Robust Network Design for Multispecies Conservation

Ronan Le Bras, Bistra Dilkina, Yexiang Xue, Carla P. Gomes
Cornell University, Ithaca, NY
{lebras,bistra,yexiang,gomes}@cs.cornell.edu

Kevin S. McKelvey, Michael K. Schwartz
US Forest Service, Missoula, MT
{kmckelvey,mkschwartz}@fs.fed.us

Claire A. Montgomery
Oregon State University, Corvallis, OR
claire.montgomery@oregonstate.edu

Abstract

Our work is motivated by an important network design application in computational sustainability concerning wildlife conservation. In the face of human development and climate change, it is important that conservation plans for protecting landscape connectivity exhibit certain level of robustness. While previous work has focused on conservation strategies that result in a connected network of habitat reserves, the robustness of the proposed solutions has not been taken into account. In order to address this important aspect, we formalize the problem as a node-weighted bi-criteria network design problem with connectivity requirements on the number of disjoint paths between pairs of nodes. While in most previous work on survivable network design the objective is to minimize the cost of the selected network, our goal is to optimize the quality of the selected paths within a specified budget, while meeting the connectivity requirements. We characterize the complexity of the problem under different restrictions. We provide a mixed-integer programming encoding that allows for finding solutions with optimality guarantees, as well as a hybrid local search method with better scaling behavior but no guarantees. We evaluate the typical-case performance of our approaches using a synthetic benchmark, and apply them to a large-scale real-world network design problem concerning the conservation of wolverine and lynx populations in the U.S. Rocky Mountains (Montana).

1 Introduction

In recent years, considerable research has been devoted to the study and design of robust and reliable networks, in particular in areas such as transportation and telecommunications. Transportation and communication networks are vulnerable to disruptions. For example, in telecommunications networks, sabotage, accidents or random equipment failure may occur and therefore it is important to consider redundancy in the network design in order to guarantee that the network remains functional in the case of disturbances. Different properties and metrics are used in network design to characterize the robustness and reliability of a network, depending on the particular application. Often, one critical aspect concerns the network *connectivity*.

An important measure of connectivity reliability is the number of (edge or node) disjoint paths connecting a pair of important nodes. This measure has been studied extensively in the area of communication networks design as the Survivable Network Design or the Generalized Steiner Network (GSN) Problem. In the Generalized Steiner Network problem, we are given a graph G with edge costs and connectivity requirements r_{uv} between node pairs u,v. The goal is to find a minimum-cost subgraph H of G that contains r_{uv} disjoint paths for all $u, v \in V$. The GSN problem has attracted considerable attention as it generalizes several well-studied combinatorial optimization problems such as the minimum Steiner tree problem, the k-connected subgraph problem, and the traveling salesman problem (Kerivin and Mahjoub 2005, for a survey).

Often the desirability of a network design goes beyond simply satisfying the connectivity requirements at a minimum cost. For example, the quality of service of a communications network depends not only on the number of redundant paths but also on the transmission delay along these paths. Similarly, in the context of transportation network design, one wants to provide alternative routes between important locations, but the travel time along these routes should also be taken into account. The network quality is especially relevant in our motivating application in computational sustainability (Gomes 2009) concerning wildlife conservation.

In wildlife conservation an important planning problem is that of allocating scarce conservation resources to create a protected network, which links core habitat areas in a way that is robust to unexpected natural or anthropological disturbances. The quality of the paths selected to fulfill the connectivity requirements captures the difficulty of animal dispersal along the path and therefore the likelihood that animals will successfully use the path to migrate from one habitat area to another. Hence, given a conservation budget one wants to select the paths that have least resistance (or *delay*) to animal movement.

To address this setting, we consider a budget-constrained node-weighted undirected network design problem with costs and delays. We refer to this problem as the *Minimum Delay Generalized Steiner Network* (MinDelay-GSN) problem: given a graph where nodes have corresponding costs and delays, a set of node pairs with corresponding connectivity requirements, and a fixed budget, find the subset of nodes that satisfy the budget constraint as well as the connectivity requirements while minimizing the sum of the delay-weighted lengths of the disjoint paths used to satisfy the connectivity requirements. Clearly, our problem is closely related to the generalized Steiner network problem, as satisfying the budget constraint and the connectivity requirements is exactly the decision version of node-weighted GSN. However, our application calls for a richer optimization problem where a second metric specified as the node delays is also to be considered. The use of two (or more) metrics when designing networks is well motivated and has been previously studied in the context of network design. For example, a related problem is the Shallow Light Steiner Tree problem with applications in delay-constrained multicast routing, where the goal is to select a subtree connecting a set of terminals at a minimum cost, while satisfying a bound on the delay-weighted diameter of the subtree. Another related problem is that of the cost-distance Steiner Tree where the goal is to minimize the total cost of the tree plus the delay on the paths. To our knowledge, the budget-constrained minimum delay network design problem with general connectivity requirements described here has not been addressed previously.

In a closely-related paper, Phillips et al. study a conservation problem where the goal is to minimize the area needed to be protected in order to guarantee a fixed number of dispersal corridors/paths through time, connecting suitable conditions between *time-periods* (Phillips et al. 2008). In this work, a path is a sequence of connected cells over time and the different paths are required to be disjoint. As a result, the optimization is performed over time-unrolled graphs (directed and acyclic) and over fixed-length paths (the time horizon). In comparison, we consider a more general setting, and in particular, general undirected graphs, and identify that this setting is much harder to solve in practice.

