollars to repair and potentially setting back economic development and exacerbating the existing vulnerabilities faced by the population. Flooding is of major concern in Senegal, due to both a recent increase in severe floods and the susceptibility of the largely unpaved road network to damage from precipitation. Preserving connectivity via the national and regional road network in the event of frequently-occurring (e.g. 5- or 10-year) floods is an important goal towards enabling recovery efforts as well as normal activity to continue in the face of these risks.

We apply the Budget-PCSF problem to retrofitting the Senegal regional road network against 5, 10, 20 and 50-year flooding scenarios. The full road network consists of 6917 vertices corresponding to road intersections or endpoints and 7175 edges corresponding to

Table 3.3: Budget-PCSF instance sizes resulting from floods of different return periods on the Senegal road network.

| Return Period | Vertices | Edges | MST Cost (km) |
|---|---|---|---|
| **5** | 161 | 178 | 60.399 |
| **10** | 285 | 321 | 122.593 |
| **20** | 444 | 518 | 211.250 |
| **50** | 654 | 758 | 353.884 |

road segments forming a single connected component. The flood risk of each road can be estimated from predicted or historical flood data. The costs of the edges are assumed to be proportional to the length of the flooded portion of the road segment, which would need to be fortified. We construct a compact representation of the flooded road network, where each vertex represents a connected component and each edge is the least cost edge between a pair of connected components. The Budget-PCSF graph sizes resulting from the 4 flood scenarios are summarized in Table 3.3. Travel demand over the network can be estimated from population data, or from fine-grained mobility data mined from call detail records, for example [39]. Unlike our synthetic instances with random uniform demands, the pairwise travel demands between vertices in the Senegal road network graphs were highly skewed, with very low travel demand between all but a few vertex pairs.

We compare the performance of the `Greedy` algorithm, `KR`, `Semigrad-GreedyMLB` and `Semigrad-KR` with a budget of $B = 0.1 \times$ MST cost in Table 3.4. The four methods perform more similarly compared to our experiments on synthetic instances; this is due to the fact that the highly skewed pairwise demands enable the `Greedy` baseline to perform reasonably well. Nevertheless, `KR` still produces solutions as good as or better than `Greedy` in a fraction of the time, providing a $140\times$ speedup on the smallest instance (5-year flood scenario) and a $> 8000\times$ speedup on the largest instance (50-year flood scenario). The semigradient-ascent based methods improve solution quality beyond either `Greedy` or `KR` while still being 5 to 20 times faster than `Greedy`. Although the higher objective values attained by `Semigrad-GreedyMLB` and `Semigrad-KR` come at the cost of more iterations and longer runtime compared to `KR`, these algorithms terminate within a

few seconds on the real-world road instances, e.g. in under 3 minutes on even the largest instance corresponding to the 50-year flood scenario.

## 3.5 Predict then Optimize: Combining Predicted Travel Flows with Optimizing Budget-PCSF

While there has been considerable prior work on predicting ground truth travel flows, there has been little to no assessment of how prediction errors might propagate and impact downstream decision making. Motivated by this important, real-world use case of travel flow prediction models, *we propose an evaluation pipeline for assessing a travel flow model's ability to guide infrastructure investments for disaster mitigation planning in urban areas*. We use the predicted travel flows as inputs to our semigradient ascent-based procedure for road network fortification planning, and assess the quality of the obtained plans with respect to the ground truth mobility flows. By examining whether the resulting recommended fortification plans are sensitive to errors in the predictions of the proposed travel demand models, we evaluate the travel demand prediction models on the basis of their suitability for use in such a decision pipeline.

We employ the semigradient ascent algorithm to select which road segments to upgrade [93]. Then, we assess the quality of each plan by using the ground truth OD flows to evaluate the number of remaining infeasible trips with the recommended upgrades in place. Figure 3.10 shows the result of using OD flows predicted by the random-forest model, the gravity model with exponential decay and travel distances, and ground truth data to inform the decision process. At the low budget levels shown, only a few critical road segments can be chosen for flood resilience upgrades and so segments that restore the most mobility must be correctly identified. In terms of the realized reduction in infeasible trips, both the random forest and the simple gravity model perform comparably to the upgrade plans designed with the ground-truth OD data itself. The predictive error of the models impacted the downstream decision task by over- or underestimating the quality of the road upgrade

Table 3.4: Best found solution objective value (number of feasible trips) and runtime (s) on the Senegal road network for 5-, 10-, 20- and 50-year floods with budget 0.1*MSTCost. Values in parentheses are gap percentages for sub-optimal solutions.

| **Method** | 5-year | | 10-year | | 20-year | | 50-year | |
|---|---|---|---|---|---|---|---|---|
| | Objective | Time (s) | Objective | Time (s) | Objective | Time (s) | Objective | Time (s) |
| Greedy | 1195531703 (0.02%) | 14.03 | 1182448994 (0.75%) | 84.00 | 1177026417 (0.08%) | 332.30 | 1161809642 (0.83%) | 1140.11 |
| KR | 1195531703 (0.02%) | 0.10 | 1182534472 (0.74%) | 0.05 | 1177389779 (0.05%) | 0.07 | 1161974353 (0.82%) | 0.13 |
| Semigrad-GreedyMLB | **1195800142** (0.00%) | 2.97 | **1191291479** (0.00%) | 7.06 | **1177993964** (0.00%) | 32.17 | 1171044232 (0.04%) | 85.84 |
| Semigrad-KR | 1195798624 (0.00%) | 2.06 | 1183590105 (0.65%) | 5.04 | 1177948435 (0.00%) | 16.63 | **1171484174** (0.00%) | 155.19 |

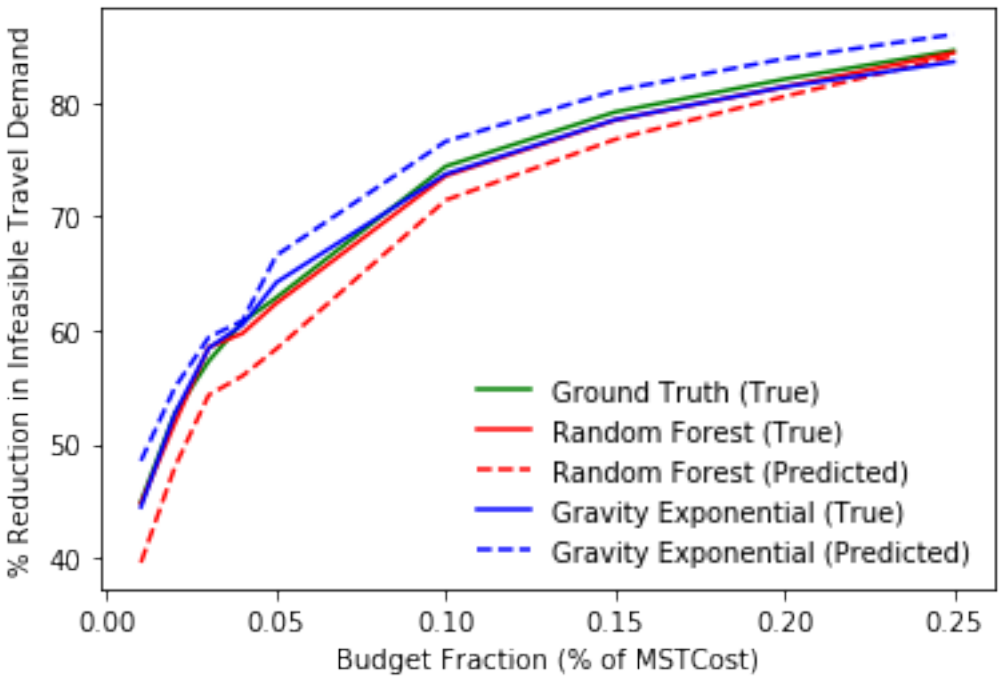



Figure 3.10: Predicted and realized recovery of infeasible trips in Frederick County, MD using ground truth travel demand data and predicted travel demand from random forest and gravity models.

plans based on the predicted travel flows. Nevertheless, these results give a strong indication that indeed predictive models for OD flows can be used in guiding urban transportation mitigation planning.

## 3.6 Conclusions

We address the problem of strategically fortifying edges in an infrastructure network against failures in order to maximize satisfied demand between vertices in the network. Unlike previous work on network design that relies on integer programming-based methods, we show that our optimization objective exhibits the property of restricted supermodularity, connecting budget-constrained prize-collecting Steiner forest to the vast literature on submodular/supermodular optimization. We demonstrate how to extend recent work on constrained supermodular maximization to our *restricted* supermodular setting. We also propose a novel, fast algorithm for maximizing modular functions subject to a knapsack and

a budget constraint. Empirically, we show that our proposed algorithms perform as well as a greedy baseline on both synthetic and real-world networks, while typically being significantly faster. Importantly, we show that supermodularity-based algorithms have the potential to scale well to solve large practical network design problems in this family.

# CHAPTER 4
# BUDGET-CONSTRAINED COMBINATORIAL OPTIMIZATION FOR NETWORK FLOWS: WILDLIFE RESERVE DESIGN

## 4.1 Background on Spatial Capture-Recapture Models

Population density and functional landscape connectivity are both central to population persistence [94, 95]. Moreover, methods that simultaneously estimate local densities and resistance to individual movement, such as the ecological distance parametrization of spatial capture-recapture models (SCR) [96], capture interdependencies between density and connectivity that could ultimately affect population viability [95]. Specifically, SCR-based landscape connectivity metrics [97, 98, 99] describe the capacity of individuals to move through the landscape with respect to their distribution across the landscape and thus are valuable objectives for a reserve-design optimization framework. For example, density-weighted connectivity [97, 98], which can be derived from population densities and functional connectivity estimated from SCR models, was recently used as an optimization objective in landscape conservation [100]. These metrics provide an alternative to traditional reserve-design approaches in which species abundance and connectivity are decoupled and treated as separate objectives [101].

Following [96], we assume that each individual in the target species population has an activity center, which depends on the biology of the species, but can be regarded as the centroid of an animal's home range or the centroid of an individual's activities during the time of sampling. We represented the landscape as a raster or grid network of $G$ pixels of unit area indexed by $g$ or $s$, where $s$ is a pixel containing an activity center. The realized population density $N(s)$ of pixel $s$ is then the number of individuals whose activity centers are located within that pixel. Each pixel $g$ also has an associated movement cost

$e^{\alpha_2 z(g)}$ related to the local resistance caused by pixel-specific landscape covariate values $z(g)$, where $\alpha_2$ parametrizes the extent to which landscape structure increases resistance to animal movement. The ecological distance $d_{\text{ecol}}(g, s)$ between a pair of pixels is measured as the sum of movement costs along the least-cost path between them [96, 97]. The probability that a pixel $g$ is used by an individual whose activity center is in pixel $s$ is modeled using a Gaussian kernel:

$$Pr(g, s) = \exp\left[-\alpha_1 d^2_{\text{ecol}}(g, s)\right] \tag{4.1}$$

where $\alpha_1$ is $\frac{1}{2\theta^2}$ and $\theta$ is the radius of a home range and the distance at which an individual could be detected from their activity center. Thus, $Pr(g, s)$ describes the probability of use for pixels based on their distance from an individual's activity center ($\alpha_1$) and the resistance to movement across pixels characterized by $\alpha_2$, resulting in an asymmetrical home-range kernel representing how individuals utilize space around their activity centers. Further details on SCR model assumptions and estimation can be found in [96] and [97].

## 4.2 Optimal Wildlife Reserve Design

Given a fixed budget, the purchasing cost of each pixel, and pixel-wise estimated local population densities and use probabilities, the goal of the budget-constrained reserve design optimization problem is to select a set of pixels to purchase that has the greatest conservation value. We formulated this problem as an integer linear program in which decisions about selecting pixels are encoded in binary variables, the limited budget is expressed as a mathematical constraint, and the value of a set of purchasing decisions is quantified in terms of the number of protected individuals and functional connectivity of the resulting reserve.

We defined a binary decision variable $x_g$ for each pixel g in the landscape to encode the decision of whether or it is purchased for the reserve (i.e., if pixel $g$ is selected, $x_g = 1$;

otherwise, $x_g = 0$). Given the purchasing cost of each pixel $c_g$, we also required that the total cost of the selected pixels not exceed the budget $B$. A feasible solution to the reserve design ILP is an assignment of binary values to each $x_g$ variable such that the total purchasing cost of those pixels with $x_g = 1$ is within the budget $B$.

We expressed our conservation goals as optimization objectives to be maximized. Our first conservation objective was to maximize the number of individuals protected by the reserve design. This is similar to the maximum coverage site selection problem [102] in which the goal is to protect as many conservation targets as possible with finite resources. The full ILP for maximizing the number of individuals within the reserve, given by the protected realized density (RD) objective is:

$$\max \sum_{g \in G} x_g \cdot N(g) \tag{4.2}$$

$$\text{s.t.} \sum_{g \in G} x_g \cdot c_g \leq B \tag{4.3}$$

$$x_g \in \{0, 1\} \quad \forall g \in G \tag{4.4}$$

The expression in Equation Equation 4.2 sums the estimated local population density over purchased pixels $g$ indicated by $x_g = 1$, excluding density from unpurchased pixels with $x_g = 0$. This is based on the assumption that purchasing a pixel $g$ is sufficient to protect the $N(g)$ individuals with activity centers located within that pixel. The constraint in Equation Equation 4.3 ensures the cost of the purchased pixels is within budget $B$. Equation Equation 4.4 constrains the $x_g$ variables to binary values.

