The Steiner Multigraph Problem: Wildlife Corridor Design for Multiple Species

Katherine J. Lai
Carla P. Gomes
Computer Science Department
Cornell University
Ithaca, NY 14850
klai, gomes@cs.cornell.edu

Michael K. Schwartz
Kevin S. McKelvey
David E. Calkin
USDA Forest Service
Rocky Mountain Research Station
800 East Beckwith, Missoula, Montana 59801
mkschwartz, kmckelvey, decalkin@fs.fed.us

Claire A. Montgomery
Department of Forest Resources
Oregon State University
Corvallis, Oregon 97331
claire.montgomery@oregonstate.edu

Abstract

The conservation of wildlife corridors between existing habitat preserves is important for combating the effects of habitat loss and fragmentation facing species of concern. We introduce the Steiner Multigraph Problem to model the problem of minimum-cost wildlife corridor design for multiple species with different landscape requirements. This problem can also model other analogous settings in wireless and social networks. As a generalization of Steiner forest, the goal is to find a minimum-cost subgraph that connects multiple sets of terminals. In contrast to Steiner forest, each set of terminals can only be connected via a subset of the nodes. Generalizing Steiner forest in this way makes the problem NP-hard even when restricted to two pairs of terminals. However, we show that if the node subsets have a nested structure, the problem admits a fixed-parameter tractable algorithm in the number of terminals. We successfully test exact and heuristic solution approaches on a wildlife corridor instance for wolverines and lynx in western Montana, showing that though the problem is computationally hard, heuristics perform well, and provably optimal solutions can still be obtained.

1 Introduction

One of the major problems in computational sustainability is that of preserving our world's diminishing biodiversity (Gomes 2009). The plight of many endangered species is largely due to the encroachment of human development activities that diminish and fragment their existing habitat. While conserving contiguous reserves is essential for setting aside core habitat areas, connectivity must be maintained between the smaller communities that form in these reserves. Habitat fragmentation can lead to small isolated populations, which in turn can lead to negative demographic and genetic consequences (Gilpin and Soule 1986; Fagan and Holmes 2006). One method to mitigate this problem is to set aside so-called wildlife corridors, or swaths of preserved land that connect important patches of habitat for the endangered species. Wildlife corridors have been used successfully in the past (Shepherd and Whittington 2006; Haddad et al. 2003). An important computational problem that arises is how to best spend limited funds on these conservation efforts. It is also important for biodiversity and economic reasons to plan conservation efforts while considering the diverse needs of the different species in a common geographical area (Beier, Majka, and Spencer 2008).

In Montana, both the wolverine (*Gulo gulo*) and the Canada lynx (*Lynx canadensis*) (see Figure 1) are classified as species of concern, with the lynx federally listed as a threatened species under the Endangered Species Act. Both these species suffer from habitat fragmentation and inhabit the Northern Continental Divide Ecosystem (NCDE) as well as the Greater Yellowstone Area (GYA) as shown in Figure 1. Preserving a wildlife corridor connecting these two major areas, as well as other areas of potential habitat, would be beneficial for both species. However, it should be noted that while the two species inhabit similar habitats, their needs are not identical.

While the habits of the few and elusive Montana wolverines are not fully known, they commonly occur in areas with persistent spring snow cover (Copeland et al. 2010). These areas are essential for reproduction, as the female wolverines make their dens in the snow itself. In the study of landscape genetics, the dispersal of wolverines has also been found to be highly correlated with areas with persistent spring snow cover (Schwartz et al. 2009). Dispersal and exploration of individual wolverines largely occurs when juveniles travel through varied suboptimal habitat before establishing a home range, possibly covering hundreds of kilometers in the process (Gardner, Ballard, and Jessup 1986; Inman et al. 2004; Moriarty et al. 2009).

Canada lynx are also known for long distance movements especially during the low of a hare population cycle (Schwartz et al. 2002), yet due to their dependency on snowshoe hares for the majority of their diet they have an ecologically limited home range when not dispersing (Squires and Ruggiero 2007). They are also known to prefer gentler slopes (less than 15 degrees), and as hunters, they are less successful at finding food in open areas and thus can only travel through denser ground cover. Thus to design a wildlife corridor to connect the NCDE and the GYA for the wolverine and lynx, we must take into account the stricter requirements of the lynx as shown in Figure 1.

Researchers (Conrad et al. 2007; Gomes, van Hoeve, and Sabharwal 2008; Dilkina and Gomes 2010) have modeled the wildlife corridor design problem as a network design problem with strict connectivity constraints between a set of