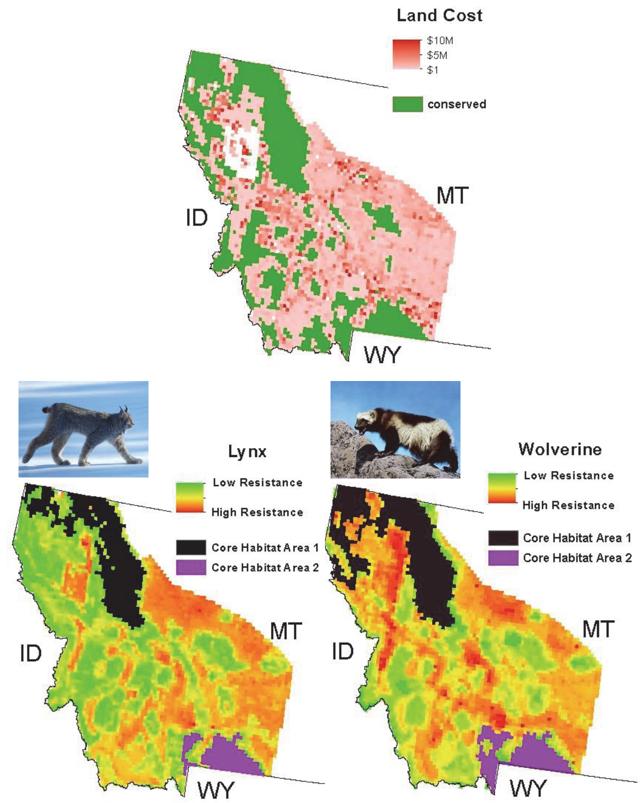

Figure 1: Western Montana case study: land cost (top), landscape resistance for lynx (bottom-left) and wolverines (bottom-right).

Our Contributions

We formalize a bi-criteria undirected network design problem with costs and delays and with general connectivity requirements, relevant to our motivating application domain in wildlife conservation planning as well as to many other domains such as QoS communication networks. We show that the problem is NP-hard with respect to both edge and node connectivity even when restricted to planar graphs or when restricted to the connectivity of a single pair of nodes. We provide a mixed integer linear programming (MILP) problem formulation that enables solution approaches with optimality guarantees. We perform a detailed typical case analysis on a synthetic benchmark. Our results show that the problem exhibits easy-hard-easy behavior and that it is hardest for budgets slightly above the minimum budget necessary to meet the connectivity requirements. However, the MILP-based approach does not scale to very large real conservation problems involving many pairs of core areas. To fill this computational gap, we also develop two novel local search methods. Designing a local search for the minimum delay generalized Steiner network problem is a challenging task due to the intricate hard constraints of the problem involving both connectivity requirements as well as a budget constraint. Our simpler local search approach is based on replacing parts of the solution with the corresponding shortest-delay path when the budget allows it. On the other hand, our hybrid local search method utilizes special moves based on solving small scale MILP subproblems to incrementally and optimally improve localized parts of the solution. We study the scaling behavior of our approaches in terms of the budget constraint as well as the number of node pairs with connectivity requirements. More details are provided in the experimental section. We apply our approaches to a large-scale network design problem concerning the conservation of wolverine and lynx populations in the U.S. Rocky Mountains, Western Montana (Lai et al. 2011)(see Fig. 1). Our results show that our MILP formulation allows us to solve optimally large-scale planning problems when they involve few pairs of habitat areas, while our local search approaches allow us to address conservation problems involving larger number of pairs when the MILP approach does not scale up.

2 Conservation Planning

As we will describe below, the domain of wildlife conservation brings new challenges to the network design problem and motivates the bi-criteria problem studied in this work.

Landscape Connectivity Under the pressures of rapid human development, wildlife habitat has largely diminished and become fragmented, compromising the ability of many species to persist (Andrén 1994; Hanski and Ovaskainen 2000). This has occurred at a time when rapid directional climate change will necessitate range shifts for many organisms (Iverson and Prasad 1998, e.g.).

To reduce the degree of isolation caused by fragmented habitat, many conservation biologists have recommended maintaining landscape connectivity between core habitat areas, existing reserves, or subpopulations (e.g., (Noss 1983; Walker and Craighead 1997; Noss and Harris 1986)). Ideally, one wants to spend the limited available economic resources efficiently in order to obtain the conservation network of highest quality. An important network design decision concerns the selection of the criteria for evaluating the landscape connectivity of a given conservation network. Ecologists have developed models of landscape resistance (or conversely, landscape permeability), where the landscape is represented as a set of small parcels or pixels, each of which has a resistance value that gives the species-specific cost of moving through the particular landscape features. Such resistance models are then used to assess the landscape connectivity between important habitat patches in a given study area. For example, under the least-cost path (LCP) model, the connectivity between two designated habitat patches is measured as the length of the shortest resistance-weighted path between them (Singleton, Gaines, and Lehmkuhl 2002). Hence, a longer resistance-weighted path corresponds to a more difficult or less likely route for the animals to take.

Connectivity of the core areas needs also to be robust to changes in the landscape (McKelvey et al. 2011, e.g.). One way to achieve that is to design conservation plans that support multiple paths of low resistance-weighted length between core areas. To date the field of wildlife connectivity modeling, while acknowledging the necessity for robustness, has failed to quantify it or bring it into an optimization framework. For example, both Circuitscape (McRae and Beier 2007) and least cost corridors (LCC) (Beier, Majka, and Newell 2009) have been highlighted as approaches that can result in delineating corridors that support multple paths, but in both cases robustness is not calculated or targeted explicitely, but is a possible outcome of the way these two approaches work.