Another objective was to maximize the extent to which conserved individuals can access the purchased pixels. Reserve-design approaches that minimize functional distance between conserved sites often use least-cost path modeling [103], which provides a way to characterize the impact of landscape features on animal movements and resulting population-level attributes, such as genetic differentiation [104]. We used potential con-

nectivity (PC) [97, 98] as a landscape-scale measure of functional connectivity. We define the protected potential connectivity (PC) objective as:

$$\max \sum_{g \in G} \sum_{s \in G} Pr(g, s) \cdot x_g \cdot x_s \tag{4.5}$$

This objective maximizes the accessibility of the reserve sites (for which $x_g = 1$) to individuals inhabiting protected pixels (for which $x_s = 1$). The full ILP for maximizing PC combines the above objective function with the constraints in Equations Equation 4.3 and Equation 4.4. Although Equation Equation 4.5 is an accurate formulation of the PC optimization objective, the expression involves products of decision variables $x_g$ and $x_s$, violating the requirements of a linear program, but the objective can be easily linearized. This allows us to leverage state-of-the-art linear programming solvers such as CPLEX, Gurobi or SCIP, that can solve problems with thousands of decision variables efficiently thanks to decades of algorithmic enhancements.

It may be preferable to evaluate a density-weighted variant of connectivity, which favors conserving areas that are highly accessible from sites with high local abundance. This can be quantified by density-weighted connectivity (DWC) [97, 98]. We defined the protected density-weighted connectivity (DWC) objective as:

$$\max \sum_{g \in G} \sum_{s \in G} Pr(g, s) \cdot x_g \cdot x_s \cdot N(s) \tag{4.6}$$

In other words, this objective maximizes the probability of selected pixels (for which $x_g = 1$) being used by individuals in the reserve (for which $x_s = 1$) and is weighted by the estimated local population density $N(s)$. We linearized the objective in Equation Equation 4.6 with the same strategy we used for the PC objective.

### 4.2.1 Home Ranges and Individual Resource Needs

In the above formulations, individuals were considered protected by a reserve design if their activity centers fell within a purchased pixel. However, in finer-resolution landscapes, the pixel containing an individual's activity center might encompass only a fraction of the area utilized by the individual to meet daily and seasonal requirements for survival. If pixels outside the reserve become inaccessible or undergo land-use change, individuals relying on those areas for resources may face increased mortality risk, even if their activity centers are located within the reserve. Thus, it may be advantageous to explicitly enforce the protection of activity centers as well as the surrounding high-use pixels to ensure that individual resource requirements are comprehensively met.

One mechanism for modeling space use by individuals is the concept of home ranges. The size and geometry of an individual's home range are directly determined by its movements about its activity center and thus depend on resistance to movement exerted by the surrounding landscape features. Thus, use probabilities (Equation Equation 4.1) provide a means of delineating the home ranges of individuals based on how they utilize space. Following [97], given an activity center in pixel s, we referred to the corresponding H% home range kernel as the set of pixels g such that the use probability $Pr(g, s) \geq 1 - \frac{H}{100}$. The 95% home range is commonly reported as delineating the entire home range of an individual [105], and we assumed that protecting the entire home range would meet all of an individual's needs to survive and persist.

We augmented our optimization model to indicate whether the full home range of an individual is protected by a set of purchased pixels. We used $A^{95\%}(s)$ to denote the set of pixels belonging to the 95% home range with activity center at $s$, comprising any pixels used by individuals at $s$ with probability at least 0.05. We also defined another binary decision variable $h_s$ for each pixel $s$ in the landscape representing whether the home range centered at pixel $s$ is protected by the reserve design. Then, if the pixels belonging to the set $A^{95\%}(s)$ are all purchased, the full 95% home range centered at pixel $s$ is conserved and

we set $h_s = 1$; otherwise, $h_s = 0$.

Under the assumption that only individuals whose full home range is within the reserve can be considered protected, the full ILP for maximizing the protected realized density objective with home-range constraints (RD-H) becomes:

$$\max \sum_{s \in G} h_s \cdot N(s) \tag{4.7}$$

$$\text{s.t.} \sum_{g \in G} x_c \cdot c_g \leq B \tag{4.8}$$

$$x_g \geq h_s \quad \forall g \in A^{95\%}(s), \ \forall s \in G \tag{4.9}$$

$$x_g \in \{0, 1\} \quad \forall g \in G \tag{4.10}$$

$$h_s \in \{0, 1\} \quad \forall s \in G \tag{4.11}$$

Equation Equation 4.9 ensures that for any protected home range ($h_s = 1$), all pixels $g$ in that home range are purchased ($x_g = 1$). Equation Equation 4.11 ensures that the decision variables $h_s$ take only binary values. We added these constraints to the protected potential connectivity and protected density-weighted connectivity maximization problems as well to get the home-range constrained versions (PC-H and DWC-H), whose objective functions are as follows:

$$\max \sum_{g \in G} \sum_{s \in G} Pr(g, s) \cdot x_g \cdot h_s \tag{4.12}$$

$$\max \sum_{g \in G} \sum_{s \in G} Pr(g, s) \cdot x_g \cdot h_s \cdot N(s) \tag{4.13}$$

### 4.2.2 Simulated Landscape Experiments and Evaluating Reserve Designs

We created simulated landscapes with the type of estimated population density and landscape resistance data that would be used in practice as inputs to the reserve design problem. We created 2, 40 $\times$ 40 pixel gridded landscapes over which we simulated a continuous

landscape covariate at 2 levels of habitat fragmentation (low or high) (Figure 4.1) approximating levels of fragmentation found for protected areas in the literature. We kept total amount of habitat constant. We simulated data for a population of $N = 100$ individuals distributed over the landscape according to an inhomogeneous point process; low values of the covariate corresponded to greater local population densities. We modeled animal movement in our landscapes after the SCR model [96, 97] in which the movement cost through a pixel $g$ with covariate $z(g)$ is given by $e^{\alpha_2 z(g)}$ and the ecological distance between pixels $g$ and $g^{'}$ is calculated by least-cost path. We set $\alpha_2 = 2.25$ [98]. The use probabilities are related to ecological distance (Equation 1) with parameter $\alpha_1 = 2.85$ and $\alpha_1 = 1.36$ for the low and high fragmentation landscapes respectively, resulting in mean home range sizes of 89 and 95 pixels for a hypothetical species. We simulated spatial capture-recapture data with a fixed detector array and then estimated pixel-wise realized densities $\hat{N}(s)$ and use probabilities $\hat{Pr}(g, s)$ with the SCR ecological distance model [96].

For each landscape, we formulated ILPs as described above based on the 3 objective functions (RD, PC, and DWC) with and without home-range constraints (95% home range or only activity center respectively). All pixels were assigned unit costs, although these could be generalized to reflect different land values or pixel availabilities. We varied the available budget from 0 to 1600 land units in increments of 100, resulting in 204 optimization problems. The resulting ILPs were solved using IBM ILOG CPLEX Studio version 12.6 in ¡5 minutes per problem. We evaluated each solution in terms of the design's protected realized density, potential connectivity, and density-weighted connectivity. Additionally, we evaluated the designs optimized without the home-range constraints against the 95% home-range area requirement to determine how disregarding home-range requirements might compromise reserve design quality. For example, we recomputed the protected realized density of the design obtained by maximizing RD without home-range constraints, but in this case only included individuals with complete home-range coverage when calculating protected realized density. This yields a more conservative estimate of

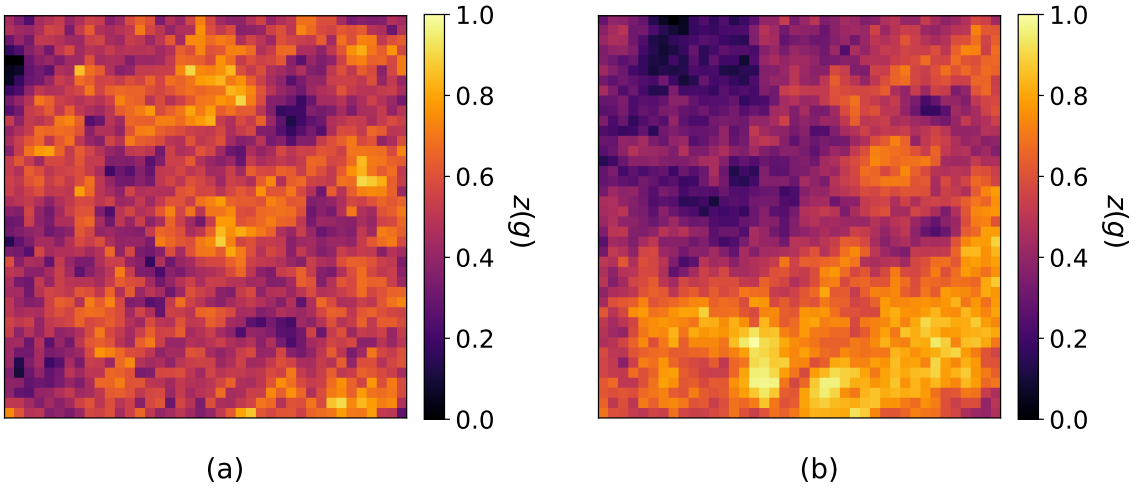



Figure 4.1: Simulated landscapes showing (a) high and (b) low habitat fragmentation, where higher values of covariate z(g) correspond to areas with less favorable habitat for a hypothetical species.

the protected density than the objective value for maximizing RD without home-range constraints by incorporating the assumption that individuals whose 95% home ranges are not fully protected by the reserve are not adequately protected by the design. Finally, we compared the spatial composition of the designs in terms of percent overlap between designs obtained by maximizing different objectives and by calculating the number of patches and aggregation index of the designs with the SDMTools R package [106].

## 4.3 Results

### 4.3.1 Conservation Objectives and Outcomes

We obtained optimal solutions for all 204 optimization problems. This meant each reserve quality measure was greatest when the corresponding objective was maximized (Figure 4.2) ; for example, protected realized density was greater for designs maximizing RD than for those maximizing PC or DWC. Without imposing home-range constraints, the purely density-driven reserve designs had the lowest protected potential connectivity, whereas the purely connectivity-driven reserve designs protected the lowest realized density out of solutions obtained using the 3 objectives. Meanwhile, maximizing the DWC objective resulted

in a compromise between maximizing the number of protected individuals and maximizing the potential connectivity between the purchased pixels.

The spatial configuration of reserves obtained by different optimization objectives was dramatically different. Maximizing RD produced reserves with the lowest aggregation index and the greatest number of patches (Table 4.1, Figure 4.3). Reserves maximizing PC always had the highest aggregation index and typically had the lowest number of patches (Table 4.1). Maximizing the DWC objective yielded reserves with intermediate aggregation index values and a comparable number of patches to the PC-optimal reserves (Table 4.1). The different objectives also prioritized different parts of the landscape for protection. Reserves maximizing RD had relatively little overlap with those maximizing PC, despite the existence of a positive correlation between high-density and high-connectivity areas within our simulated dataset. The DWC-optimal reserves overlapped significantly with both RD-optimal and PC-optimal reserves, partly by protecting nearly all of the pixels important to both RD and PC.

### 4.3.2 Individual Resource Requirements

We compared reserves obtained assuming that purchasing the activity center is adequate to protect an individual with those obtained with the 95% home-range requirements. Adding home-range constraints created more aggregated reserve designs by requiring the incorporation of pixels surrounding activity centers into the design. This was especially evident in reserves maximizing the RD objective: without 95% home-range constraints these designs were patchy because there is nothing inherent in density that naturally provides for connectivity or aggregates designs. With home-range constraints, the designs had far fewer patches (Figure 4.3). Moreover, maximizing density through the RD objective and including 95% home-range constraints achieved very different designs from combining density and connectivity by using the DWC objective alone (without home-range requirements).

Imposing home-range constraints on the reserve-design process made it more challeng-

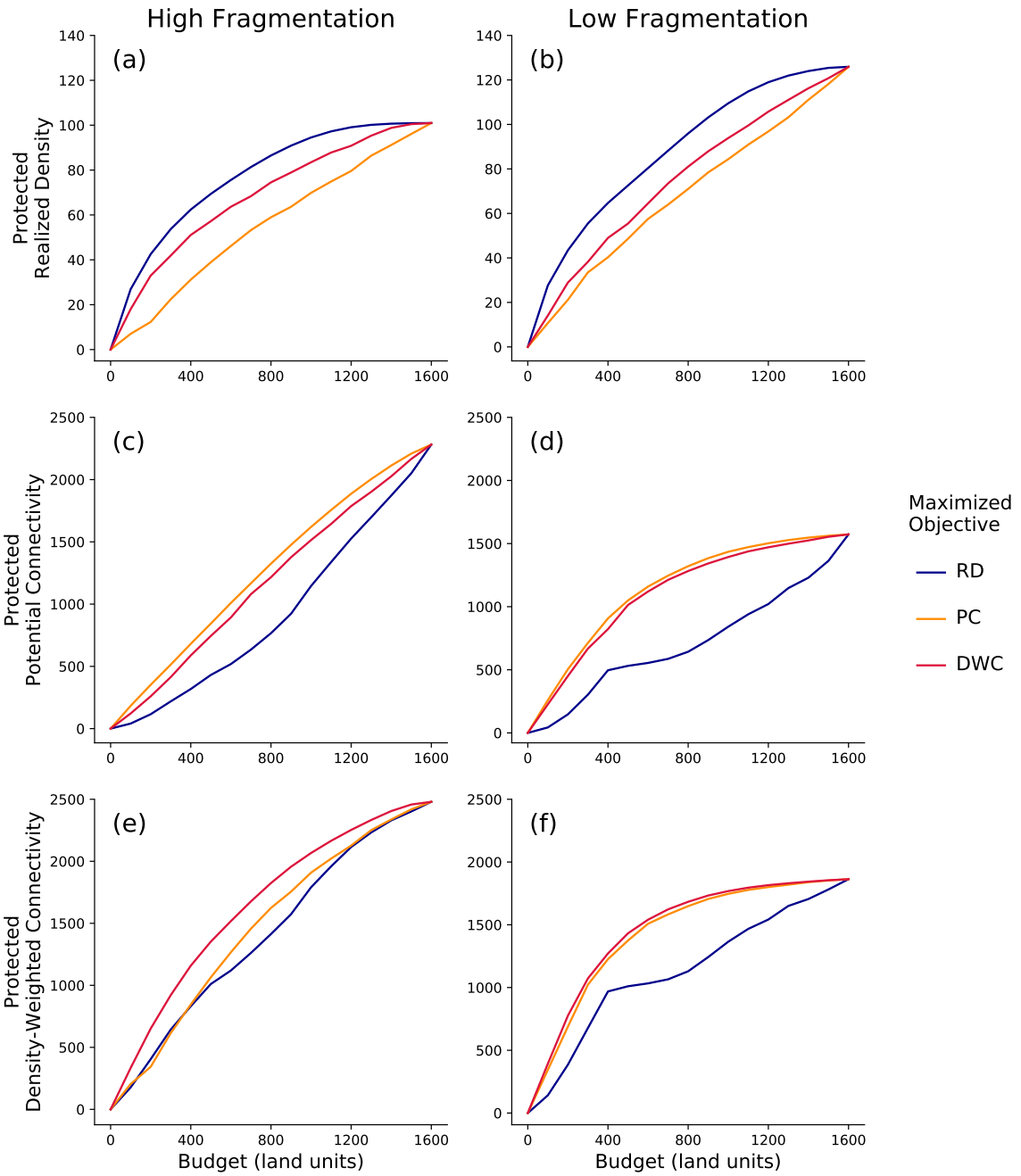



Figure 4.2: Protected realized density (a and b), protected potential connectivity (c and d) and protected density-weighted connectivity (e and f) of reserves obtained by maximizing either realized density (RD), potential connectivity (PC), or density-weighted connectivity (DWC) with different land-unit budgets. Results are for a simulated landscape with high habitat fragmentation and a simulated landscape with low habitat fragmentation.