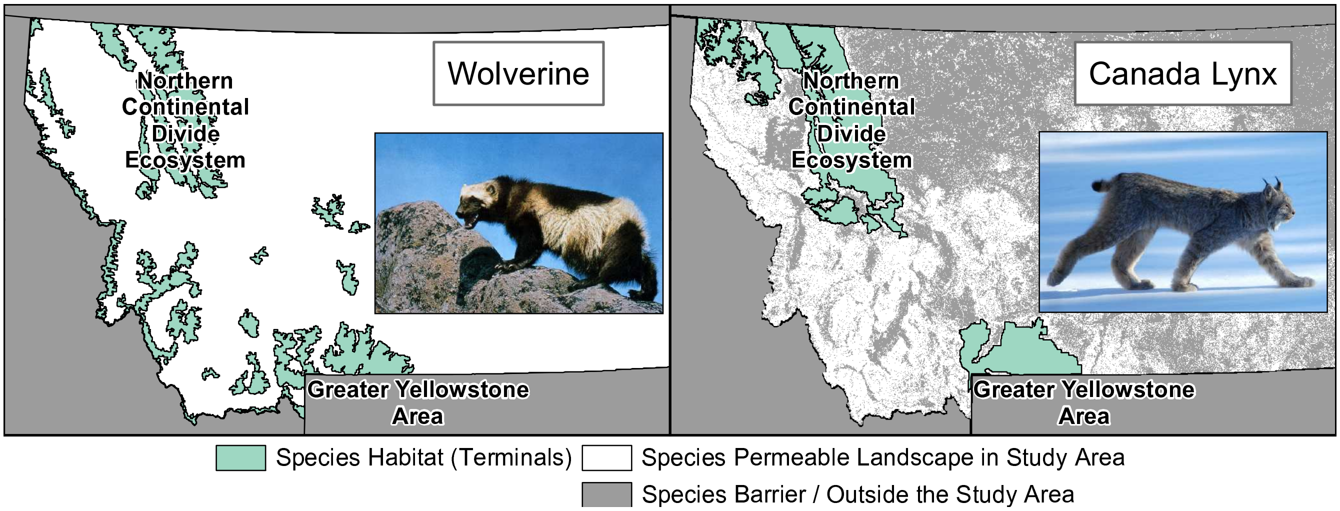


Figure 1: *Both the wolverines (left, photo from the National Park Service) and the Canada lynx (right, photo taken by Keith Williams, 2010) are listed as species of concern in Montana. The lynx are federally listed as a threatened species under the Endangered Species Act. The areas of wolverine terminals are the areas with persistent spring snow cover (acquired using MODIS snow cover data). The lynx terminals shown are the areas in the intersection of Montana and the Critical Habitat designated by the US Fish and Wildlife in 2000. As shown above, the lynx have more restrictions with respect to land cover. Not pictured: barriers common to both species, e.g. bodies of water and urban areas.*

locations called *terminals* which represent the core habitat areas to be connected. The land parcels under consideration are modeled as nodes in a network with costs equal to their purchase price and utility values measuring their habitat suitability for a particular species. Edges are drawn between any two parcels that share a border. A feasible corridor is then a subset of parcels (nodes) that connect any two habitat areas (terminals) for a total cost that satisfies a given budget amount. To make the most of the available budget, they define the optimal solution to be the wildlife corridor which contains the maximum total utility possible.

Unfortunately, we cannot use the model posed by (Conrad et al. 2007) to solve the wildlife corridor design problem simultaneously for the wolverines and the lynx. While we can certainly use it to solve it once for each species and then hope to combine the solution for a single wildlife corridor, this may not find the most economically efficient solution. The crux of the problem lies in the fact that some areas that are practically barriers for the lynx (such as steep slopes) should not be used to connect habitat for the lynx, but they should still be considered for connecting wolverine habitat. In this paper, we introduce the *Steiner Multigraph Problem* (SMP) to address the limitation of previous corridor models by taking into account the set of land parcels that can be used to connect each species' set of terminals. This paper studies the cost-minimization problem as a first step towards studying the analogous multiple-species wildlife corridor problem with budget constraints and utility values.

SMP can also be used to model other settings such as wireless sensor networks and social networks. For example, graph nodes can represent sensors that can communicate if they have been placed close enough (Salhieh, Weinmann, and Kochhal 2001). Finding a minimum-cost subgraph in a graph of potential sensor locations corresponds to finding the cheapest backbone for multicast communication in the network. As an SMP instance, we can now model a network with different devices that may only have the required hardware for certain communication capabilities

In social networks, nodes represent individuals, and edges represent relationships between them. Finding the smallest connected subgraph in a social network can give information on the minimum number of people who need to have gossiped for news to travel from one end of the network to the other (Faloutsos, McCurley, and Tomkins 2004). As one might imagine, a social network is actually composed of smaller communities that are connected and overlap (Ahn, Bagrow, and Lehmann 2010). SMP can model such different communities by taking into account the fact that different types of information may freely travel within different subgraphs of the overall social network.

Our Contributions

In Section 2, we formally define the Steiner Multigraph Problem. In Section 3, we characterize its computational complexity under different restrictions. As a generalization of Steiner forest, it follows directly that SMP is also NP-hard. For a constant number of terminals, the Steiner forest problem is also fixed-parameter tractable (FPT) and admits a polynomial time algorithm. However, we show that SMP is NP-hard even for two sets of terminals with two terminals each. In a positive result, we show that when the subsets V_i have a nested structure, SMP admits an optimal solution that does not contain a cycle. This special case of SMP is in FPT with respect to the number of terminals. In Sections 4 and 5, we present and test both exact and heuristic approaches to solving SMP on a set of synthetic instances and finally on the wildlife corridor problem in Montana. On the synthetic instances, the heuristics perform well, coming within 10% of optimal for most of the tested instances. For the wildlife corridor instance in Montana, we found that the inherent

$T_1, T_2 = \{t_1, t_2\}$
$V_1 = \{t_1, u, t_2\}$
$V_2 = \{t_1, v, t_2\}$

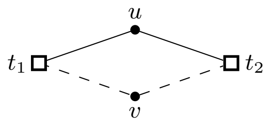


Figure 2: *In this example, a feasible solution requires all the nodes and edges in the graph, resulting in a cycle.*

structure of the land parcel costs and the additional constraints placed on the problem by the lynx made the problem tractable, allowing us to obtain provably optimal solutions.