Underlying Graph To capture the conservation planning problem as a network design problem, we model the parcels of land as nodes in a graph, and edges are drawn between parcels that share boundaries. The resistance of each parcel is the corresponding node delay, and the conservation cost of a parcel is the corresponding node cost. For each pair of important habitat patches (i.e. existing conserved areas or subpopulations) that need to be connected, the corresponding pair of nodes in the graph are associated with the necessary connectivity requirement. Solving the combinatorial optimization problem designs a conservation strategy that minimizes the resistance-weighted length of connectivity paths (or corridors).

Relationship to Corridor Design Recently, wildlife corridor design was addressed using the so-called Connected Subgraph Problem (Conrad et al. 2007; Gomes, van Hoeve, and Sabharwal 2008; Dilkina and Gomes 2010; Conrad et al. 2012). In that work, the goal was to maximize the total utility of the set of bought parcels while ensuring that the parcels connected a designated set of reserves and that the total cost did not exceed a specified budget. Since the Connected Subgraph model only enforces connectivity without redundancy or reliability requirements, it can result in a conservation strategy where the loss of a single parcel disconnects the conservation network. Furthermore, the goal of maximizing the sum of utilities of all selected parcels could result in a strategy where the quality of the best path between two core areas is low, while the available budget is spent buying cheap parcels of good quality that do not directly contribute to connectivity in terms of the best path. The model presented in this work addresses these challenges by providing a way to obtain conservation strategies that lead to a *robust corridor design* and that optimizes directly the quality of the connecting paths between the core areas. If k is the minimum connectivity requirement enforced between the core areas, then even if parcels on $k-1$ of the paths are lost, the conservation network still remains connected.

3 Problem Definitions and Complexity

In this section we formally define the Minimum Delay Generalized Steiner Network problem, and analyze its computational complexity. In the following, given a graph $G = (V, E)$, we denote $G(V')$ as the subgraph of G induced by the vertices in $V' \subseteq V$. Moreover, we write $f(V')$ in short for $\sum_{v \in V'} f(v)$ for any real function $f : V \to \mathbb{R}+$ and $V' \subseteq V$. We start with reviewing the definition of the decision version of the node-weighted Generalized Steiner Network problem.

Generalized Steiner Network (GSN)

Given: an undirected graph $G = (V, E)$, a cost function on the nodes $c : V \to \mathbb{R}^+$, a budget value $B \geq 0$, and a set of terminal node pairs $P = \{(s_1, t_1), ..., (s_k, t_k)\} \subset V \times V$ with connectivity requirements $r_{s_i t_i}$

Find: a set of nodes $V' \subseteq V$ such that $c(V') \leq B$ and for each pair $(s,t) \in P$ the subgraph $G(V')$ has at least r_{st} edge (node) disjoint paths between s and t. We refer to these two variants as edge- and node-connectivity GSN Problems.

We extend this definition to include node delays and a delay-minimization objective. Here, given a delay function on the nodes $d : V \to \mathbb{R}^+$ the length of a path P is defined as $d(P) = \sum_{v \in P} d(v)$. Given a solution V' inducing subgraph $G' = G(V')$, for a terminal pair (s,t) with connectivity requirement r, suppose $P_1, \ldots, P_r$ are the r shortest paths between s and t in G', then the delay for pair (s,t) is defined as $d_{st}(G') = \sum_{i=1}^{r} d(P_i)$. Finally, the delay of a subgraph is $delay(G') = \sum_{i=1}^{k} d_{s_i t_i}(G')$.

Minimum Delay Generalized Steiner Network

Given: The specification of a GSN problem, and a delay function on the nodes $d : V \to \mathbb{R}^+$

Find: a set of nodes $V' \subseteq V$ satisfying the budget and connectivity constraints while minimizing the total delay $delay(G(V'))$.

Note that both edge or node connectivity variants of the GSN problem are NP-hard when there are two or more terminal pairs, and even on planar graphs, because they generalize the node-weighted Minimum Steiner Tree (MST) problem, which is NP-hard even on planar graphs (Borradaile, Klein, and Mathieu 2009).

Now consider the case when there is a single terminal pair (s,t). The node-connectivity version of GSN can be reduced to solving a Minimum Edge-Cost Flow problem on a directed graph with unit capacities. This problem is known to be polynomial time solvable in the case of unit capacities (Garey and Johnson 1979)[ND32]. Therefore, the node-connectivity GSN for one pair is in $\mathcal{P}$. In contrast, in Thm. 3.1 we show that the edge-connectivity GSN with one pair of terminals is strongly NP-complete.

Theorem 3.1. *The decision version of the edge-connectivity Generalized Steiner Network (ec-GSN) problem with a single pair of terminals is strongly NP-complete.*

Proof Sketch. The problem is clearly in NP. To show NP-completeness, we reduce the well-known NP-complete *Vertex Cover Problem* to the *ec-GSN* problem. Consider an instance of the Vertex Cover, i.e. a graph $G_c = (V_c, E_c)$ and an integer k. We build an instance of the *ec-GSN* decision problem, i.e. a graph $G = (V, E)$, a cost function c on nodes, a budget value B, and a single pair of terminals (s,t) with a connectivity requirement $m \geq 0$, as follows. For each node v_i in V_c, we introduce a node v_i in V as well as h nodes $d_i^1, d_i^2, ..., d_i^h$ where h is the degree of v_i in G_c. For each edge e_j in E_c, we introduce a node e_j in V. In addition, we augment V with a pair (s,t) of terminals. Moreover, we create an edge between node s and each node d_i^l, an edge between node t and any node e in V as well as an edge between v and e whenever e is incident to v in G_c. Furthermore, we define the costs of the nodes in V to be 0, except for the nodes v_i whose costs are set to 1. Finally, we set the budget value B to be k and the connectivity requirement m to be $|E_c|$. Figure 2 illustrates this construction.