Table 4.1: Number of reserve patches and aggregation index (AI) as calculated by SDMTools for optimal reserves obtained by maximizing realized density (RD), potential connectivity (PC), or density-weighted connectivity (DWC) without home-range constraints at select budgets and percent overlap between optimal reserves obtained for each objective.

| Fragmentation & Budget | Maximize RD | | Maximize PC | | Maximize DWC | | % Overlap | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | no. patches | AI | no.patches | AI | no. patches | AI | RD-PC | RD-DWC | PC-DWC | RD-PC-DWC |
| High | | | | | | | | | | |
| 400 | 22 | 72.60 | 2 | 95.65 | 2 | 91.57 | 34.00 | 63.00 | 49.00 | 29.25 |
| 600 | 22 | 77.32 | 2 | 95.57 | 3 | 92.70 | 40.83 | 63.17 | 62.17 | 36.17 |
| 800 | 14 | 80.10 | 2 | 95.33 | 2 | 92.68 | 51.25 | 65.38 | 82.25 | 50.38 |
| 1000 | 5 | 86.21 | 1 | 96.75 | 2 | 93.80 | 62.00 | 75.00 | 84.60 | 61.20 |
| 1200 | 1 | 91.42 | 1 | 98.20 | 1 | 95.75 | 73.67 | 84.50 | 84.58 | 69.17 |
| Low | | | | | | | | | | |
| 400 | 14 | 79.71 | 1 | 99.47 | 1 | 95.26 | 54.50 | 67.25 | 85.25 | 54.00 |
| 600 | 13 | 79.84 | 1 | 99.22 | 3 | 96.61 | 51.67 | 58.50 | 89.83 | 51.17 |
| 800 | 13 | 82.96 | 2 | 97.54 | 3 | 96.18 | 50.63 | 56.88 | 90.88 | 50.63 |
| 1000 | 7 | 85.43 | 1 | 98.40 | 5 | 96.07 | 60.50 | 69.30 | 90.40 | 60.40 |
| 1200 | 4 | 88.50 | 3 | 98.15 | 2 | 96.31 | 70.92 | 78.17 | 90.75 | 69.08 |

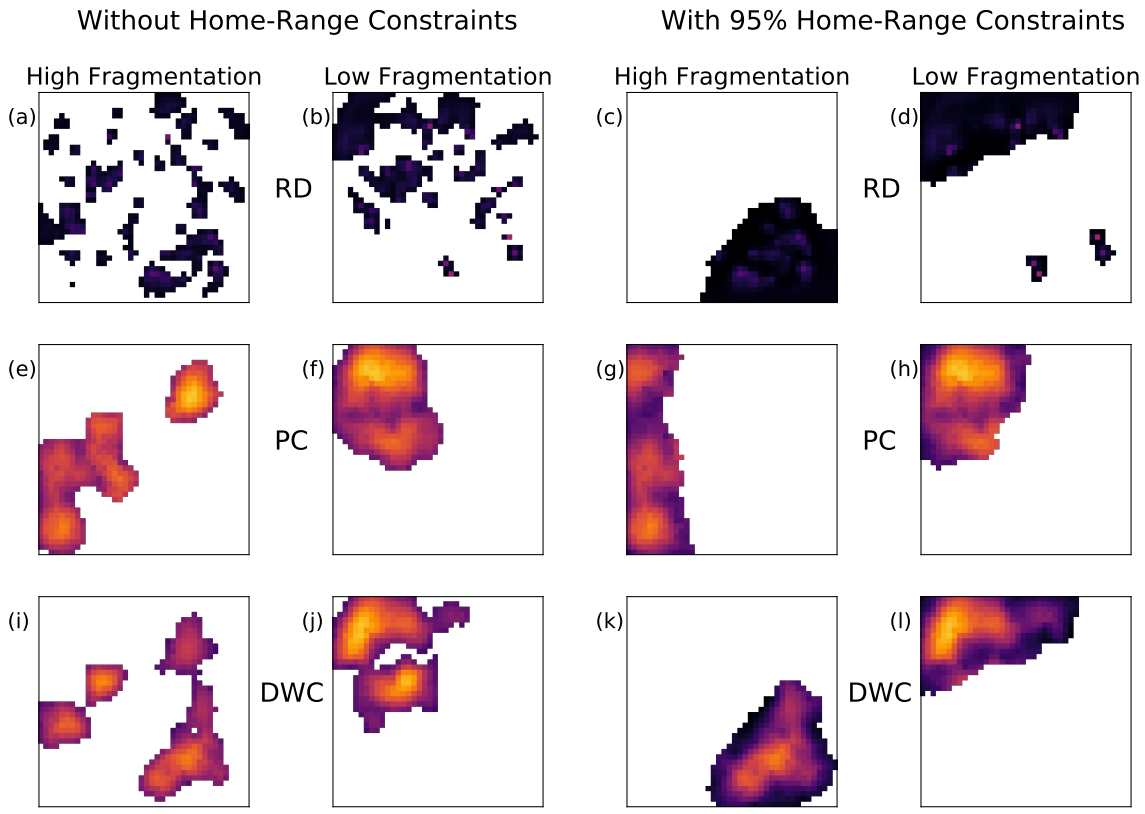



Figure 4.3: Reserve designs for simulated landscapes with high and low fragmentation obtained by maximizing the protected realized density (RD) in terms of total number of protected target species individuals in the reserve (a–d), potential connectivity (PC) of the reserve (e–h), and density-weighted connectivity (DWC) of the reserve (i–l) with a budget of 400 land units and either no home-range constraints or 95% home-range constraints.

ing to achieve reserve designs with high objective value scores, reflecting the increased cost of protecting each individual's home range compared to just their activity center. For any given budget level, the optimal reserve design objective value (for RD, PC and DWC) was lower with home-range constraints than without them (Figure 4.4, 95% versus activity center) because only the density or connectivity from fully-conserved 95% home ranges counts toward the objective when home-range requirements are considered. When reevaluating the reserve designs obtained without the home-range constraints in terms of their objective values when the 95% home range was used as the criterion for protection, designs obtained without home-range constraints had drastically lower protected RD, PC, and DWC over only the fully-protected home ranges, particularly at low budget values (Figure 4.4, activity center reevaluated with 95% versus activity center). The reduction in reserve-quality measures was substantially greater compared with when home-range area constraints were incorporated in the optimization (Figure 4.4, activity center reevaluated with 95% versus 95%).

## 4.4 Conclusions

Given the high economic and political costs associated with designing reserves, it is important to test sensible reserve design objectives that are related to population persistence. Our reserve designs considered both local population density and connectivity objectives with the goal of designing reserves that protect individuals and provide functional connectivity for those individuals. Aspects of species behavior such as resource selection and movement determine how individual animals interact with the surrounding landscape and thus influence both short-term survival of individuals and long-term persistence of the population. Our results showed that designing reserves based solely on population density can result in fragmented, patchy designs with low connectivity between reserve parcels, whereas designs that maximize only functional connectivity may achieve a small protected population size. Greater amounts of patch isolation can deter long distance dispersal [107, 108] and

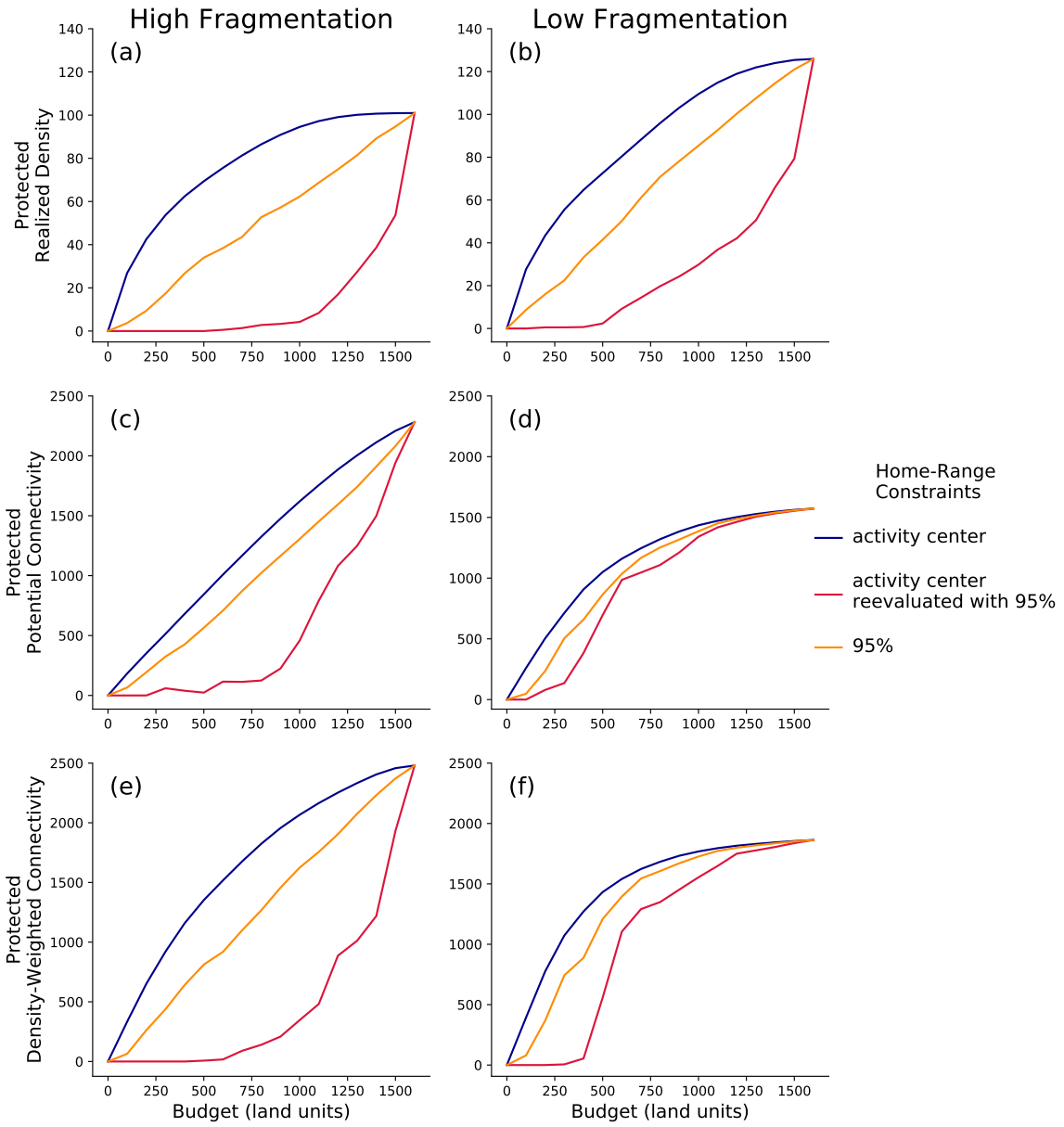



Figure 4.4: Protected realized density (a and b), protected potential connectivity (c and d), and protected density-weighted connectivity (e and f) of reserves obtained with different land-unit budgets by maximizing protected realized density (RD), potential connectivity (PC), or density weighted-connectivity (DWC) respectively without home-range constraints (activity center), with 95% home-range constraints (95%), or without home-range constraints and reevaluating RD, PC or DWC in terms of only the full home ranges in the design (activity center reevaluated with 95%).

result in decreased population sizes, inbreeding, and genetic drift when both immigration and emigration are limited [109]. This is of particular concern when areas excluded from the reserve could undergo land-use changes that could further increase resistance to movement. While resistance to dispersal may differ from resistance to daily home range movements [110], failed dispersal attempts through a dangerous matrix may alter the learned or evolved behavior of future dispersers, effectively reducing the connectivity or magnifying the isolation of a reserve network over time [111]. Instead, maximizing an objective that combines both density and connectivity or preemptively imposing home-range constraints on the reserve design are 2 ecologically meaningful strategies that yield reserves composed of spatially compact sets of parcels covering high-density areas in the landscape with high functional connectivity between them.