2 Problem Definition

We will focus on the node-weighted formulation of SMP because our motivating application in wildlife corridor design is modeled more naturally using node weights. Though we will not discuss it in detail, our hardness and algorithmic results hold for the edge-weighted case as well.

STEINER MULTIGRAPH PROBLEM (SMP)

Given: an undirected graph $G = (V, E)$, an index set P, node sets $V_i \subseteq V$ for all $i \in P$, sets of terminals $T_i \subseteq V_i$ for all $i \in P$, and costs c_v for all $v \in V$.

Find: a node set $W \subseteq V$ such that the induced subgraph $G(W \cap V_i)$ connects T_i for each $i \in P$, and the total cost $\sum_{v \in W} c_v$ is minimized.

The name of this problem is inspired by the network design problems that it generalizes, most notably the minimum-weight *Steiner tree* and *Steiner forest* problems. A Steiner tree instance is an instance of SMP where $|P| = 1$; there is only one species to be accommodated by the wildlife corridor. A Steiner forest instance is an instance of SMP where all V_i are the same; there can be multiple species, and all land is equally permeable for all of them. Intuitively, a solution to SMP can be thought of as the union of feasible Steiner trees that connect the sets of terminals T_i in their respective induced subgraphs $G(V_i)$. As a generalization of minimum-weight Steiner forest, however, an optimal solution for SMP may contain cycles. An example of such a case is shown in Figure 2, where there are two (non-disjoint) pairs of terminals. Each pair is limited to a different path connecting the two terminals. As a result, the only feasible solution is the entire graph which contains a cycle.

In the rest of this paper, we will consider a few different special cases of the problem. In particular, given that the wolverine can traverse a superset of the landscape that the lynx can, we will also define and discuss the *Laminar Steiner Multigraph Problem* (LSMP) where the sets V_i are required to form a laminar family. By definition, a family of sets is *laminar* if for every two member sets V_i and V_j, it is the case that $V_i \cap V_j = \emptyset$, $V_i \subseteq V_j$, or $V_j \subseteq V_i$.

3 Hardness Results

3.1 General Case

As a generalization of Steiner forest, SMP is NP-hard and is at least as hard to approximate as Steiner forest. For the node-weighted case, this means that the best approximation factor possible is $\Omega(\log |T|)$ for general graphs.

3.2 Two Sets of Terminals

If the node-weighted Steiner forest problem is restricted to two pairs of terminals, it is solvable in polynomial time. This is because the solution can only be either (a) a single Steiner tree connecting all the terminals, or (b) a forest consisting of the shortest paths for each of the two pairs of terminals. In contrast, SMP is NP-hard even under the same restriction.

Theorem 1. *SMP is NP-hard for instances on two pairs of terminals, i.e. where* $|P| = 2$, $|T_1| = 2$, *and* $|T_2| = 2$.

This result is obtained via a reduction from the NP-complete problem 3SAT. Our construction is similar to the reduction showing that the integer multicommodity flow problem is NP-hard for two commodities (Even, Itai, and Shamir 1976). For the full proof, see (Lai et al. 2011).

Many graph problems like Steiner forest are easier on planar graphs, not only in practice but also in terms of theoretical guarantees. Unfortunately, SMP on two terminal pairs is still NP-hard on planar graphs. For this result, we reuse the concepts of the nonplanar reduction and reduce from the restricted planar 3SAT, another NP-complete problem (Lichtenstein 1982). Again, see (Lai et al. 2011) for a full proof.

Theorem 2. *Theorem 1 also applies to planar graphs.*

3.3 Laminar Subgraphs

In the 3SAT reductions, the two node sets V_1 and V_2 overlap in a convoluted manner. However, if we restrict one set to be contained in the other, the problem becomes fixed-parameter tractable. As described earlier, this special case comes up in the problem of designing a wildlife corridor for wolverines and lynx. Social networks and biological networks are also often argued to have and interpreted to have a hierarchical or modular structure in many cases (Newman and Girvan 2004; Simon 1962). We refer the reader to the full version of the paper for the proof of the following result (Lai et al. 2011).

Theorem 3. *LSMP admits some optimal solution with no cycles and is therefore FPT with respect to* $|\cup_{i \in P} T_i|$.

4 Exact and Heuristic Solutions

We describe several algorithms to solve SMP. To solve the problem exactly, we can employ a mixed integer program (MIP) formulation. If the instance is laminar, then it admits a dynamic programming (DP) algorithm. We also describe faster approaches that may not solve the problem optimally.

4.1 An Exact MIP Formulation

We can encode SMP by formulating each species' connectivity requirements as a multicommodity flow mixed integer program, binding each of these sets of constraints together by a common set of variables x_v that indicate which nodes are to be bought. For each species, we must connect its terminals T_i using only nodes in V_i, and we can model this as having to choose $|T_i| - 1$ paths from a designated source $s_i \in T_i$ to each of the other (sink) terminals $T_i' = T_i \setminus \{s_i\}$. For each pair of terminals (s_i, t) where $t \in T_i'$, we have a set of node and directed edge flow variables f_v^{it} and f_{uv}^{it} for all nodes and edges in $G(V_i)$. Using $\Gamma_i(v)$ to indicate the nodes

$$\min \sum\nolimits_{v\in V} c_v x_v \quad (3)$$

$$\text{s.t.} \quad in^{it}_{s_i} = 0 \qquad \forall i \in P, t \in T'_i \quad (4)$$