Suppose G_c has a vertex cover set $D = \{v_{i_1}, ..., v_{i_k}\}$ of size k. Then, we can select $V' = \{V \setminus V_c\} \cup D$, whose total cost is exactly k and such that there are $|E_c|$ $s-t$ edge-disjoint paths in $G(V')$. Conversely, assume there exists a set of nodes V' of total cost at most k such that the number of $s-t$ edge-disjoint paths in $G(V')$ is $|E_c|$. Then, it is easy to show that $D = V' \cap V_c$ is a vertex cover for graph G_c and has size at most k. $\square$

Now we are ready to establish the complexity of Minimum Delay GSN problem.

Theorem 3.2. *The Minimum Delay Generalized Steiner Network problem is NP-hard for node-connectivity and edge-connectivity requirements, even for a single pair of terminals.*

Proof. To see that this problem is in NP, suppose we are given a set of nodes V', finding the delay for a given pair

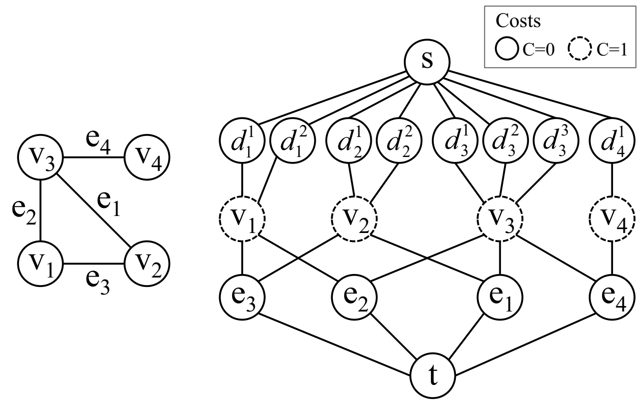

Figure 2: Reduction from the Vertex Cover Problem. Left: Instance of the Vertex Cover. Right: Corresponding instance of the node-weighted edge-con nectivity GSN problem.

(s,t) in $G(V')$ consists of finding a mincost flow of r_{st} units in the corresponding unit-capacity network, which can be done in polynomial time. Moreover, the fact that the Minimum Delay GSN problem has the GSN problem embedded completes the proof for all cases but the variant concerning node-connectivity and a single pair of terminals nodes. This variant is still hard as it is a generalization of the NP-hard problem *Shortest Weight-Constrained Path* (Garey and Johnson 1979)(ND30) where given a graph with costs and weights on nodes, the goal is to find a single path between s and t such that the cost of the path is within a given budget and the length (in terms of weights) is minimized. $\square$

Finally, we present a slight generalization of the problem to explicitly capture the problem of conservation planning of multiple species, where each species i experiences species-specific resistance or delay $d^i : V \to \mathbb{R}^+$ at each parcel or node of the landscape. Additionally, as some parts of the landscape (network) might correspond to complete barriers to movement for a particular species i, we restrict the paths for that species to a given subgraph $G^i = (V^i \subseteq V, E^i \subseteq E)$.

4 MIP formulation

To solve the MinDelay-GSN problem, we exploit the fact that finding a given number r_j of node-disjoint paths between a pair of vertices (s_j, t_j) in a network is equivalent to finding a flow of size r_j in the network augmented with unit edge capacities and constraining the flow through each node to be at most one. Based on this correspondence, we use a flow-based MILP formulation. For this formulation, we transform the given undirected graph G to a directed graph $\hat{G}$ where each undirected edge is replaced with two directed edges. Let $\delta^-(v)$ denote the incoming edges for vertex v, and $\delta^+(v)$ denote the outgoing edges for v in $\hat{G}$. We also compute delays on edges instead of on nodes using the formula $d(e = (u,v)) = [d(u) + d(v)]/2$, which preserves the delay on each path plus a constant contribution from the terminals $[d(s) + d(t)]/2$ that can be ignored for optimality

purposes.
The variables used in our formulation are the following:

- x_v: a binary variable which is 1 when node $v \in V$ is bought, and 0 otherwise;
- f_{pe}: a continuous variable indicating the flow of commodity p on edge e, i.e. whether edge e is chosen to be on a path for the terminal pair p.

The overall *flow-based MILP* encoding of the problem is:

$$\min \quad \sum_{i\in S}\sum_{p\in P^i} flowcost_p^i \tag{1}$$

$$\text{s.t. } \Pi_p \qquad \forall i \in S, \forall p \in P^i \tag{2}$$

$$\sum_{v\in V} c(v)x_v \leq B \tag{3}$$

$$x_v \in \{0,1\} \qquad \forall v \in V \tag{4}$$

For a particular pair $p = (s,t)$ for species i, Π_p stands for the flow-related set of constraints, enforcing the connectivity requirement, and $flowcost_p^i$ is an auxiliary variable that stands for the delay incurred by the paths connecting pair p of species i. Both are defined below:

$$\Pi_p = \sum_{e\in\delta^-(v)} f_{pe} \leq x_v \qquad \forall v \in V \setminus \{s,t\} \tag{5}$$