Density-weighted connectivity fuses functional connectivity with local population densities in an ecologically meaningful manner, rather than treating density and connectivity as 2 separate objectives in a reserve-design optimization framework. Using DWC as a conservation objective ensures that the resulting reserved parcels offer the most utility to the target population or that the probability of the conserved areas being used by protected individuals is maximized. Density-weighted connectivity is similar to the "realized connectivity" quantity described by [112], which has been linked to metapopulation persistence. Although we do not explicitly consider population dynamics or dispersal distances in this work, the DWC objective could easily be extended to this setting, for instance, by only counting connectivity between protected sites that are close enough in ecological distance for dispersal. Additionally, both the connectivity-based objectives in our model naturally result in compact reserves by maximizing the total probability of species moving between selected sites or equivalently minimizing the functional distance between selected sites [103]. When the stronger condition of contiguity is required for, say, a terrestrial species, the potential connectivity and density-weighted connectivity metrics can be modified to count only connectivity between sites that have a fully-protected path between them [113,

114].

Our model allowed us to examine the impact of designing reserves with or without explicit provisions for individual resource needs. In early reserve-design models, species occurrence was largely treated as static and the patches or sites under consideration were much larger than average home ranges [115]. To estimate landscape resistance from individual movement using SCR, we used relatively fine-resolution landscapes in which it is more realistic to model individuals as using multiple sites. Our model can accommodate varying the area or fraction of the home range to use more or less conservative thresholds (for example, 95% versus 85% home-range extents) for whether or not an individual is considered protected by the reserve design. This can be a useful framework for conservation policy makers tasked with deciding how much habitat to protect in order to support a given population. Reserves designed with home-range constraints exchange the capacity to cover a large population for potentially greater certainty that a smaller population will persist. However, reserves designed without these constraints could overestimate their conservation value, which could be undesirable for a risk-averse planner. Our approach makes these trade-offs clear and thus helps decision makers compare a range of alternatives that can be obtained by varying the home-range-extent requirement. Conceivably, then, the framework we proposed could be implemented as part of a study focused on an umbrella species.

Our approach could potentially be extended to encompass more varied reserve design goals. For example, one can address the design of reserves for multiple species using techniques from multiobjective optimization. Given estimated pixel-wise densities and pixel-to-pixel use probabilities for several target species within the landscape of interest, one can assess the objective value (such as DWC) of a given reserve design for each species separately, as we did for our single hypothetical target species. Optimizing the reserve design for multiple species simultaneously requires a weighting or ranking of target species in order of conservation priority. For a relatively small number of target species, one can construct an optimization objective as a weighted sum of objectives for each species [116].

With linear objectives and constraints, the same powerful ILP solver tools can be applied to this modified problem. Alternatively, spatial contiguity [114] could be incorporated in addition to home-range constraints. Our framework for reserve design provides decision makers with a tool for obtaining optimal designs that protect ecologically significant space-use patterns at the individual and landscape scales.

# CHAPTER 5

# DEEP LEARNING FOR CAMERA TRAP IMAGE CLASSIFICATION

Motion-activated cameras are gaining popularity as tools for conducting effective, large-scale automated biodiversity surveys [117]. Camera trap projects generate hundreds of thousands to millions of images per deployment [118] that are then reviewed by expert ecologists or a community of volunteers to detect and identify species. This takes a great deal of human annotation effort; for example, the Snapshot Serengeti project attracted 30,000 volunteers to collectively donate over 14 years of 40 h/week effort to label 5.5 million images [119].

In order to ease the burden of reviewing these data, machine learning researchers have proposed to use deep learning-based models to automatically detect and classify species from camera trap images [120]. However, there are several challenges to achieving this in practice. Deep neural networks typically require large amounts of labeled data to train, so significant annotation effort is still necessary. Moreover, image classification models typically do not generalize well to images taken in new environments [121], even when there are shared classes between the training and target datasets. This means that at present, new camera trap projects cannot directly apply pre-trained models to identify species, and must invest additional labeling effort towards training a custom model.

There are several machine learning subfields focused on developing learning strategies with smaller amounts of data with supervised information [122]. One such subfield is **transfer learning**, which aims to leverage knowledge in the form of examples or model parameters from a previously learned (source) task where training data was abundant to a new, related (target) task with limited training data. A common deep transfer learning strategy is to pre-train a model on a large, labeled source dataset to automatically learn useful feature embeddings for classification. Then, the pre-trained model is used as a feature ex-

tractor for training a new shallow model for the target task, or fine-tuned by freezing all but the final few layers of the neural network and retraining with a set of labeled samples from the target task. Another subfield is **semi-supervised learning**, which learns from a small number of labeled samples along with a large number of unlabeled samples. **Active learning** selects which samples should be labeled based on how informative they are anticipated to be for the learner.

In this chapter, we will explore several of these strategies in the context of training a classifier for identifying species from camera trap images. Since there are initially **no labeled data** from a new camera trap deployment and **obtaining labels is a labor-intensive process**, our goal is to **minimize the number of samples that need labeling to train a model for the target task to high accuracy**.

## 5.1 Few-Shot Learning via Representation Learning

### 5.1.1 Problem Statement

We are interested in classifying animals appearing in image crops from a new deployment of camera traps. These image crops can be thought of as samples from the space of all possible images drawn from some distribution. Formally, a **domain** $\mathcal{D}$ is a tuple $(\mathcal{X}, P(X))$ composed of a feature space $\mathcal{X}$ and a marginal distribution $P(X)$ over a set of instances $X = \{\mathbf{x}_i | \mathbf{x}_i \in \mathcal{X}, i = 1, \ldots, n\}$. Hence, camera trap images from one project can be considered as one image domain. To build a classifier on these images, we must build a function that maps samples from the domain distribution to species names or classes. A **task** $\mathcal{T}$ is a tuple $(\mathcal{Y}, f)$ composed of a label space $\mathcal{Y}$ and a decision function $f : \mathcal{X} \rightarrow \mathcal{Y}$ learned from the sample data $X$. As is common when working with image domains, we consider training a convolutional neural network for our task. In other words, we consider $f$ to be a feed-forward neural network $f(\mathbf{x}; \theta, \phi) \triangleq h(g(\mathbf{x}; \theta); \phi)$ consisting of a feature extractor $g(\cdot; \theta)$ that maps an input $\mathbf{x}$ to a $d$-dimensional vector embedding, and a classifier $h(\cdot; \phi)$ that maps $g(\mathbf{x}; \theta)$ to a label in $\mathcal{Y}$.

Increasingly, researchers and organizations are making pre-trained neural network models with a wide variety of architectures available for use by others directly or through transfer learning. This refers to the setting in which information learned from one or more source domains and corresponding source tasks $\{(\mathcal{D}_{S_i}, \mathcal{T}_{S_i})|i = 1, \ldots, m_S\}$ is exploited to improve performance on one or more target tasks in their target domains $\{(\mathcal{D}_{T_j}, \mathcal{T}_{T_j})|j = 1, \ldots, m_T\}$. One of the most common workflows for transfer learning with deep neural networks is to freeze layers from a model trained on the source domain(s) and task(s), append some new trainable layers, and train these on the target domain(s) and task(s). In other words, a feature extractor $g_S(\cdot; \theta_S)$ trained on the source task(s) is composed with the transformation $g_T(\cdot; \theta_T)$ applied by the appended layers to obtain inputs $g_T(g({}_S\mathbf{x}; \theta_S); \theta_T)$ to the final prediction function $h_T(\cdot; \phi_T)$.

Our goal is to learn parameters $\theta_T$ and $\phi_T$ (and possibly fine-tune parameters $\theta_S$) such that the accuracy of classifier $h_T(g_T(g_S(\mathbf{x}; \theta_S); \theta_T); \phi_T)$ on the target task of classifying species from a new camera trap project is maximized while the number of labeled samples from the target domain is minimized.

*Related Work*

**Unsupervised and Self-Supervised Representation Learning:** Given the abundance of unlabeled image data in the real world coupled with the difficulty of obtaining labels for these data for supervised learning, there has been a great deal of recent work on learning useful feature representations in a fully unsupervised manner, or alternatively by training on algorithmically generated pretext tasks with known labels for self-supervision. Our proposed work is more closely related to the former, in particular to recent work on deep clustering models such as [123, 124]. These works suggest that learning patterns in the distribution of inputs can prove useful for downstream predictive tasks. In particular, [123] show that even representations learned based on clustering inputs can improve performance on transfer tasks. We explore the potential of Gaussian mixture models, a density estimation

model commonly used for clustering, for unsupervised and semi-supervised representation learning.

**Semi-Supervised Representation Learning:** Semi-supervised learning deals with training models using a small number of labeled samples combined with relatively abundant unlabeled samples [125]. It aims to combine supervised learning, in which the goal is to learn a functional mapping between points $\boldsymbol{X}$ and their corresponding labels $\boldsymbol{Y}$, and unsupervised learning, in which data consist of only points $\boldsymbol{X}$ and one aims to learn some underlying structure in these points. Recent work by [126] illustrates the potential for semi-supervised training for deep neural networks to learn generalizable representations from limited data.

### 5.1.2 Gaussian Mixture Negative Log-Likelihood as a Loss

*Gaussian Mixture Model Preliminaries*

A Gaussian Mixture Model (GMM) is a probabilistic model for data drawn i.i.d. from $K$ multivariate Gaussian components. Each data point $\boldsymbol{x}_i \in \mathbb{R}^d$ belongs to one of the $K$ components, and we represent the component that data point $\boldsymbol{x}_i$ belongs to by the latent variable $z_i \in \{1, \ldots, K\}$. The data are assumed to be generated by first sampling $z_i \sim$ Multinomial$(\boldsymbol{\pi}, 1)$, where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ represents the probability of selecting each component, and then sampling $\boldsymbol{x}_i | z_i = k \sim \mathcal{N}(\boldsymbol{x}_i; \mu_k, \Sigma_k)$, where $\mu_k \in \mathbb{R}^d$ and $\Sigma_k \in \mathbb{R}^{d \times d}$ are the mean and covariance matrix of component $k$ respectively. The joint probability of

data point $\boldsymbol{x}_i$ and latent variable $z_i$ is given by:

$$\begin{aligned}
p(\boldsymbol{x}_i, z_i = k) &= p(z_i = k) \cdot p(\boldsymbol{x}_i | z_i = k) \\
&= \pi_k \cdot \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \Sigma_k) \\
&= \pi_k \cdot \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right] \\
&= \exp\left[\log\left\{\pi_k \cdot \frac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right]\right\}\right] \\
&= \exp\left[\log \pi_k - \frac{d}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma_k|) - \frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right]
\end{aligned} \tag{5.1}$$

The GMM thus represents a distribution with density over $\boldsymbol{x}$ given by:

$$p(\boldsymbol{x_i}) = \sum_{k=1}^{K} p(z_i = k) \cdot p(\boldsymbol{x}_i | z_i = k) \tag{5.2}$$

and the negative log-likelihood of a single point $\boldsymbol{x}_i$ is then:

$$-\log\, p(\boldsymbol{x}_i) = -\log \sum_{k=1}^{K} p(z_i = k) \cdot p(\boldsymbol{x}_i | z_i = k) \tag{5.3}$$

$$= -\log \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\boldsymbol{x}_i; \boldsymbol{\mu}_k, \Sigma_k) \tag{5.4}$$

*Minimizing the GMM Negative Log-Likelihood*

Given a neural network, $f$ with feature extractor $g(\cdot; \theta)$ that takes as input a minibatch of $N$ samples $\{\boldsymbol{x}_i | i = 1, \ldots, N\}$ and generates a $d$-dimensional embedding for each input $\{g(\boldsymbol{x}_i; \theta) | i = 1, \ldots, N\}$. We suppose we have a GMM with $K$ components and fixed parameters $\boldsymbol{\mu} \in \mathbb{R}^{K \times d}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{K \times d \times d}$, and $\boldsymbol{\pi} \in \mathbb{R}^K$. For conciseness, the GMM parameters are collectively referred to by the symbol $\boldsymbol{\phi}$. We first propose to use the negative log-likelihood of this GMM as an unsupervised loss for training the neural network. That is, we will update the parameters $\theta$ of the feature extractor such that the log-likelihood of the

resulting dataset embedding under the given fixed GMM is maximized. The unsupervised GMM loss over the minibatch of $N$ embedded data points, $\{z_1, \ldots, z_N\}$, is therefore:

$$\mathcal{L}_{\text{GMMNLL}} = -\frac{1}{N} \sum_{i=1}^{N} \log\, p(g(\boldsymbol{x}_i; \theta)). \tag{5.5}$$

We then can update the parameters of the network in a standard way with stochastic gradient descent and backpropagation. For illustration, this will start by computing the partial derivative of the loss with respect to an individual embedded point $g(\boldsymbol{x}_i)$ (dropping the dependence on $\theta$ in notation for brevity):

$$\frac{\partial \mathcal{L}_{\text{GMMNLL}}}{\partial g(\boldsymbol{x}_i)} = -\frac{1}{N} \frac{1}{\sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(g(\boldsymbol{x}_i); \mu_k, \Sigma_k)} \sum_{j=1}^{K} \pi_j \mathcal{N}(g(\boldsymbol{x}_i); \mu_j, \Sigma_j) \cdot \Sigma_j^{-1}(g(\boldsymbol{x}_i) - \mu_j) \tag{5.6}$$

Note that for a single point $\boldsymbol{x}_i$, there are stationary points when $g(\boldsymbol{x}_i) = \mu_j$ for any $j$. This highlights a tendency for the training to adapt the embedding in such a way as to move points closer to the predefined GMM component means. In fact, the embedding with the lowest negative log-likelihood would collapse all the samples onto the mean of the component with the largest value of $\frac{\pi_k}{|\Sigma_k|^{1/2}}$.