$$out^{it}_{s_i} = 1 \qquad \forall i \in P, t \in T'_i \quad (5)$$

$$f^{it}_{s_i} = 1 \qquad \forall i \in P, t \in T'_i \quad (6)$$

$$in^{it}_{t} = 1 \qquad \forall i \in P, t \in T'_i \quad (7)$$

$$out^{it}_{t} = 0 \qquad \forall i \in P, t \in T'_i \quad (8)$$

$$f^{it}_{t} = 1 \qquad \forall i \in P, t \in T'_i \quad (9)$$

$$in^{it}_{v} = f^{it}_v \qquad \forall i \in P, t \in T'_i, v \neq s_i, t \in V_i \quad (10)$$

$$f^{it}_{v} = out^{it}_v \qquad \forall i \in P, t \in T'_i, v \neq s_i, t \in V_i \quad (11)$$

$$f^{it}_{v} \leq x_v \qquad \forall i \in P, t \in T'_i, v \in V_i \quad (12)$$

$$x_v \in \{0, 1\} \qquad \forall v \in V \quad (13)$$

$$f^{it}_{v}, f^{it}_{uv} \geq 0 \qquad \forall i \in P, \forall v \in V_i, \forall u \in \Gamma_i(v) \quad (14)$$

Figure 3: *Using the expressions from* (1) *and* (2)*, this multicommodity flow MIP exactly captures SMP. For each species, a unique commodity is defined for each sink terminal* $t \in T'_i = T_i \setminus \{s_i\}$*, and exactly 1 unit of flow of this type must travel from the source* s_i *to the sink* t*. The function* $\Gamma_i(v)$ *denotes the neighbors of* v *in* V_i*.*

in V_i which neighbor v, we define the following expressions for incoming and outgoing flow:

$$in^{it}_v = \sum\nolimits_{u\in\Gamma_i(v)} f^{it}_{uv} \quad (1)$$

$$out^{it}_v = \sum\nolimits_{u\in\Gamma_i(v)} f^{it}_{vu} \quad (2)$$

The complete MIP is shown in Figure 3. Constraints (4)-(5) force s_i to be the source of 1 unit of flow for each remaining terminal for species i. The corresponding sink constraints are captured by (7)-(8). Constraints (6) and (9) force the endpoints s_i and t of each flow to be in the solution. Constraints (10) and (11) capture flow conservation for all other nodes while preventing flow through a node v if f^{it}_v is set to 0. By (12), the node flow variable f^{it}_v can only be positive if v has been chosen for the solution, i.e. x_v has been set to 1. The last two constraints (13) and (14) force the indicator variables to be integer and the flow to be nonnegative. An optimal solution to this MIP is a minimum Steiner multigraph. Each set of terminals must be connected by the set of nodes $V' = \{v \in V : x_v = 1\}$ since there must be a feasible flow from s_i to every other terminal $t \in T'_i$, and any such flow must use some path in V_i that can only travel via nodes that have nonzero capacity, i.e. nodes for which $x_v = 1$.

We note that we relax the variables f^{it}_v to be continuous variables because in any feasible solution for which a variable f^{it}_v has been set to a fractional value, we can round its value up without violating any constraints nor affecting the objective value. The formulation also technically allows the existence of cycles in the multicommodity flow, but these will not appear in the optimal solution when the total cost is minimized unless the cycle is effectively free, in which case it has no bearing on the solution. We also considered a single commodity flow formulation that uses only a single set of flow variables for each species, and we report some experimental results in Section 5 for comparison.

4.2 Exact Laminar DP Algorithm for $|P| = 2$

For LSMP, we describe a DP algorithm for when $|P| = 2$; we will refer to this as the *Laminar DP Algorithm* (LDP). For $|P| = 2$, laminarity implies either (a) $V_1 \cap V_2 = \emptyset$ or (b) $V_1 \subseteq V_2$ without loss of generality. For (a), it is sufficient to find the optimal Steiner tree for the two disjoint sub-instances with a DP algorithm which runs in time $\tilde{O}(3^{|T_1|} + 3^{|T_2|})$ (Dreyfus and Wagner 1972). For (b), the optimal solution is either (1) the union of two disjoint Steiner trees, or (2) one connected component. We therefore solve for both cases and choose the cheaper solution. We now describe how to solve for the latter case.

As one connected component, the optimal solution consists of a single tree which can be viewed as a star of connected components: the center is a tree $H_1 \subseteq V_1$ connecting all of T_1 (and possibly some members of T_2), and the leaf components are optimal Steiner trees H_k on $T^k_2 \cup \{v^k_1\}$ where $T^k_2 \subseteq T_2$ and $v^k_1 \in H_1$. To find this solution, we extend the Dreyfus-Wagner algorithm to recursively compute the optimal trees that connect all possible sets $X_1 \cup X_2 \cup \{v_1\}$ where $X_1 \subseteq T_1$, $X_2 \subseteq T_2$, $v_1 \in V_1$, and all paths between nodes in X_1 and v_1 use only nodes in V_1.

Without loss of generality, we assume that $T_1 \cap T_2 = \emptyset$. For any terminal t in the intersection, any feasible solution will still be feasible if t is removed from T_2 while forcing the sets T_1 and T_2 to be connected. Using the Dreyfus-Wagner DP algorithm, we recursively calculate the following functions for all $X_2 \subseteq T_2$ and $v \in V_2$.