$$\sum_{e\in\delta^+(s)} f_{pe} = \sum_{e\in\delta^-(t)} f_{pe} \tag{6}$$

$$\sum_{e\in\delta^-(v)} f_{pe} = \sum_{e\in\delta^+(v)} f_{pe} \qquad \forall v \in V \setminus \{s,t\} \tag{7}$$

$$\sum_{e\in\delta^+(s)} f_{pe} = r_{st} \tag{8}$$

$$flowcost_p^i = \sum_{e\in E} d^i(e) * f_{pe} \tag{9}$$

$$f_{pe} = 0 \qquad \forall e \in E\backslash E^i \tag{10}$$

$$0 \leq f_{pe} \leq 1 \qquad \forall e \in E^i \tag{11}$$

Constraint 5 requires that a node needs to be purchased in order to receive a non-zero incoming flow. It also enforces that paths need to be node-disjoint, by allowing at most one incoming edge. Note that to encode the edge-disjoint variant, one only needs to modify this constraint to include a Big-M constant multiplier for x_v.Constraint 9 captures the total delay for pair $p = (s,t)$ as the delay along all edges carrying flow. Finally, Constraint 10 confines species i to the subgraph G^i.

5 Local Search Approach

In order to scale to very complex large-scale real-world networks, we also introduce a local search approach whose neighborhood involves solving the proposed MIP encoding for small subproblems.

As described earlier, this network design problem combines an intricate combinatorial structure (to enforce budget-constrained connectivity requirements) with a complex path-based optimization component. On one hand, for constraint satisfaction problems, specialized global constraints have been proposed to maintain connectivity when generating a network (Brown et al. 2005; Prosser and Unsworth 2006), but they typically suffer from a poor propagation. On the other, approximation algorithms (Williamson et al. 1995; Goemans et al. 1994, e.g.) and local search strategies (Cancela, Robledo, and Rubino 2003; Verhoeven, Severens, and Aarts 1996) have been proposed to address optimization in network design problems. Building upon (Cancela, Robledo, and Rubino 2003), we maintain the connectivity requirements by construction at each step of the search. In particular, we use the following notion of *key-path* to define a suitable structure for the neighborhood.

Key-paths Given a feasible solution $\mathcal{G}$ of the MinDelay-GSN problem, a key-path is a path in $\mathcal{G}$ such that all intermediate nodes are Steiner nodes (i.e. non-terminal nodes) of degree 2 in $\mathcal{G}$, and whose end nodes are either terminal nodes or Steiner nodes of degree at least 3.

Neighborhood Given a feasible solution $\mathcal{G}$ of the MinDelay-GSN problem and a key-path p in $\mathcal{G}$, we define a neighbor solution of $\mathcal{G}$ as $\hat{\mathcal{G}} = (\mathcal{G} \setminus p) \cup \hat{p}$, where $\hat{p}$ is a shortest path (SP) or a set of budget-constrained shortest paths (MIP) connecting the end points of p and maintaining the feasibility in the new solution $\hat{\mathcal{G}}$.

Given a key-path p, we denote by $\mathcal{V}_p(\mathcal{G})$ the set of terminal pairs for which there is at least one path going through p in $\mathcal{G}$. To maintain the node-disjoint connectivity requirements, when substituting a key-path p for a path $\hat{p}$, the nodes of $\hat{p}$ should not include any node used by the paths in $\mathcal{V}_p(\mathcal{G})$, except for the ones in p itself. When adapting (Cancela, Robledo, and Rubino 2003) to the MinDelay-GSN problem, $\hat{p}$ corresponds to the delay-weighted shortest path (SP) between the end points of p. When p serves the connectivity of more than one terminal pair, we consider the shortest path with respect to the joint node delays, obtained by summing the delay of that node incurred for each of the pairs in $\mathcal{V}_p(\mathcal{G})$. Yet, in our setting, the delay-shortest path is likely to violate the budget constraint and be rejected as a possible local search move. Therefore, in addition, we propose a more involved neighborhood that solves a budget-constrained shortest paths problem for each key-path of $\mathcal{G}$ using a MIP formulation. This allows us to find $\hat{p}$ within the available remaining budget and to consider solutions with potentially multiple paths when $|\mathcal{V}_p(\mathcal{G})| > 1$. We embed this MIP neighborhood (resp. SP neighborhood) in a hill-climbing procedure (HC-MIP) (resp. HC-SP).

Finally, this approach requires a feasible initial solution. Therefore, we propose a randomized greedy algorithm to provide initial solutions, which iterates over a randomized sequence of all terminal pairs and for each subsequent pair satisfies the connectivity requirement while minimizing the additional cost incurred. In addition, we enhance this greedy solution by performing improving cost-shortest path moves, as described in (Cancela, Robledo, and Rubino 2003). Note that this initialization procedure might fail to provide a solution within the budget.

The proposed local searches can be classified as Large Neighborhood Searches (LNS) (Pisinger and Ropke 2010). In particular, the method HC-MIP is reminiscent of the contraint-based LNS proposed in (Shaw 1998) for vehicle routing problems. In our approach, following the LNS terminology, an initial solution is gradually improved by *destroying* a key-path of the solution, and *repairing* the solution using the shortest path (HC-SP) or a budget-constrained set of shortest paths (HC-MIP).