If our goal is to learn an embedding such that the density of dataset samples is modeled well by the given GMM, this behavior is undesirable. This type of degeneracy is a common problem in unsupervised loss functions. For example, [123] observe similar behavior when trying to simultaneously learn embedding features and clustering model parameters, and incorporate algorithmic mechanisms to prevent learning empty clusters and failing to learn clusters with few samples. To avoid this, we add a regularization term to Equation Equation 5.5 to encourage the network to embed points such that their resulting membership of points across GMM components closely matches the distribution defined by the given mixture weights $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$. For a minibatch of points, we let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_K)$ where $\gamma_k = \sum_{i=1}^{N} \frac{1}{N} p(z_i = k | g(\boldsymbol{x}_i))$. The vector $\boldsymbol{\gamma}$ describes how the points in the batch

are distributed between the $K$ components. The regularization term is defined based on the cross entropy between these two discrete probability distributions:

$$\begin{aligned}\mathcal{L}_{\text{reg}} &= H(\boldsymbol{\pi}, \boldsymbol{\gamma}) \\ &= -\sum_{k=1}^{K} \pi_k \log \gamma_k\end{aligned} \tag{5.7}$$

Finally, we can write the unsupervised GMM loss for the minibatch as a weighted sum of the negative log-likelihood and the regularization term:

$$L(\mathbf{X}) = \lambda_1 \mathcal{L}_{\text{GMMNLL}} + \lambda_2 \mathcal{L}_{\text{reg}} \tag{5.8}$$

Alternatively, suppose we have a frozen neural network $f$ with fixed feature extractor $g(\cdot;\theta)$, and are interested in learning GMM parameters $\phi$ to fit the dataset embedding produced by the network. Traditionally, these parameters are fit using expectation-maximization (EM), but it is nevertheless possible to learn the parameters through gradient descent (Figure 5.1).

$$\frac{\partial \mathcal{L}_{\text{GMMNLL}}}{\partial \mu_k} = \frac{1}{N} \sum_{i=1}^{N} \frac{\pi_k \cdot \mathcal{N}(g(x_i); \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(g(x_i); \mu_j, \Sigma_j)} \Sigma_k^{-1} (g(x_i) - \mu_k) \tag{5.9}$$

$$\frac{\partial L}{\partial \Sigma_k^{-1}} = -\frac{1}{2N} \sum_{i=1}^{N} \frac{\pi_k \cdot \mathcal{N}(g(x_i); \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(g(x_i); \mu_j, \Sigma_j)} \cdot \left\{ \Sigma_k^{-1} - \Sigma_k^{-1} (g(x_i) - \mu_k)(g(x_i) - \mu_k)^\top \Sigma_k^{-1} \right\} \tag{5.10}$$

$$\frac{\partial \mathcal{L}_{\text{GMMNLL}}}{\partial \pi_k} = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathcal{N}(g(x_i); \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \cdot \mathcal{N}(g(x_i); \mu_j, \Sigma_j)} \tag{5.11}$$

The learnable parameters corresponding to the GMM are unconstrained. Since $\boldsymbol{\pi}$ must be positive and sum to one, we employ the softmax function to transform the unconstrained parameters corresponding to mixture weights into the realized values. Similarly, the covariance matrices must be symmetric and positive definite. We use GMMs with diagonal covariance matrices only, which are by design symmetric; to ensure positive definiteness,

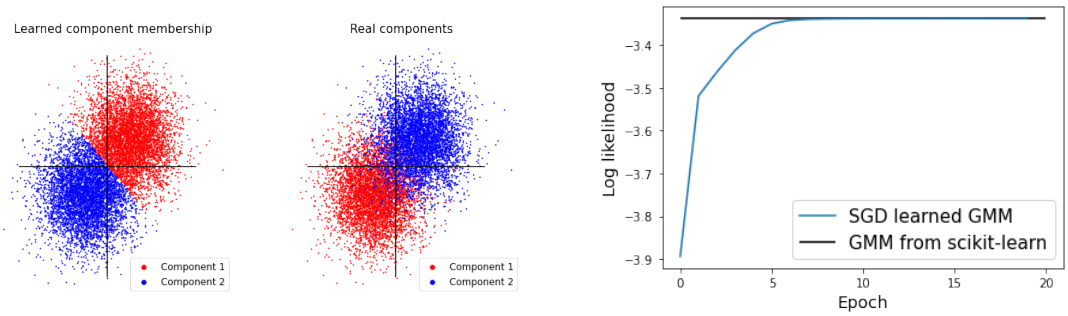
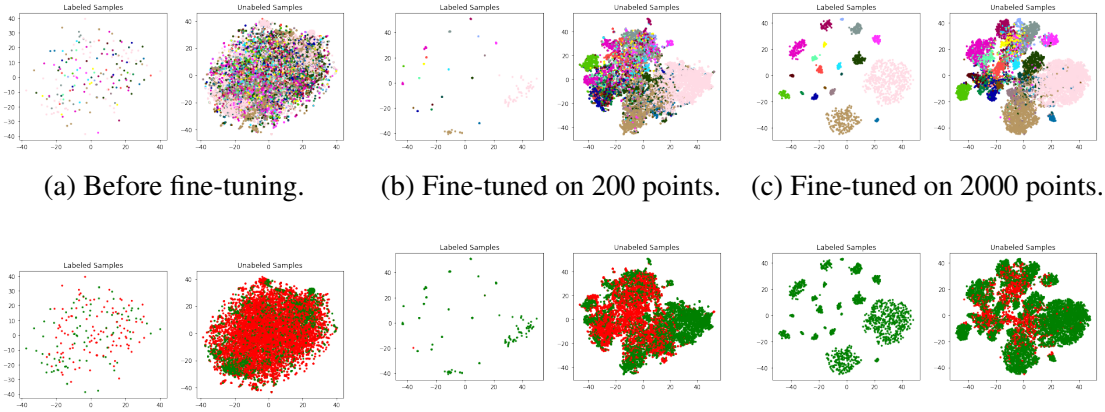



Figure 5.1: Predicted component membership and log-liklihood after fitting GMM means, covariances and mixture weights by gradient descent for a toy dataset.

we take the square of the unconstrained parameters to obtain only positive values for the diagonal elements of the covariance matrices.

### 5.1.3 Semi-Supervised Representation Learning with GMM Loss




(a) Before fine-tuning. (b) Fine-tuned on 200 points. (c) Fine-tuned on 2000 points.

(d) Errors before fine-tuning. (e) Errors after fine-tuning on 200 points. (f) Errors after fine-tuning on 2000 points.

Figure 5.2: t-SNE projections of embedding showing labeled and unlabeled points before fine-tuning (a), after fine-tuning with 200 points (b), and after fine-tuning with 2000 points (c), using cross-entropy loss for 50 epochs. In each embedding, correct (green) versus incorrect (red) predictions made by a k-nearest neighbor classifier trained on the labeled points are shown (d-f).

Deep neural networks are prone to overfitting, especially when trained using supervised learning methods on relatively small amounts of labeled data [126]. This can also occur when an existing model trained on a large source dataset is adapted to a small target dataset

[127]. To illustrate, consider fine-tuning a convolutional neural network (ResNet-50) pre-trained on a large camera trap image database—eMammal [128]—to a much smaller camera trap dataset—Missouri Camera Traps [129]—by freezing most of the network backbone and training only the weights of a shallow head network. The head is trained using cross-entropy loss on a small set of labeled samples from the target dataset that are chosen at random, so that before training the labeled samples are distributed similarly to the unlabeled points in the head network embedding (Figure 5.2a). After training the network for 50 epochs with 200 labeled points, the labeled points from each class are tightly clustered in the learned embedding even though the unlabeled points are not distributed in this way. Using more labeled points to train the network mitigates this effect. Fine-tuning the same network with 2000 labeled points results in a learned embedding in which the high-density regions corresponding to clusters of labeled points are also densely populated by unlabeled points (Figure 5.2c). Importantly, differences in the distribution of labeled and unlabeled points in the learned embedding can have consequences for performance on the target task. Classifiers trained on the labeled points in the learned embedding may be prone to making errors on unlabeled points falling in parts of the learned embedding space where there are no seen training points (Figure 5.2e), since the unlabeled points are now significantly different from those used to fit the classifier. This motivates us to fine-tune the head such that it learns an embedding that is discriminative with respect to the classes in the training set but also ensures that the distribution of the labeled training points remains similar to the distribution of the unlabeled test points.

*Proposed Method*

We propose to leverage the Gaussian Mixture Model negative log-likelihood loss described earlier to limit the degree to which the distribution of the labeled points diverges from that of the unlabeled points in the embedding representation that is learned while training the network to separate classes. Given a target density $p(g(\boldsymbol{x};\theta))$ for the embedding layer de-

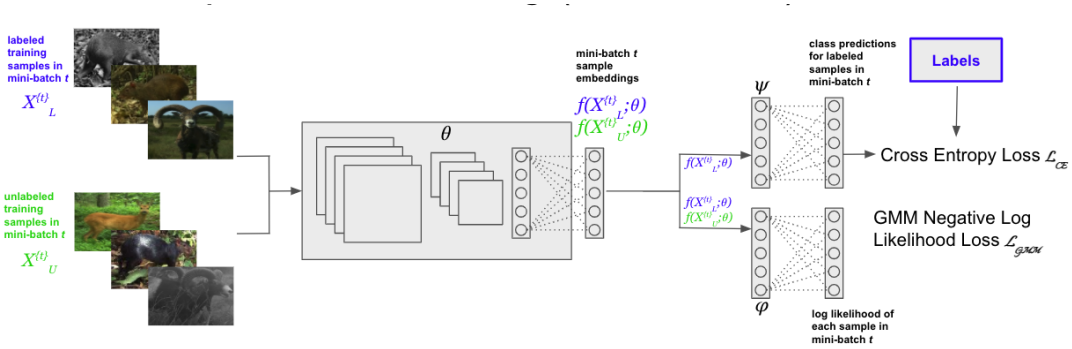



Figure 5.3: The network consists of a feature extractor with learnable parameters $\theta$, a cross-entropy head with 1 hidden layer with learnable parameters $\psi$, and a GMM head with learnable weights $\phi$.

fined by a fixed Gaussian mixture model, we can use Equation Equation 5.8 to maximize the likelihood of both the labeled and unlabeled data under the GMM. This can be considered a form of density-based regularization, which has recently been explored as a means for improving generalization in deep metric learning [130]. At the same time, we employ supervised learning to train the network to separate classes.

We employ an alternating training procedure involving supervised and semisupervised training phases. In the supervised training phase, a loss such as cross-entropy is used to train the network using only the labeled points $L$ for a fixed number of epochs $E_{\text{semisup}}$. During this phase, the training procedure is free to update the embedding in order to separate the classes in the labeled set, even if this update results in the density distribution of labeled points shifting away from that of the unlabeled points. In the semisupervised training phase, we readjust the embedding such that the labeled points and the unlabeled points together maximize the likelihood of a single target Gaussian mixture density. Specifically, for $\mathcal{L}_{\text{semisup}}$ in Algorithm Algorithm 3, we use the regularized negative log likelihood in Equation Equation 5.8, which we evaluate with respect to fixed GMM parameters $\phi$, and minimize by updating the parameters $\theta$ of the feature extractor.

**Algorithm 3** Semi-supervised network training procedure

**Input** Model $\mathcal{M}$, labeled samples $L$, unlabeled samples U

**for** $e = 1, \ldots, E$ **do**
  **for** $e_1 = 1, \ldots, E_{\text{sup}}$ **do**
    Train $\mathcal{M}$ with $\mathcal{L}_{\text{sup}}(L)$
  **for** $e_2 = 1, \ldots, E_{\text{semisup}}$ **do**
    Train $\mathcal{M}$ with $\mathcal{L}_{\text{semisup}}(L \cup U)$

*Experiments*

**Data:** We use data from two camera trap datasets: Missouri Camera Traps [129] and Caltech Camera Traps [121]. The images are preprocessed by generating crops from high-confidence bounding boxes from a pre-trained object detector. This processing pipeline allows us to also use images with multiple species or individuals of the same species; such images result in multiple bounding boxes and subsequent cropped inputs that we add to our dataset. Missouri Camera Traps contains 20 classes, while Caltech Camera Traps is filtered to remove examples annotated as "unknown" to obtain 13 classes. Both these target datasets exhibit significant class imbalance.

Table 5.1: Properties of camera trap datasets used as target classification tasks.

| | Missouri Camera Traps | Caltech Camera Traps |
|---|---|---|
| Number of Images | 24675 | 244584 |
| Number of Crops | 33137 | 91219 |
| Number of Classes | 20 | 13 |
| Number of Locations | Unknown | 140 |

**Network Architecture:** We use a ResNet-50 architecture pretrained on eMammal [128], a large annotated dataset of camera trap images. The network is frozen up to the second block of the 4th layer of ResNet-50, and the remainder of the network is replaced with 3 fully-connected layers with final layer output dimensionality 10. The outputs of the frozen ResNet backbone are of size 2048x7x7, and are first compressed using average pooling to produce 2048-dimensional vector inputs for the shallow feedforward network.