- $s^2(X_2, v_2)$: cost of the optimal tree connecting $X_2 \cup \{v_2\}$
- $s^2_v(X_2, v_2)$: cost of the optimal tree connecting $X_2 \cup \{v_2\}$ where v_2 has degree at least 2

To compute the full star of connected components, we now define the following functions.

- $s^1(X_1, X_2, v_1)$: cost of the optimal tree connecting $X_1 \cup \{v_1\}$ using only nodes in V_1 while also connecting $X_1 \cup \{v_1\} \cup X_2$ using any nodes in the graph
- $s^1_v(X_1, X_2, v_1)$: cost of the optimal tree connecting $X_1 \cup \{v_1\}$ using only nodes in V_1 and connecting $X_1 \cup \{v_1\} \cup X_2$ using any nodes in the graph such that the node v_1 has degree at least 2.

Note that $s^1(\emptyset, X_2, v_1) = s^2(X_2, v_1)$ and $s^1_v(\emptyset, X_2, v_1) = s^2_v(X_2, v_1)$. Let $d_1(u, v)$ be the shortest path distance between nodes u and v using only nodes in V_1, including the weight of u and v. We then have the following base case.

$$s^1(t_1, \emptyset, v_1) = d_1(t_1, v_1) \qquad \forall t_1 \in T_1, v_1 \in V \quad (15)$$

The function $s^1_v(X_1, X_2, v_1)$ is recursively computed by finding the pair of sets $(X'_1 \subseteq X_1, X'_2 \subseteq X_2)$ that satisfies $\emptyset \neq X'_1 \cup X'_2 \neq X_1 \cup X_2$ and minimizes

$$s^1(X'_1, X'_2, v_1) + s(X_1 \setminus X'_1, X_2 \setminus X'_2, v_1) - c_v \quad (16)$$

The function $s^1(X_1, X_2, v_1)$ is recursively defined by taking the cheaper option from (a) requiring v_1 to have degree at least 2 (which has cost $s^1_v(X_1, X_2, v_1)$), and (b) connecting v_1 to a node w in a precomputed tree connecting X_1 and X_2. In this latter case, v_1 and w must both be connected to X_1 using only nodes in V_1. Specifically, we find the node $w \in V_1$ that minimizes the following.

$$\begin{cases} d_1(v,w) + s^1(X_1 \setminus \{w\}, X_2, w) - c_w & \text{if } w \in X_1 \\ d_1(v,w) + s^1(X_1, X_2 \setminus \{w\}, w) - c_w & \text{if } w \in X_2 \\ d_1(v,w) + s^1_v(X_1, X_2, w) - c_w & \text{otherwise} \end{cases} \tag{17}$$

This algorithm's running time is dominated by the computations for $s^1_v(X_1, X_2, v_1)$ where it looks up previously-computed values for all nonempty proper subsets of $X_1 \cup X_2$. In aggregate, this amounts to $O(3^{|T_1 \cup T_2|})$ subset combinations since each terminal $t \in T_i$ must be in one of three sets: X'_i, $X_i \setminus X'_i$, or $T_i \setminus X_i$. The total running time is dominated by the exponential terms and takes $\tilde{O}(3^{|T_1 \cup T_2|})$ time.

4.3 Approximation Algorithms

We can use Steiner tree algorithms as building blocks for approximation algorithms for SMP. Since any solution to an SMP instance must include a Steiner tree for each subgraph $G(V_i)$, we can compute a Steiner tree for each subgraph and take their union. Furthermore, the optimal Steiner tree for each sub-instance is a lower bound on the overall cost of the optimal SMP solution, so the union of the optimal Steiner trees costs at most the number of Steiner trees, or $|P|$, times OPT. By using the exact Dreyfus-Wagner DP algorithm, this achieves an approximation ratio of $|P|$ in $\tilde{O}(3^{|T_m|})$ where $m = \arg\max_{i \in P} |T_i|$. We can further improve this iterative solution by taking past decisions into account: after each Steiner tree has been computed, we mark used nodes as free for later Steiner tree computations. We will refer to this approximation algorithm as the *Iterative DP Algorithm* (IDP). For instances with more terminals where the DP algorithm is no longer practical, faster approximation algorithms can be substituted for an exact Steiner tree algorithm for an overall SMP approximation guarantee of $\alpha|P|$ when the approximation algorithm used has a guarantee of α.

4.4 Primal-Dual Heuristic

One Steiner forest algorithm that can be naturally adapted to SMP is the primal-dual 6-approximation algorithm for planar Steiner forest (Demaine, Hajiaghayi, and Klein 2009). We will refer to this as the *Primal-Dual Algorithm* (PD). The algorithm maintains a (primal) infeasible solution F of nodes along with a (dual) set of variables y_{iS} for all $i \in P, S \subseteq V_i$. It proceeds as follows.

1. Initialize $F = \emptyset$ and all $y_{iS} = 0$.
2. Define $z_{iv} = \sum_{S \subseteq V_i : v \in \Gamma_i(S)} y_{iS}$ for all $i \in P, v \in V_i$, where $\Gamma_i(S)$ indicates the node neighbors of S in $G(V_i)$.
3. Let $Viol(F)$ be the connected components in each $F \cap V_i$ for which connectivity is not yet satisfied (i.e. they contain some proper nonempty subset of T_i).
4. While F is not a feasible solution, increase y_{iS} for all S in $Viol(F)$ until $\sum_{i \in P} z_{iv} = c_v$ for some v. Add v to F, and update $Viol(F)$ accordingly.
5. In reverse order in which the nodes were added, delete the nodes in F that are not needed for connectivity.