6 Experiments

All experiments were run on a Linux (version 2.6.18) cluster where each node has an Intel Xeon Processor X5670, with dual-CPU, hex-core @2.93GHz, 12M Cache, 48GB RAM. Both MILP and HC-MIP approaches used IBM ILOG CPLEX version 12 (ILOG, SA 2011). We tuned the solver using the IBM ILOG Cplex internal tuning method on 1,000 models from the synthetic benchmark described below for both approaches (MILP and HC-MIP). Nonetheless, no non-default parameter values appeared predominantly after tuning, and we observed no significant improvement in runtime between *untuned* and *tuned* parameter settings. For the instances concerning the case study in Montana, instead of tuning on a benchmark of instances, we use the IBM ILOG Cplex tuning method (with a cutoff time of 900 seconds) to tune the parameters of each model individually. More advanced automated algorithm configuration methods such as ParamILS (Hutter et al. 2009) could potentially improve the performance of the CPLEX solver and will be addressed in future work.

Synthetic Instances In order to evaluate the hardness profile of the MinDelay-GSN problem, we used a synthetic problem generator, proposed in (Dilkina et al. 2013), that captures the characteristics of the motivating landscape connectivity conservation planning problem. A detailed description of the generator, as well as the source code and the synthetic dataset used in this work can be found online[1].

We analyzed the performance on 300 random instances (30x30 grid graphs with 4 species each), and report median results. We vary the number of species considered by restricting each instance to only 1, 2, 3 or 4 species. A connectivity requirement of k *node-disjoint* paths is placed between the two habitat cores of each species, and it is the same for all species. For each problem instance restricted to a given number of species and a connectivity requirement, we compute the minimum budget necessary to satisfy the requirements, and then vary the budget in increments of 10% of the minimum budget. The top panel of Fig. 3 shows the scaling behavior of our MILP approach. One can clearly observe an easy-hard-easy pattern in median runtime with respect to the budget, with the most difficult instances being slightly above the minimum budget required to meet the connectivity constraints. Note that similar hardness profiles were obtained in the case of corridor design (Conrad et al. 2012). In addition, increasing the number of species has a sharply negative effect on the runtime. The right panel of the figure presents results for our local search algorithms in terms of the median optimality gap (over 300 instances) between the best solution found by the local search over 10 runs, x, and the MILP global minimum, x^*, using $\frac{(x-x^*)100\%}{x^*}$. The results show the advantage of the complex move structure used by HC-MIP, which results in better solutions than the HC-SP and Greedy. The quality of the solutions found by HC-MIP is between 7.10% and 16.67% of optimal, as compared to between 7.23% and 21.85% for HC-SP. In terms of running times, HC-SP is nevertheless extremely fast, taking only a few tens of seconds, and therefore is highly scalable, while HC-MIP and MILP take up to 220 seconds and 1119 seconds respectively.

[1]http://www.cis.cornell.edu/ics/Datasets

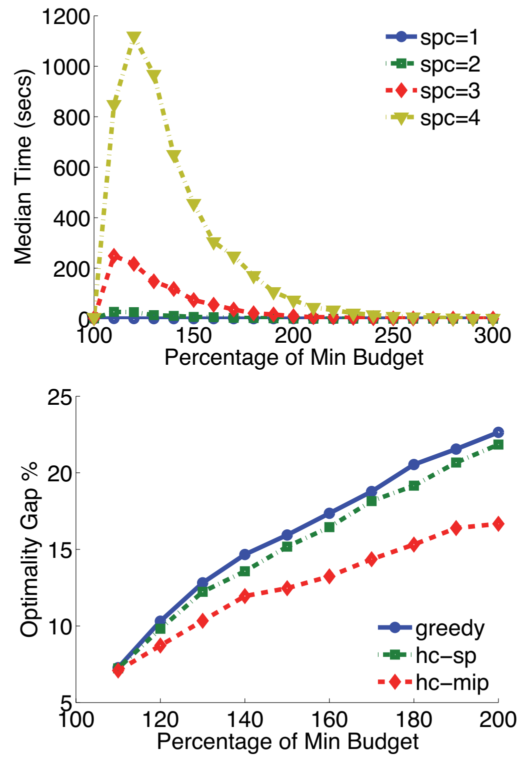

Figure 3: Typical-case analysis for the *minimum delay generalized Steiner Network* problem on synthetically generated instances with one pair per species (30x30 grid graphs; 300 instances per data point). (Top) Scaling behavior with increasing number of species/pairs for connectivity of 4. Plot shows easy-hard-easy pattern in median runtimes with respect to the budget. (Bottom) Median optimality gap of local search for 4 species and connectivity of 4.

Conservation Planning for Western Montana In Montana, the wolverine (Gulo gulo) and the Canada lynx (Lynx canadensis) are classified as species of concern, with the lynx federally listed as a threatened species under the Endangered Species Act (Federal Register 65 FR 16053 16086) and the wolverine recommended for listing as Threatened (Federal Register 77 FR 69993 70060). Both species are found at low densities in patches of optimal habitat throughout the Northern U.S. Rocky Mountains (Squires et al. 2007; 2013). For populations to persist at low densities and not suffer negative effects of inbreeding or other stochastic demographic processes, adequate connectivity among patches must exist. Several high profile efforts from both the private and public sector have attempted to preserve and enhance connectivity throughout this region for the benefit of wildlife. One of the most critical areas for connectivity is between the Northern Continental Divide Ecosystem and the Greater Yellowstone Area as show in Fig. 1. Preserving connectivity between these two major areas would be beneficial for both species. We use the dataset developed in (Lai et al. 2011), available online [2].