**Baseline Methods:** Following [131], we use two common, purely supervised methods for training neural networks to learn helpful embedding representations. The first way is by training the network to classify the labeled samples with cross-entropy loss. This requires the addition of a prediction layer to map the learned feature embeddings to class membership probabilities for each class in the target dataset. The second way is by training the network explicitly to move samples from the same class closer together in the learned embedding space than samples from different classes via metric learning. Triplet loss [132] is one of the most widely-used metric learning losses, and operates directly on the embedding features and therefore does not require any further modifications to the network architecture. We also employ a semi-supervised metric learning loss called neighbor embedding [126] specifically aimed at leveraging unlabeled data to improve the generalization error of a classifier trained with a small labeled dataset. This method draws on the *cluster assumption* of semi-supervised learning [125], which states that data points belonging to the same cluster should belong to the same class, or alternatively that the decision boundaries of a classifier should pass through low-density parts of the feature space.

**Evaluation:** For each dataset, we randomly select 200 samples to use as labeled data and the rest of the samples are treated as unlabeled samples available at training time. The learned representations are then used to train a k-nearest neighbor classifier ($n = 5$) using the initial 200 samples. This classifier is used to generate predictions on the unlabeled samples, which are used to compute the test accuracy Table 5.2. We also evaluate the normalized mutual information (NMI) of clusters in the learned dataset embeddings, by fitting a GMM to the dataset embeddings and assigning each point to the cluster with highest responsibility.

Table 5.2: Evaluation metrics for fine-tuning with 200 labeled samples.

| | Missouri Camera Traps | | Caltech Camera Traps | |
|---|---|---|---|---|
| Method | Accuracy | NMI | Accuracy | NMI |
| Cross-Entropy | 0.5516 | 0.4111 | 0.4654 | 0.2612 |
| Triplet | 0.5528 | 0.4244 | 0.4772 | 0.2812 |
| Neighbor Embedding | 0.5091 | 0.4027 | 0.4420 | 0.2638 |
| Cross-Entropy-GMM | 0.5271 | 0.4303 | 0.4514 | 0.2646 |
| Triplet-GMM | 0.5702 | 0.4518 | 0.4636 | 0.2760 |

Table 5.3: Evaluation metrics for fine-tuning with 500 labeled samples.

| | Missouri Camera Traps | | Caltech Camera Traps | |
|---|---|---|---|---|
| Method | Accuracy | NMI | Accuracy | NMI |
| Cross-Entropy | 0.6503 | 0.5216 | 0.5769 | 0.3529 |
| Triplet | 0.6615 | 0.5193 | 0.5855 | 0.3679 |
| Neighbor Embedding | 0.6361 | 0.4964 | 0.5628 | 0.3549 |
| Cross-Entropy-GMM | 0.6406 | 0.5045 | 0.5586 | 0.3535 |
| Triplet-GMM | 0.6567 | 0.5269 | 0.5948 | 0.3701 |

*Results*

Our proposed semi-supervised training procedure outperforms the other semi-supervised embedding training procedure (neighbor embedding) on both of our target camera trap datasets both in terms of test accuracy and in terms of cluster purity as measured by NMI. However, the purely supervised baselines achieved the highest test accuracies. Numerous studies have reported performance degradation from the introduction of unlabelled data in the training procedure, and it is impossible to determine a priori whether semi-supervised learning will help or harm task performance[125].

## 5.2 Few-Shot Learning via Active Learning

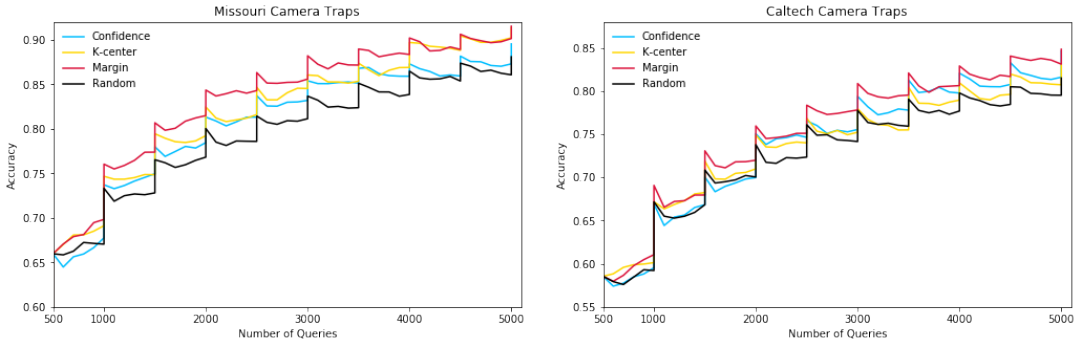
Active learning has also been explored as a strategy for improving species classification from camera trap images, in conjunction with representation learning. Here we build upon the comparison of active learning query strategies in [131].

There have been numerous categorizations proposed for active learning query methods. One distinction that can be made is between exploitation- versus exploration-bases strategies [133]. A commonly used framework for designing queries is based on uncertainty sampling: here a classifier trained on a set of labeled points can generate predictions as well as uncertainty scores for the unlabeled points, and query those points with the highest uncertainty. For example, confidence sampling queries points for which the classifier's predictions have low probability. Margin sampling queries points for which the difference between the most probable and next most probable class is slim, prioritizing decision boundaries between pairs of classes. Entropy sampling prioritizes querying points for which the classifier does not predict as belonging to any one class more than the others. All of these strategies rely exclusively on the classifier's predictions in order to inform labeling, or can be thought of as *exploiting* the classifier and trying to refine its decision boundaries. However, a major limitation of these approaches is the possibility that the learner never discovers points for which the classifier is highly confident but still incorrect. This motivates the use of *exploration*-based query strategies, in which samples are chosen based on how representative they are of unlabeled points. The k-center query strategy [134] is an example of a purely exploration-based approach.

Table 5.4: Categorization of batch active learning query strategies.

| | Exploitation-based | Mixed | Exploration-based |
|---|---|---|---|
| Myopic | Confidence | - | - |
| | Margin | - | - |
| | Triplet Embedding | - | - |
| Batch-Aware | - | AL2 | K-Center |

Another aspect that distinguishes the commonly-used uncertainty sampling strategies mentioned above from the k-center strategy in the context of pool-based batch active learning is the potential for redundancy in a batch of queried samples. In the uncertainty sampling methods, individual points are ranked in terms of the chosen uncertainty score and



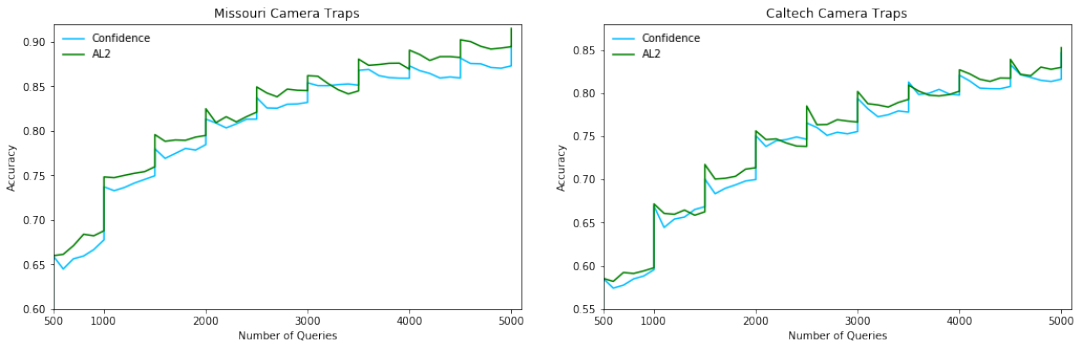

(a) Missouri Camera Traps. (b) Caltech Camera Traps.

Figure 5.4: Commonly-used active learning query strategies coupled with embedding fine-tuning with triplet loss.

points are myopically added to construct a batch of queries without attention to the impact that labeling any one point could have on the uncertainty of the rest of the points in the batch. For example, it is possible that all of the points in the batch are very similar to one another, and after labeling just one of the points the uncertainty in the remaining query points is largely resolved. In contrast, the k-center query strategy is batch-aware in the sense that the query batch is designed to collectively provide the best coverage over the remaining unlabeled points.

*Experiments*

We compare these purely exploration- and exploitation-based query strategies on the same two camera trap target datasets described earlier (Figure 5.4). Following [131], we also periodically finetune our embedding network using triplet loss. Among these strategies, margin uncertainty sampling results in the highest test accuracy across both datasets. Interestingly, confidence uncertainty sampling is a far less effective query strategy. The margin uncertainty sampling aids the classifier by resolving uncertainty around decision boundaries between pairs of classes. These points may additionally aid the construction of hard triplets during the embedding retraining phase, since around decision boundaries there are nearby points belonging to different classes.



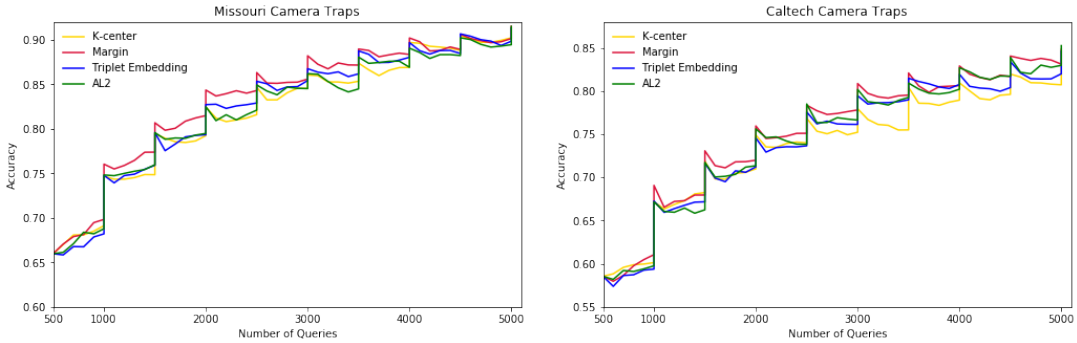

(a) Missouri Camera Traps. (b) Caltech Camera Traps.

Figure 5.5: Effect of adapting confidence uncertainty sampling to a batch-aware alternative (AL2).

Another interesting empirical finding of our analyses is that the purely exploration-driven k-center query strategy is not the best active learning method out of the set of commonly-used methods shown here. This is agreement with the findings of [131], where although k-center was the leading query method after approximately 30,000 queries, in the range of query budgets we consider here (5000) it was not as effective as margin uncertainty sampling, but more effective than confidence or random sampling.

In addition to the commonly-used query strategies described so far, we also analyze the performance of some more sophisticated query methods. The first is AL2 [135], a batch-aware query strategy that combines elements of exploration- and exploitation-based sampling. AL2 uses features such as proximity to other points and agreement among neighboring labeled samples to train a regression model for the classifier's confidence score. This enables the construction of a query batch that collectively resolves uncertainty about the rest of the unlabeled points. The performance boost of AL2 relative to confidence sampling shown in Figure 5.5 confirms the hypothesis that the myopic confidence sampling strategy selects redundant points, and that adapting the query strategy to choose a *set* of points can improve the value of information in the queried batch.

Finally, we propose a query strategy aimed at improving classifier performance indirectly via improving the effectiveness of embedding finetuning by triplet loss, which we




(a) Missouri Camera Traps. (b) Caltech Camera Traps.

Figure 5.6: Comparison of AL2 and triplet embedding against the best-performing baseline active learning query methods.

refer to as "triplet embedding". We design a query strategy that aims to select samples that result in the construction of hard triplets. The classifier is used to assign a pseudolabel of the most likely class to each unlabeled point. Then, the samples are used to mine hard triplets, and each unlabeled sample is scored according to the total loss of all mined triplets in which the sample participates. This is also a myopic, exploitation-based strategy. However, in contrast to the uncertainty sampling methods discussed earlier, triplet embedding falls into to a less common family of active learning query strategies based on expected model change [136].

Figure 5.6 compares the performance of AL2 and triplet embedding to k-center and margin sampling (random and confidence sampling are the worst performing methods and are omitted for clarity). Despite its simplicity, margin-based sampling is consistently the most effective active learning strategy for training a classifier with few labeled samples. Given that margin sampling is also a myopic strategy, it is possible that it could be further improved by adapting to the batch active learning setting, similar to the performance gains of AL2 relative to confidence sampling.

# CHAPTER 6
# CONCLUSIONS

In this dissertation we study three sustainability-related application domains as sources of compelling new computational problems at the junction of large-scale spatiotemporal modeling and prediction and discrete optimization in networks: 1) limiting the spread of invasive alien species; 2) improving the climate resilience of critical infrastructure; and 3) protecting ecologically important sites in wildlife reserves. We use spatially explicit network models to capture the underlying system dynamics of interest in each setting, and contribute discrete optimization problem formulations for maximizing sustainability objectives with finite resources.

In all three problem domains, a natural extension of our work on proposing *upfront* optimized plans is to instead optimize actions taken over *multiple time steps*. This brings significant algorithmic and scalability challenges. For instance, it is unrealistic to expect that an optimized road fortification plan can be implemented at once, and more likely the upgrades will be conducted in stages. This generates the alternate optimization problem of finding the optimal schedule for implementing the plan, requiring novel solutions involving sequential decision-making. Another practical consideration in settings when actions are unrolled over time is the potential for the underlying system dynamics to change between actions. This motivates the need to develop adaptive management approaches or policies leveraging models such as Markov Decision Processes.

Another important direction for future work has to do with the fact that the optimization problems in each setting rely on data from predictive models, driving the need to explore the impacts of errors or uncertainty in the predictions on decision-making. For example, a decision-maker may wish to take account of a discrete set of predicted scenarios to develop an optimized plan that has the best outcomes in expectation. Alternatively, one might model

the system uncertainty by allowing each model parameter to take values within a given range or interval, and the goal may be to find the plan with the best worst-case outcome, by drawing on algorithmic design approaches from the field of robust optimization. A different and increasingly popular direction for improving the performance of optimization algorithms that use model predictions as inputs is to directly train the predictive model in such a way that its predictions lead to good decision outcomes. This is a new area of study known as decision-focused learning [137, 138, 139, 140] and can be especially useful for guiding the predictive model building stage given the limited data to train models.