Algorithm	Time (s)	Max Gap	Med. Gap
Laminar DP (LDP)	1.600	0%	0%
Iterative DP (IDP)	1.980	11.1%	2.2%
Primal-Dual (PD)	0.260	27.6%	6.9%

Table 1: *We tested the LDP and the approximate approaches on 100 25x25 ($|V| = 625$) grid graphs where $|P| = 2$, $|T_1| = |T_2| = 3$, $V_2 = V$, and $|V_1| = 575 = 0.92|V|$. Reported here are the median running times and optimality gaps. On these instances, the MIP solver often took more than a few hours and was effectively superseded by the LDP algorithm.*

Unfortunately, this algorithm has no theoretical guarantees and can perform arbitrarily badly even for planar and laminar SMP instances. For an example of such a case, see (Lai et al. 2011). Nevertheless, this heuristic still performed very well in practice, as we report in the next section.

5 Experimental Results

5.1 Synthetic Instances

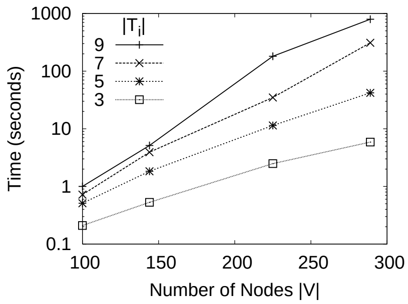
To test our solution approaches fully, we created synthetic planar instances on square grid graphs with node costs generated uniformly at random in the range [50, 1000]. Each set of terminals was chosen by placing two terminals in the opposite corners of the grid graph and adding the rest by sampling the grid graph uniformly at random.

We first tested the performance of the multicommodity flow MIP formulation (Section 4.1) using IBM ILOG CPLEX, as well as the IDP and PD algorithms described in Sections 4.3 and 4.4, respectively. Figure 4a shows the median MIP solver running times for different size grid graphs on two sets of terminals of varying sizes. Each V_i was chosen uniformly at random from the graph and was of size $0.85|V|$. As might be expected, the time required to solve the MIP scales exponentially with $|V|$, though it is notable that it still finishes within minutes for even hundreds of nodes. The performance of the IDP and PD algorithms were also tested on these instances (see Figure 4c). Though the IDP algorithm is only guaranteed to have a maximum optimality gap of 100%, it performed surprisingly very close to optimal (within 3%) more than half the time. The IDP and the PD algorithms were also fast on these instances, always finishing within 6 seconds and 1 second, respectively.

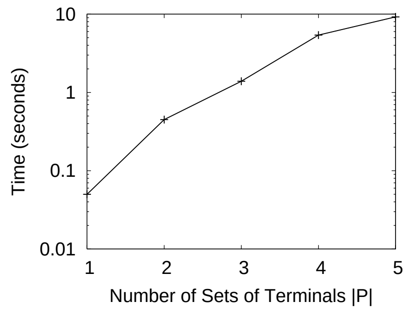
We tested the LDP algorithm described in Section 4.2 on even larger instances, and we compared it against the IDP and PD algorithms. The results are tabulated in Table 1. While the PD algorithm is very fast, it generally performed worse than the IDP algorithm. However, for most of the cases, both heuristics performed within 10% of optimal.

5.2 Wildlife Corridor Design in Montana

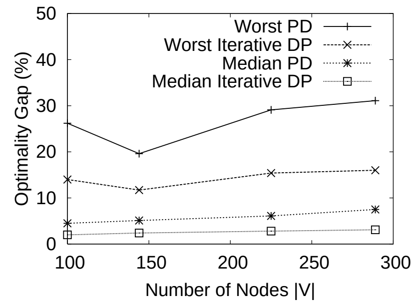
We used our solution approaches to solve the wildlife corridor design problem for wolverines and Canada lynx in Montana. Because the lynx can use only a subset of the land that the wolverines can traverse, this is exactly an instance of LSMP. We studied the problem using a 6km resolution of square cells, with each cell having a cost and binary values for wolverine and lynx permeability. The 6km cell size



(a) Median MIP running times



(b) Median MIP running times



(c) Performance of non-exact algorithms

Figure 4: *(a) The MIP solver was run on 100 samples each of square grid graph instances with* $|P|=2$, $|T_1|=|T_2|$, *and the subgraphs chosen uniformly at random such that* $|V_1|=|V_2|=0.85|V|$ *(b) The MIP solver was run on 100 samples each of 15x15 square grid graphs with* $|V_i|=191=0.85|V|$, $|T_i|=3$ *(c) Optimality gaps for the IDP and PD algorithms (same instances as in (a)).*