First, we consider conservation planning scenarios only for the wolverine. There are several habitat areas with persistent spring snow that support wolverine populations, but they are scattered across the state of Montana. We consider a planning scenario with 15 pairs. We run CPLEX on the MILP formulation with a 10 hour cpu time cutoff. No feasible solutions are found except for the 3 largest budgets (140M, 145M and 150M) , even when setting the Cplex parameter of the mip emphasis to focus on finding incumbents (CPLX_PARAM_MIPEMPHASIS=1). However, we record the tightest lower bound on the objective function derived by CPLEX within this time limit, and use this value to obtain an upper bound on the optimality gap of the local search solutions. The quality of the best solutions obtained by local search over 20 runs is presented in the top panel of Fig. 4. HC-SP takes at most 270 secs per run achieving 25-45% optimality gap, while HC-MIP takes up to 4200 secs per run and has 22-40% optimality gap across all budgets. Again, note that the optimality gaps reported are upper bounds due to the lack of optimal solutions from the MILP model.

Next, we turn to multispecies conservation planning optimizing jointly for the wolverine and the lynx and considering only the connectivity between the two major areas of the Northern Continental Divide Ecosystem and the Greater Yellowstone Area. It should be noted that while the two species inhabit similar habitats, their needs are not identical, evident from the difference in their species-specific resistance values across the landscape (see Fig.1). Results when planning for both species under different budgets and connectivity requirements are shown in the bottom panel of Fig 4. For two species and a total of two pairs of nodes to connect, we are able to solve the planning problems to optimality with our MILP formulation. We compare solutions across different connectivity levels at a fixed budget using the average delay per path in the obtained solution. As we demand more disjoint paths to ensure more robust connectivity, the average quality of the paths that can be conserved with the same budget deteriorates. This systematic evaluation of the tradeoff between achieving higher level of robustness in terms of number of disjoint paths and the average delay per path can be a very informative tool for conservation planners.

[2]http://www.cis.cornell.edu/ics/Datasets

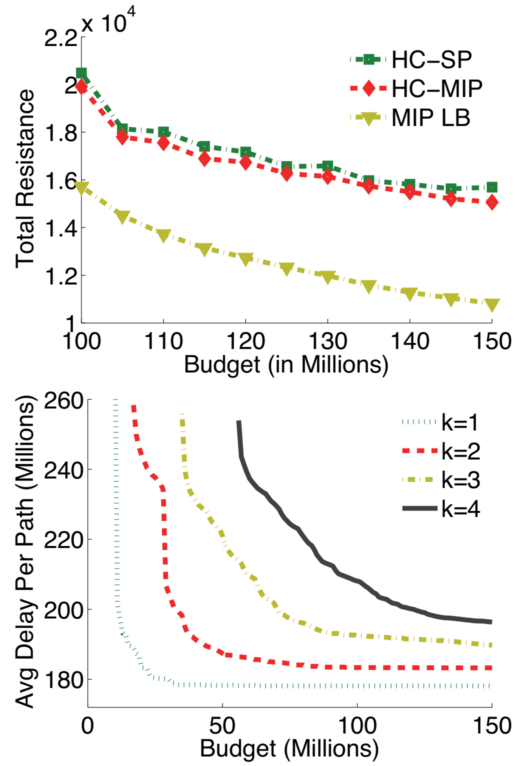

Figure 4: Western Montana case study: (Top) Results for wolverines with connectivity requirement of 4 and 15 habitat pairs comparing HC-MIP and HC-SP to the lower bound obtained from the MILP. (Bottom) Results for two species (lynx and wolverines) in terms of the average path delay over different connectivity requirements k and budgets.

7 Conclusion

In this paper we introduce the *minimum delay generalized Steiner network* problem, to address the problem of designing conservation strategies for robust landscape connectivity for endangered species. We formally show that the problem is NP-Hard and provide a mixed integer programming problem formulation, as well as local search approaches. We provide a detailed typical case analysis of the solution approaches using a synthetic problem generator. We also apply our approach to solve a large-scale real-world network design problem concerning the conservation of wolverine and lynx populations in the U.S. Rocky Mountains (Montana). The results reveal that the MILP approach is capable of providing optimal solutions for large-scale planning problems when only a few pairs of nodes need to be connected. In addition, our local search approaches are capable of addressing instances with a large number of pairs that are computationally infeasible for the MILP. This work opens up several follow up questions, in particular investigating new algorithmic approaches that will allow us to further scale up our results. We are also interested in investigating approximation techniques and in particular approaches that exploit planarity.

Acknowledgments

We are thankful to the reviewers for their constructive feedback. This work was supported by the National Science Foundation (NSF Expeditions in Computing award for Computational Sustainability, grant 0832782) and the USDA Forest Service, Rocky Mountain Research Station (grant 10-JV-11221635-24). All experiments were run on an infrastructure supported by the NSF Computing research infrastructure for Computational Sustainability grant (grant 1059284).

References

Andrén, H. 1994. Effects of habitat fragmentation on birds and mammals in landscapes with different proportions of suitable habitat: a review. *Oikos* 71(3):355–366.

Beier, P.; Majka, D. R.; and Newell, S. L. 2009. Uncertainty analysis of least-cost modeling for designing wildlife linkages. *Ecological Applications* 19(8):2067–2077.

Borradaile, G.; Klein, P.; and Mathieu, C. 2009. An o (n log n) approximation scheme for steiner tree in planar graphs. *ACM Transactions on Algorithms (TALG)* 5(3):31.

Brown, K.; Prosser, P.; Beck, J.; and Wu, C. 2005. Exploring the use of constraint programming for enforcing connectivity during graph generation. In *5th workshop on modelling and solving problems with constraints (at IJCAI05)*.