A final group of future directions that apply to all the methods developed in this dissertation has to do with facilitating the adoption of these computational techniques by end-users. There is a growing interest in building trust in artificial intelligence and machine learning approaches. One straightforward way to build trust in predictive models is to validate them on more case studies as additional data becomes available. Trust in machine learning models can also be increased by improving their interpretability and explainability, for example by analyzing relative feature importances or by using a simpler model to approximate the behavior of a more complicated one. Similar efforts are needed to improve the explainability of optimization model outputs, which will require require contributions from optimization researchers, problem domain experts, and human-computer interaction scientists. Finally, in order for these computational advances to truly have real-world impact, they must also be embedded into tools and interfaces that are accessible to the practitioners and decision-makers targeted as end users of this technology.

## REFERENCES


[1] G. H. Brundtland, M. Khalid, S. Agnelli, S. Al-Athel, and B. Chidzero, "Our common future", *New York*, vol. 8, 1987.

[2] T. G. Dietterich, E. W. Dereszynski, R. A. Hutchinson, and D. R. Sheldon, "Machine learning for computational sustainability.", in *International Green Computing Conference*, 2012, p. 1.

[3] M. Kimura, K. Saito, and H. Motoda, "Minimizing the spread of contamination by blocking links in a network.", in *AAAI*, vol. 8, 2008, pp. 1175–1180.

[4] D. Sheldon, B. Dilkina, A. Elmachtoub, R. Finseth, A. Sabharwal, J. Conrad, C. Gomes, D. Shmoys, W. Allen, O. Amundsen, *et al.*, "Maximizing the spread of cascades using network design", in *Proc. of the 26th Conference on Uncertainty in Artificial Intelligence*, AUAI Press, 2010, pp. 517–526.

[5] E. B. Khalil, B. Dilkina, and L. Song, "Scalable diffusion-aware optimization of network topology", in *Proc. of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2014, pp. 1226–1235.

[6] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network", in *Proc. of the th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2003, pp. 137–146.

[7] J. Yang and J. Leskovec, "Modeling information diffusion in implicit networks", in *ICDM*, IEEE, 2010, pp. 599–608.

[8] D. Romero, B. Meeder, and J. Kleinberg, "Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter", in *Proc. of the 20th International Conference on World Wide Web*, ACM, 2011, pp. 695–704.

[9] K. Eames and M. Keeling, "Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases", *PNAS*, vol. 99, no. 20, pp. 13 330–13 335, 2002.

[10] M.-A. Rizoiu, S. Mishra, Q. Kong, M. Carman, and L. Xie, "Sir-hawkes: On the relationship between epidemic models and hawkes point processes", *arXiv preprint arXiv:1711.01679*, 2017.

[11] M. Farajtabar, N. Du, M. G. Rodriguez, I. Valera, H. Zha, and L. Song, "Shaping social activity by incentivizing users", in *NIPS*, 2014, pp. 2474–2482.

[12] H. Kim, *Spatio-temporal point process models for the spread of avian influenza virus (H5N1)*. UC Berkeley, 2011.

[13] E. Balderama, F. P. Schoenberg, E. Murray, and P. Rundel, "Application of branching models in the study of invasive species", *Journal of the American Statistical Association*, vol. 107, no. 498, pp. 467–476, 2012.

[14] X. Wu, D. Sheldon, and S. Zilberstein, "Efficient algorithms to optimize diffusion processes under the independent cascade model", *NIPS Work. on Networks in the Social and Information Sciences*, 2015.

[15] M. Farajtabar, X. Ye, S. Harati, L. Song, and H. Zha, "Multistage campaigning in social networks", in *NIPS*, 2016, pp. 4718–4726.

[16] Y. Wang, E. Theodorou, A. Verma, and L. Song, "Steering opinion dynamics in information diffusion networks", *arXiv preprint arXiv:1603.09021*, 2016.

[17] A. Zarezade, U. Upadhyay, H. Rabiee, and M. Gomez-Rodriguez, "Redqueen: An online algorithm for smart broadcasting in social networks", in *Proc. of the 10th ACM International Conference on Web Search and Data Mining*, ACM, 2017, pp. 51–60.

[18] Y. Ogata and J. Zhuang, "Space–time etas models and an improved extension", *Tectonophysics*, vol. 413, no. 1, pp. 13–23, 2006.

[19] A. Hastings, K. Cuddington, K. Davies, C. Dugaw, S. Elmendorf, A. Freestone, S. Harrison, *et al.*, "The spatial spread of invasions: New developments in theory and evidence", *Ecology Letters*, vol. 8, no. 1, pp. 91–101, 2005.

[20] M. Fitzpatrick, E. Preisser, A. Porter, J. Elkinton, and A. Ellison, "Modeling range dynamics in heterogeneous landscapes: Invasion of the hemlock woolly adelgid in eastern north america", *Ecol Appl*, vol. 22, no. 2, pp. 472–486, 2012.

[21] A. Pauchard, A. Escudero, R. A. Garcıa, M. Cruz, B. Langdon, L. A. Cavieres, and J. Esquivel, "Pine invasions in treeless environments: Dispersal overruns microsite heterogeneity", *Ecology and Evolution*, vol. 6, no. 2, pp. 447–459, 2016.

[22] A. Sakai, F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, *et al.*, "The population biology of invasive species", *Annual Review of Ecology and Systematics*, vol. 32, no. 1, pp. 305–332, 2001.

[23] M. Arim, S. Abades, P. Neill, M. Lima, and P. Marquet, "Spread dynamics of invasive species", *PNAS*, vol. 103, no. 2, pp. 374–378, 2006.

[24] C. Halpern, *Ecology and restoration of montane meadows at bunchgrass ridge near the andrews experimental forest, 1999-2009. bunchgrass study.* 2012.

[25] M. Wittmann, D. Metzler, W. Gabriel, and J. Jeschke, "Decomposing propagule pressure: The effects of propagule size and propagule frequency on invasion success", *Oikos*, vol. 123, no. 4, pp. 441–450, 2014.

[26] C. Taylor and A. Hastings, "Finding optimal control strategies for invasive species: A density-structured model for spartina alterniflora", *J Appl Ecol*, vol. 41, no. 6, pp. 1049–1057, 2004.

[27] "Transforming our world: The 2030 agenda for sustainable development", *Division for Sustainable Development Goals: New York, NY, USA*, 2015.

[28] P. Chinowsky and C. Arndt, "Climate change and roads: A dynamic stressor–response model", *Review of Development Economics*, vol. 16, no. 3, pp. 448–462, 2012.

[29] M. Panteli and P. Mancarella, "Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events", *IEEE Systems Journal*, 2015.

[30] T. Huang and B. Dilkina, "Enhancing seismic resilience of water pipe networks", in *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies*, 2020, pp. 44–52.

[31] E. Pournaras, R. Taormina, M. Thapa, S. Galelli, V. Palleti, and R. Kooij, "Cascading failures in interconnected power-to-water networks", *ACM SIGMETRICS Performance Evaluation Review*, vol. 47, no. 4, pp. 16–20, 2020.

[32] G. Forzieri, A. Bianchi, F. B. e Silva, M. A. M. Herrera, A. Leblois, C. Lavalle, J. C. Aerts, and L. Feyen, "Escalating impacts of climate extremes on critical infrastructures in europe", *Global environmental change*, vol. 48, pp. 97–107, 2018.

[33] O. E. Wing, P. D. Bates, A. M. Smith, C. C. Sampson, K. A. Johnson, J. Fargione, and P. Morefield, "Estimates of present and future flood risk in the conterminous united states", *Environmental Research Letters*, vol. 13, no. 3, p. 034 023, 2018.

[34] *Nws annual flood loss summary reports to u.s. army corps of engineers*, https://www.weather.gov/water/, Accessed: 2020-09-15, 2020.

[35] T. Bles, M. van der Doef, R. van Buren, J. Buma, R. Brolsma, A. Venmans, and J. van Meerten, "Investigation of the blue spots in the netherlands national highway network", *Deltares rapport*, pp. 1 205 568–000, 2012.

[36] W. Kron, "Flood risk= hazard• values• vulnerability", *Water international*, vol. 30, no. 1, pp. 58–68, 2005.

[37] B. Hankin, I. Craigen, W. Rogers, J. Morphet, A. Bailey, and M. Whitehead, "Flood risk to the strategic road network in england", in *E3S Web of Conferences*, EDP Sciences, vol. 7, 2016, p. 10 001.

[38] S. Kulp and B. H. Strauss, "Rapid escalation of coastal flood exposure in us municipalities from sea level rise", *Climatic Change*, vol. 142, no. 3-4, pp. 477–489, 2017.

[39] A. Gupta, C. Robinson, and B. Dilkina, "Infrastructure resilience for climate adaptation", in *Proc. of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, ACM, 2018, p. 28.

[40] R. Pant, J. W. Hall, and S. P. Blainey, "Vulnerability assessment framework for interdependent critical infrastructures: Case-study for great britain's rail network", *European Journal of Transport and Infrastructure Research*, vol. 16, no. 1, 2016.

[41] Y. Casali and H. R. Heinimann, "A topological characterization of flooding impacts on the zurich road network", *PLoS one*, vol. 14, no. 7, e0220338, 2019.

[42] N. Zhang and A. Alipour, "Integrated framework for risk and resilience assessment of the road network under inland flooding", *Transportation research record*, vol. 2673, no. 12, pp. 182–190, 2019.

[43] S. Peeta, F. S. Salman, D. Gunnec, and K. Viswanath, "Pre-disaster investment decisions for strengthening a highway network", *Comput Oper Res*, vol. 37, no. 10, pp. 1708–1719, 2010.

[44] K. Kumar, J. Romanski, and P. Van Hentenryck, "Optimizing infrastructure enhancements for evacuation planning.", in *AAAI*, 2016, pp. 3864–3870.

[45] X. Wu, D. Sheldon, and S. Zilberstein, "Optimizing resilience in large scale networks.", in *AAAI*, 2016, pp. 3922–3928.

[46] D. T. Aksu and L. Ozdamar, "A mathematical model for post-disaster road restoration: Enabling accessibility and evacuation", *Transportation Research Part E: Logistics and Transportation Review*, vol. 61, pp. 56–67, 2014.

[47] F. Liberatore, M. T. Ortuño, G. Tirado, B. Vitoriano, and M. P. Scaparra, "A hierarchical compromise model for the joint optimization of recovery operations and distribution of emergency goods in humanitarian logistics", *Comput Oper Res*, vol. 42, pp. 3–13, 2014.

[48] *Fact sheet: Flooding–our nation's most frequent and costly natural disaster*, https://www.hsdl.org/?view&did=27767, Accessed: 2020-09-15, 2010.

[49] P. Chinowsky, A. Schweikert, N. Strzepek, K. Manahan, K. Strzepek, and C. A. Schlosser, "Climate change adaptation advantage for african road infrastructure", *Climatic change*, vol. 117, no. 1-2, pp. 345–361, 2013.

[50] *Openstreetmap*, Accessed: 2018-03-01, 2018.

[51] https://msc.fema.gov/portal/advanceSearch, 2019.

[52] C. C. Sampson, A. M. Smith, P. D. Bates, J. C. Neal, L. Alfieri, and J. E. Freer, "A high-resolution global flood hazard model", *Water resources research*, vol. 51, no. 9, pp. 7358–7381, 2015.

[53] SSBN, *Ssbn-global senegal*, http://www.ssbn.co.uk/ssbnglobal, Accessed: 2017-06-25, 2017.

[54] *Nasa modis*, Accessed: 2018-03-01, 2018.

[55] *Modeling zones*, https://datahub.cmap.illinois.gov/dataset/cmap-modeling-zone-systems, 2009.

[56] *Traffic analysis zone - 2010*, https://www.psrc.org/gis-shapefiles, 2010.

[57] *Taz*, http://rtdc-mwcog.opendata.arcgis.com/datasets/taz, 2019.

[58] Orange Data for Development Challenge, *Book of abstract : Scientific paper*, http://www.d4d.orange.com/fr/content/download/43453/406503/version/1/file/D4DChallengeSenegal_Book_of_Abstracts_Scientific_Papers.pdf, Accessed: 2017-06-24, 2013.

[59] *Data for climate action challenge data*, Accessed: 2018-03-01, 2017.

[60] Z. Liu, F. Miranda, W. Xiong, J. Yang, Q. Wang, and C. Silva, "Learning geo-contextual embeddings for commuting flow prediction", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 808–816.

[61] G. Spadon, A. C. de Carvalho, J. F. Rodrigues-Jr, and L. G. Alves, "Reconstructing commuters network using machine learning and urban indicators", *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[62] M. Lenormand, S. Huet, F. Gargiulo, and G. Deffuant, "A universal model of commuting networks", *PloS one*, vol. 7, no. 10, e45985, 2012.

[63] G. Qiu, A. Gupta, C. Robinson, S. Feng, and B. Dilkina, "Learning-based travel prediction in urban road network resilience optimization", 2020.

[64] M. Lenormand, A. Bassolas, and J. J. Ramasco, "Systematic comparison of trip distribution laws and models", *Journal of Transport Geography*, vol. 51, pp. 158–169, 2016.

[65] H. C. Carey, *Principles of social science*. JB Lippincott & Company, 1867, vol. 3.

[66] G. K. Zipf, "The p 1 p 2/d hypothesis: On the intercity movement of persons", *American sociological review*, vol. 11, no. 6, pp. 677–686, 1946.

[67] S. Erlander and N. F. Stewart, *The gravity model in transportation analysis: theory and extensions*. Vsp, 1990, vol. 3.

[68] S. A. Stouffer, "Intervening opportunities: A theory relating mobility and distance", *American sociological review*, vol. 5, no. 6, pp. 845–867, 1940.