was chosen to be consistent with statistically downscaled climate data used to model wolverine habitat (McKelvey et al. In press). The data preprocessing was performed using ArcGIS 9.3. For most of the data used, we used the publicly available data sets downloadable from Montana's website. This included elevation (USGS National Elevation), land cover (USGS GAP analysis 2010), and roads (Census 2010 TIGER/Line). Housing density was derived from Census 2010 data. To calculate the permeability layer for the Canada lynx, we combined the factors of slope, land cover, road density, and housing density as per the formula reported by (Bates and Jones 2007). To generate the costs, we regarded all already-conserved land as free and otherwise used the taxable land value data from 2007. For cells which overlap some part of a primary road, we added a cost estimate for installing a wildlife bypass such as an overpass. To define the wolverines' terminals, we found the contiguous areas with persistent spring snow cover (queried from MODIS as per (Copeland et al. 2010)) already under conservation, resampled them to 6km, and then eliminated areas that were not large enough for a female wolverine home range. Terminals already connected by preserved land were consolidated for a total of 13 wolverine terminals. The terminals for the lynx were defined as the already-conserved areas in the NCDE and GYA designated as critical habitat by the US Fish and Wildlife Service; these terminals were consolidated into 4 terminals. After pruning the study area to exclude eastern Montana, lakes, and other barriers such as urban areas, the resulting graph contains a total of 4514 cells of 6km by 6km, 3326 of which are accessible to the lynx.

Given that the budget may not be large enough to connect all the terminals, we created scenarios using different subsets of the defined wolverine and lynx terminals, prioritizing the larger terminals. We report results on our solution approaches on these scenarios in Table 2. The PD algorithm was the fastest of all our solution approaches. While it has no performance guarantees, it performed close to optimal on the real data. We show the result of the PD heuristic as compared to the optimal solution in Figure 5. For exact solutions, we tested both the single-commodity and multicommodity flow encodings for the MIP (of which we had not seen much of a difference on the smaller randomized instances). We found

(L,W) Instance	SCF Gap	MCF Time	LDP Time	PD Time	PD Gap
(0,8)	(13.6h)	3.13m	5.25m	5.46s	.02%
(2,8)	17.5%	4.08m	11.5m	7.01s	.02%
(4,8)	(17.45h)	1.93m	27.7m	8.67s	.02%
(0,13)	20.9%	37.34h	3.28h	5.70s	2.8%
(2,13)	31.1%	29.04h	7.87h	7.22s	2.8%
(4,13)	24.8%	42.2m	19.35h	9.10s	6.7%

Table 2: *Different instances of the wildlife corridor design problem in Montana were solved for different subsets of the species' terminals. They are indicated as pairs (L,W) for L lynx terminals and W wolverine terminals. The best-known optimality gaps are reported for the single-commodity flow (SCF) after 40 hours (or the final running time if less than 40 hours). Running times are reported for solving the multicommodity flow (MCF) encoding, the LDP algorithm (LDP) (not including preprocessing time for calculating all-pairs shortest paths: 6.55 minutes when L=0, 9.10 minutes when L>0), and the PD algorithm (PD). The last column reports the optimality gap for the PD algorithm.*

that the multicommodity flow encoding, while adding more variables and constraints, helped introduce tighter bounds on the problem and allowed the solver to prove optimality faster. For these experiments, we set the CPLEX parameters according to those found by the CPLEX automatic tuning utility. Comparing the different scenarios, increasing the number of lynx terminals in the instance oftentimes actually decreased the running times. As this was an LSMP instance, we tested the LDP algorithm which also finds an optimal solution. As expected, its running times increased exponentially with the total number of lynx and wolverine terminals. Surprisingly, because of the inherent structure of the real-world instance, solving the MIP was sometimes faster than the LDP algorithm, especially as the number of terminals increased. While heuristics such as the PD algorithm can give fast solutions that are close to optimal, real-world instances of SMP can still be practical to solve to optimality.

6 Previous Work

In this paper, we present the Steiner Multigraph Problem, a new network design problem under a multigraph setting. This problem is closely related to problems in network design and multicommodity flow. The *Steiner tree* problem

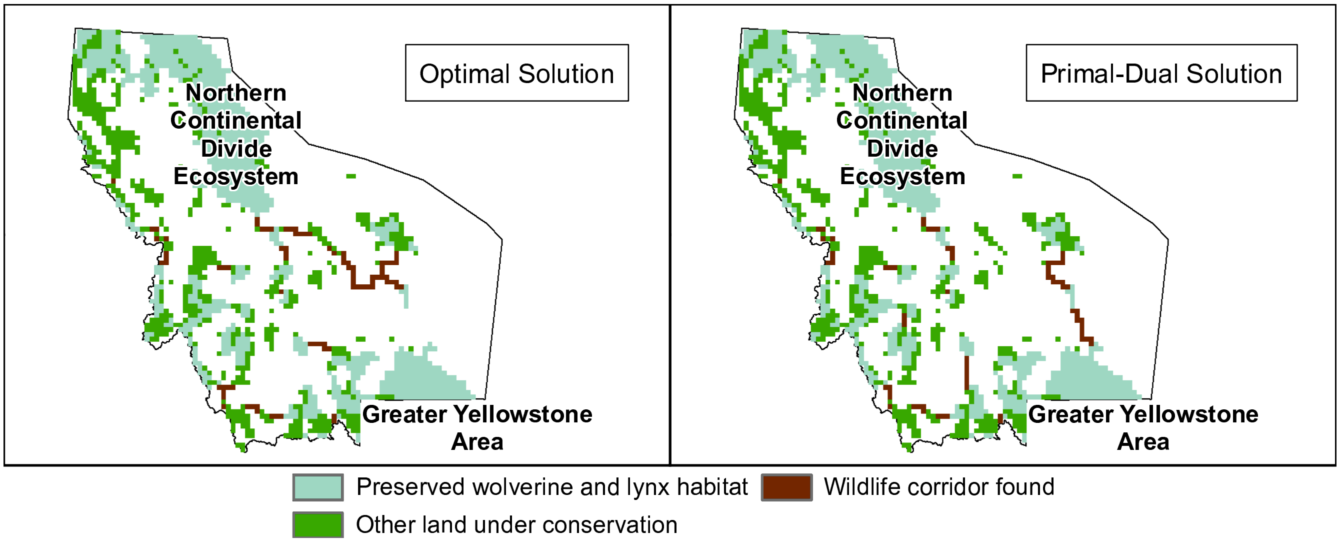


Figure 5: *For 4 lynx terminals and 13 wolverine terminals, the MIP solver found an optimal solution in a surprisingly practical amount of time: 42.2 minutes (left). The PD heuristic, while much faster at 9.1 seconds, found a suboptimal solution that cost 6.7% more (right).*