Cancela, H.; Robledo, F.; and Rubino, G. 2003. Network design with node connectivity constraints. In *IFIP/ACM Latin America conference on Towards a Latin American agenda for network research (LANC)*, 13–20.

Conrad, J.; Gomes, C. P.; van Hoeve, W.-J.; Sabharwal, A.; and Suter, J. 2007. Connections in networks: Hardness of feasibility versus optimality. In *CPAIOR: International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, 16–28.

Conrad, J.; Gomes, C. P.; van Hoeve, W.-J.; Sabharwal, A.; and Suter, J. 2012. Wildlife corridors as a connected subgraph problem. *J. of Environmental Economics and Management* 63(1):1 – 18.

Dilkina, B., and Gomes, C. P. 2010. Solving connected subgraph problems in wildlife conservation. In *CPAIOR: International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, 102–116.

Dilkina, B.; Gomes, C. P.; Lai, K.; Le Bras, R.; McKelvey, K. S.; Sabharwal, A.; Schwartz, M. K.; Suter, J.; and Xue, Y. 2013. Large conservation landscapesynthetic and real-world datasets. In *AAAI*.

Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York, NY, USA: W. H. Freeman & Co.

Goemans, M.; Goldberg, A.; Plotkin, S.; Shmoys, D.; Tardos, E.; and Williamson, D. 1994. Improved approximation algorithms for network design problems. In *SODA: ACM-SIAM Symposium on Discrete algorithms*, 223–232.

Gomes, C. P.; van Hoeve, W.-J.; and Sabharwal, A. 2008. Connections in networks: A hybrid approach. In *CPAIOR: International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, 303–307.

Gomes, C. 2009. Computational sustainability: Computational methods for a sustainable environment, economy, and society. *The Bridge, National Academy of Engineering* 39(4).

Hanski, I., and Ovaskainen, O. 2000. The metapopulation capacity of a fragmented landscape. *Nature* 404(6779):755–758.

Hutter, F.; Hoos, H. H.; Leyton-Brown, K.; and Stutzle, T. 2009. ParamILS: an automatic algorithm configuration framework. *Journal of Artificial Intelligence Research* 36:267–306.

ILOG, SA. 2011. CPLEX 12.3 Reference Manual.

Iverson, L. R., and Prasad, A. M. 1998. Predicting abundance of 80 tree species following climate change in the eastern united states. *Ecological Monographs* 68(4):465–485.

Kerivin, H., and Mahjoub, A. R. 2005. Design of Survivable Networks: A survey. *Networks* 46(1):1–21.

Lai, K. J.; Gomes, C. P.; Schwartz, M. K.; McKelvey, K. S.; Calkin, D. E.; and Montgomery, C. A. 2011. The steiner multigraph problem: Wildlife corridor design for multiple species. In *AAAI*.

McKelvey, K. S.; Copeland, J. P.; Schwartz, M. K.; Littell, J. S.; Aubry, K. B.; et al. 2011. Climate change predicted to shift wolverine distributions, connectivity, and dispersal corridors. *Ecological Applications* 21(8):2882–2897.

McRae, B. H., and Beier, P. 2007. Circuit theory predicts gene flow in plant and animal populations. *Proceedings of the National Academy of Sciences* 104(50):19885–19890.

Noss, R. F., and Harris, L. D. 1986. Nodes, networks, and mums: Preserving diversity at all scales. *Environmental Management* 10:299–309.

Noss, R. F. 1983. A regional landscape approach to maintain diversity. *BioScience* 33(11):700–706.

Phillips, S. J.; Williams, P.; Midgley, G.; and Archer, A. 2008. Optimizing dispersal corridors for the cape proteaceae using network flow. *Ecological Applications* 18(5):1200–1211.

Pisinger, D., and Ropke, S. 2010. Large neighborhood search. In *Handbook of metaheuristics*. Springer. 399–419.

Prosser, P., and Unsworth, C. 2006. A connectivity constraint using bridges. *Frontiers in Artificial Intelligence and Applications* 141:707.

Shaw, P. 1998. Using constraint programming and local search methods to solve vehicle routing problems. In *CP*, 417–431.

Singleton, P. H.; Gaines, W. L.; and Lehmkuhl, J. F. 2002. Landscape permeability for large carnivores in washington: a geographic information system weighted-distance and least-cost corridor assessment. *Res. Pap. PNW-RP-549: USDA, Forest Service, Pacific Northwest Research Station*.

Squires, J. R.; Copeland, J. P.; Ulizio, T. J.; Schwartz, M. K.; and Ruggiero, L. F. 2007. Sources and patterns of wolverine mortality in western montana. *The Journal of Wildlife Management* 71(7):2213–2220.

Squires, J. R.; DeCesare, N. J.; Olson, L. E.; Kolbe, J. A.; Hebblewhite, M.; and Parks, S. A. 2013. Combining resource selection and movement behavior to predict corridors for canada lynx at their southern range periphery. *Biological Conservation* 157:187–195.

Verhoeven, M.; Severens, M.; and Aarts, E. 1996. Local search for steiner trees in graphs. *Modern Heuristics Search Methods* 117–129.

Walker, R., and Craighead, L. 1997. Analyzing wildlife movement corridors in montana using gis. In *ESRI Users Conference*, 1–19. ESRI.

Williamson, D.; Goemans, M.; Mihail, M.; and Vazirani, V. 1995. A primal-dual approximation algorithm for generalized steiner network problems. *Combinatorica* 15(3):435–454.