[69] M. Schneider, "Gravity models and trip distribution theory", *Papers in Regional Science*, vol. 5, no. 1, pp. 51–56, 1959.

[70] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns", *Nature*, vol. 484, no. 7392, p. 96, 2012.

[71] Y. Yang, C. Herrera, N. Eagle, and M. C. González, "Limits of predictability in commuting flows in the absence of data for calibration", *Scientific reports*, vol. 4, p. 5662, 2014.

[72] J. E. Anderson, "The gravity model", National Bureau of Economic Research, Working Paper 16576, Dec. 2010.

[73] I. S. Lowry, *Migration and metropolitan growth: two analytical models*. Chandler Pub. Co., 1966.

[74] C. Blanco, "The determinants of interstate population movements", *Journal of Regional Science*, vol. 5, no. 1, pp. 77–84, 1963.

[75] W. Alonso, *A theory of movements*. Inst. of Urban and Regional Devel., 1976.

[76] F. Gargiulo, M. Lenormand, S. Huet, and O. B. Espinosa, "Commuting network models: Getting the essentials", *Journal of Artificial Societies & Social Simulation*, vol. 15, no. 2, p. 6, 2012.

[77] M. T. Hajiaghayi and K. Jain, "The prize-collecting generalized steiner tree problem via a new approach of primal-dual schema", in *Proc. of the 17th annual ACM-*

*SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 2006, pp. 631–640.

[78] N. Tuncbag, A. Braunstein, A. Pagnani, S.-S. C. Huang, J. Chayes, C. Borgs, R. Zecchina, and E. Fraenkel, "Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem", *J Comput Biol*, vol. 20, no. 2, pp. 124–136, 2013.

[79] S. Chawla, K. Garimella, A. Gionis, and D. Tsang, "Backbone discovery in traffic networks", *International Journal of Data Science and Analytics*, vol. 1, no. 3-4, pp. 215–227, 2016.

[80] M. Bateni, C. Chekuri, A. Ene, M. T. Hajiaghayi, N. Korula, and D. Marx, "Prize-collecting steiner problems on planar graphs", in *Proc. of the 22nd annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 2011, pp. 1028–1049.

[81] D. Segev and G. Segev, "Approximate k-steiner forests via the lagrangian relaxation technique with internal preprocessing", in *European Symposium on Algorithms*, Springer, 2006, pp. 600–611.

[82] Y.-S. Myung and H.-j. Kim, "A cutting plane algorithm for computing k-edge survivability of a network", *European Journal of Operational Research*, vol. 156, no. 3, pp. 579–589, 2004.

[83] T. C. Matisziw and A. T. Murray, "Modeling s–t path availability to support disaster vulnerability assessment of network infrastructure", *Comput Oper Res*, vol. 36, no. 1, pp. 16–26, 2009.

[84] K. Viswanath and S. Peeta, "Multicommodity maximal covering network design problem for planning critical routes for earthquake response", *Transportation Research Record: Journal of the Transportation Research Board*, no. 1857, pp. 1–10, 2003.

[85] I. Shames and T. H. Summers, "Rigid network design via submodular set function optimization", *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 3, pp. 84–96, 2015.

[86] N. Mehr and R. Horowitz, "A submodular approach for optimal sensor placement in traffic networks", in *2018 Annual American Control Conference (ACC)*, IEEE, 2018, pp. 6353–6358.

[87] D.-Z. Du, R. L. Graham, P. M. Pardalos, P.-J. Wan, W. Wu, and W. Zhao, "Analysis of greedy approximations with nonsubmodular potential functions", in *Proc. of the*

*19th annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 2008, pp. 167–175.

[88] A. Aouad, R. Levi, and D. Segev, "Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences", 2015.

[89] W. Bai and J. Bilmes, "Greed is still good: Maximizing monotone submodular+supermodular (bp) functions", in *ICML*, 2018, pp. 314–323.

[90] R. Iyer, S. Jegelka, and J. Bilmes, "Fast semidifferential-based submodular function optimization", in *ICML*, 2013, pp. 855–863.

[91] K. K. Sarpatwar, B. Schieber, and H. Shachnai, "Constrained submodular maximization via greedy local search", *arXiv preprint arXiv:1705.06319*, 2017.

[92] C. Chekuri, J. Vondrak, and R. Zenklusen, "Dependent randomized rounding via exchange properties of combinatorial structures", in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, IEEE, 2010, pp. 575–584.

[93] A. Gupta and B. Dilkina, "Budget-constrained demand-weighted network design for resilient infrastructure", in *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2019, pp. 456–463.

[94] L. Tischendorf and L. Fahrig, "How should we measure landscape connectivity?", *Landscape ecology*, vol. 15, no. 7, pp. 633–641, 2000.

[95] S. A. Cushman, B. W. Compton, and K. McGarigal, "Habitat fragmentation effects depend on complex interactions between population size and dispersal ability: Modeling influences of roads, agriculture and residential development across a range of life-history characteristics", in *Spatial complexity, informatics, and wildlife conservation*, Springer, 2010, pp. 369–385.

[96] J. A. Royle, R. B. Chandler, K. D. Gazenski, and T. A. Graves, "Spatial capture–recapture models for jointly estimating population density and landscape connectivity", *Ecology*, vol. 94, no. 2, pp. 287–294, 2013.

[97] C. Sutherland, A. K. Fuller, and J. A. Royle, "Modelling non-euclidean movement and landscape connectivity in highly structured ecological networks", *Methods in Ecology and Evolution*, vol. 6, no. 2, pp. 169–177, 2015.

[98] D. J. Morin, A. K. Fuller, J. A. Royle, and C. Sutherland, "Model-based estimators of density and connectivity to inform conservation of spatially structured populations", *Ecosphere*, vol. 8, no. 1, e01623, 2017.

[99] J. A. Royle, A. K. Fuller, and C. Sutherland, "Unifying population and landscape ecology with spatial capture–recapture", *Ecography*, vol. 41, no. 3, pp. 444–456, 2018.

[100] Y. Xue, X. Wu, D. Morin, B. Dilkina, A. Fuller, J. A. Royle, and C. P. Gomes, "Dynamic optimization of landscape connectivity embedding spatial-capture-recapture information", in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[101] J. C. Williams, C. S. ReVelle, and S. A. Levin, "Using mathematical optimization models to design nature reserves", *Frontiers in Ecology and the Environment*, vol. 2, no. 2, pp. 98–105, 2004.

[102] R. Church and C. R. Velle, "The maximal covering location problem", *Papers in regional science*, vol. 32, no. 1, pp. 101–118, 1974.

[103] Y. Wang and H. Önal, "Optimal design of compact and connected nature reserves for multiple species", *Conservation biology*, vol. 30, no. 2, pp. 413–424, 2016.

[104] S. A. Cushman and J. S. Lewis, "Movement behavior explains genetic differentiation in american black bears", *Landscape ecology*, vol. 25, no. 10, pp. 1613–1625, 2010.

[105] B. G. Dickson and P. Beier, "Home-range and habitat selection by adult cougars in southern california", *The Journal of Wildlife Management*, pp. 1235–1245, 2002.

[106] J. VanDerWal, L. Falconi, S. Januchowski, L. Shoo, C. Storlie, and M. J. VanDerWal, "Package 'sdmtools'", *R package*, 2014.

[107] L. Fahrig, "Non-optimal animal movement in human-altered landscapes", *Functional ecology*, vol. 21, no. 6, pp. 1003–1015, 2007.

[108] J. Cote, E. Bestion, S. Jacob, J. Travis, D. Legrand, and M. Baguette, "Evolution of dispersal strategies and dispersal syndromes in fragmented landscapes", *Ecography*, vol. 40, no. 1, pp. 56–73, 2017.

[109] L. F. Keller and D. M. Waller, "Inbreeding effects in wild populations", *Trends in ecology & evolution*, vol. 17, no. 5, pp. 230–241, 2002.

[110] A. T. Keeley, P. Beier, B. W. Keeley, and M. E. Fagan, "Habitat suitability is a poor proxy for landscape connectivity during dispersal and mating movements", *Landscape and Urban Planning*, vol. 161, pp. 90–102, 2017.

[111] M. Baguette and H. Van Dyck, "Landscape connectivity and animal behavior: Functional grain as a key determinant for dispersal", *Landscape ecology*, vol. 22, no. 8, pp. 1117–1129, 2007.

[112] J. R. Watson, D. A. Siegel, B. E. Kendall, S. Mitarai, A. Rassweiller, and S. D. Gaines, "Identifying critical regions in small-world marine metapopulations", *Proceedings of the National Academy of Sciences*, vol. 108, no. 43, E907–E913, 2011.

[113] H. Önal, Y. Wang, S. T. Dissanayake, and J. D. Westervelt, "Optimal design of compact and functionally contiguous conservation management areas", *European Journal of Operational Research*, vol. 251, no. 3, pp. 957–968, 2016.

[114] N. Jafari, B. L. Nuse, C. T. Moore, B. Dilkina, and J. Hepinstall-Cymerman, "Achieving full connectivity of sites in the multiperiod reserve network design problem", *Computers & Operations Research*, vol. 81, pp. 119–127, 2017.

[115] M. Cabeza and A. Moilanen, "Design of reserve networks and the persistence of biodiversity", *Trends in ecology & evolution*, vol. 16, no. 5, pp. 242–248, 2001.

[116] B. Dilkina, R. Houtman, C. P. Gomes, C. A. Montgomery, K. S. McKelvey, K. Kendall, T. A. Graves, R. Bernstein, and M. K. Schwartz, "Trade-offs and efficiencies in optimal budget-constrained multispecies corridor networks", *Conservation biology*, vol. 31, no. 1, pp. 192–202, 2017.

[117] O. R. Wearn and P. Glover-Kapfer, "Snap happy: Camera traps are an effective sampling tool when compared with alternative methods", *Royal Society open science*, vol. 6, no. 3, p. 181 748, 2019.

[118] E. H. Fegraus, K. Lin, J. A. Ahumada, C. Baru, S. Chandra, and C. Youn, "Data acquisition and management software for camera trap data: A case study from the team network", *Ecological Informatics*, vol. 6, no. 6, pp. 345–353, 2011.

[119] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer, "Snapshot serengeti, high-frequency annotated camera trap images of 40 mammalian species in an african savanna", *Scientific data*, vol. 2, p. 150 026, 2015.

[120] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning", *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, E5716–E5725, 2018.

[121] S. Beery, G. V. Horn, and P. Perona, "Recognition in terra incognita", in *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XVI*, 2018, pp. 472–489.

[122] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning", *ACM Computing Surveys (CSUR)*, vol. 53, no. 3, pp. 1–34, 2020.

[123] M. Caron, P. Bojanowski, A. Joulin, and M. Douze, "Deep clustering for unsupervised learning of visual features", in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 132–149.

[124] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis", in *International conference on machine learning*, 2016, pp. 478–487.

[125] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning", *Machine Learning*, vol. 109, no. 2, pp. 373–440, 2020.

[126] E. Hoffer and N. Ailon, "Semi-supervised deep learning by metric embedding", *arXiv preprint arXiv:1611.01449*, 2016.

[127] A. V. M. Barone, B. Haddow, U. Germann, and R. Sennrich, "Regularization techniques for fine-tuning in neural machine translation", *arXiv preprint arXiv:1707.09920*, 2017.

[128] S. G. Schuttler, W. McShea, R. Costello, T. Forrester, M. Baker, E. Kalies, A. Parsons, Z. He, J. Millspaugh, T. Han, *et al.*, "Emammal citizen science camera trapping–collecting big data to answer wildlife questions", in *TDWG 2015 ANNUAL CONFERENCE*, 2015.

[129] Z. Zhang, Z. He, G. Cao, and W. Cao, "Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification", *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2079–2092, 2016.

[130] Y. Li, T. Yao, Y. Pan, H. Chao, and T. Mei, "Deep metric learning with density adaptivity", *IEEE Transactions on Multimedia*, vol. 22, no. 5, pp. 1285–1297, 2019.

[131] M. S. Norouzzadeh, D. Morris, S. Beery, N. Joshi, N. Jojic, and J. Clune, "A deep active learning system for species identification and counting in camera trap images", *Methods in Ecology and Evolution*, vol. 12, no. 1, pp. 150–161, 2021. eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13504.

[132] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[133] T. Osugi, D. Kim, and S. Scott, "Balancing exploration and exploitation: A new algorithm for active machine learning", in *Fifth IEEE International Conference on Data Mining (ICDM'05)*, IEEE, 2005, 8–pp.

[134] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach", *arXiv preprint arXiv:1708.00489*, 2017.

[135] B. Dilkina, T. Damoulas, C. P. Gomes, and D. Fink, "Al2: Learning for active learning", in *Workshop" Machine Learning for Sustainability" in the 25th Conference of Neural Information Processing Systems (NIPS), Granada, Spain*, Citeseer, 2011.

[136] B. Settles, "Active learning literature survey", University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2009.

[137] B. Wilder, B. Dilkina, and M. Tambe, "Melding the data-decisions pipeline: Decision-focused learning for combinatorial optimization", in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 1658–1665.

[138] J. A. Killian, B. Wilder, A. Sharma, V. Choudhary, B. Dilkina, and M. Tambe, "Learning to prescribe interventions for tuberculosis patients using digital adherence data", in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2430–2438.

[139] A. Ferber, B. Wilder, B. Dilkina, and M. Tambe, "Mipaal: Mixed integer program as a layer", vol. 34, no. 02, pp. 1504–1511, Apr. 2020.

[140] A. Perrault, B. Wilder, E. Ewing, A. Mate, B. Dilkina, and M. Tambe, "End-to-end game-focused learning of adversary behavior in security games", vol. 34, no. 02, pp. 1378–1386, Apr. 2020.