is a special case of SMP and at the core of many other network design problems in combinatorial optimization. In its simplest form, the Steiner tree problem is a generalization of the minimum spanning tree: on an undirected graph $G = (V, E)$ with nonnegative costs on the edges, find a minimum-cost set of edges that connect a particular subset of nodes $T \subseteq V$, or the *terminals* of the graph. This problem is well-studied and is known to be APX-complete, i.e. there is some constant factor within which it cannot be approximated in polynomial time unless P = NP (Promel and Steger 2002). The current best approximation factor is 1.39 for general graphs (Byrka et al. 2010), though there exist polynomial time approximation schemes (or PTAS) for Euclidean Steiner trees, planar graphs, and graphs of bounded treewidth, which have a guarantee of $1 + \epsilon$ for any $\epsilon > 0$ (Arora 1998; Bateni, Hajiaghayi, and Marx 2010). The problem is also fixed-parameter tractable (FPT), admitting an exact dynamic programming (DP) algorithm that runs in time exponential in $|T|$ (Dreyfus and Wagner 1972).

In the variant called the *node-weighted Steiner tree* problem, costs appear on the nodes instead. This is more appropriate for capturing certain applications such as modeling wireless sensor networks or adjacent parcels of land. While this reformulation may seem like a trivial change to the problem, it is computationally harder. For general graphs, the best possible approximation factor is not a constant but $\Theta(\log |T|)$ (Klein and Ravi 1995), unless P = NP. However, just as in the case of the original edge-weighted Steiner tree problem, the problem is still in FPT, admitting a similar DP algorithm. Better approximation guarantees can also be found for restricted classes of graphs that appear in practice. For example, there is a primal-dual 6-approximation algorithm for the node-weighted Steiner tree problem in planar graphs (Demaine, Hajiaghayi, and Klein 2009).

Steiner forest replaces the set of terminals T with a family of multiple disjoint sets $T_1, \ldots, T_k \subseteq V$ (Agrawal, Klein, and Ravi 1991). A feasible solution must connect each set of terminals T_i, but terminals in different sets need not be connected. An optimal solution is a forest of trees, each tree connecting some of the terminal sets.

Currently, the best theoretical result for the connection subgraph problem posed by (Conrad et al. 2007) for single-species wildlife corridor design is an approximation algorithm which finds a tree with utility at most a factor $O(\frac{1}{\epsilon} \log |V|)$ worse than optimal and may violate the budget constraint by up to a factor of $\frac{1}{\epsilon}$ (Rabani and Scalosub 2008).

7 Conclusions and Future Work

In this paper, we introduced the Steiner Multigraph Problem, a new combinatorial optimization problem that captures the connectivity requirements for multiple species in a common wildlife corridor. We showed that this problem is harder than Steiner forest in general, but it is fixed-parameter tractable for the special case where the demand subgraphs have special structure. We presented a few different solution approaches and tested them on a set of synthetic instances as well as the wildlife corridor problem in Montana. While solving the MIP generally takes significantly longer than the inexact approaches and scales exponentially, the running times when solving the real-world instance in Montana with thousands of nodes were still practical and gave a guarantee on the optimality of the solution. Most notably, the additional consideration of the lynx decreased the running time as compared to solving the minimum-cost wildlife corridor design problem for just the wolverines.

This work opens up several followup questions. For the purposes of wildlife corridor design, one can add a budget constraint to the problem and maximize a function of the species-specific habitat suitability values for the land included in the corridor. Another open question is to find an algorithm that has a stronger approximation guarantee than the one from iteratively solving the Steiner tree instances. It would also be interesting to look into other graph problems that can be meaningfully defined in this multigraph setting.

Acknowledgments

This research was supported by NSF Expeditions in Computing award for Computational Sustainability (Grant 0832782) and by the USDA Forest Service, Rocky Mountain Research Station (10-JV-11221635-24). Katherine J. Lai is supported by a Graduate Research Fellowship from the National Science Foundation. We thank Maarika Teose and Rachel Houtman for their help with the data, as well as the anonymous reviewers for their thoughtful comments.

References

Agrawal, A.; Klein, P.; and Ravi, R. 1991. When trees collide: An approximation algorithm for the generalized Steiner problem on networks. In *STOC*, 134–144.

Ahn, Y.-Y.; Bagrow, J. P.; and Lehmann, S. 2010. Link communities reveal multiscale complexity in networks. *Nature* 466(7307):761–764.

Arora, S. 1998. Polynomial time approximation schemes for euclidean traveling salesman and other geometric problems. *J. ACM* 45(5):753–782.

Bateni, M.; Hajiaghayi, M.; and Marx, D. 2010. Approximation schemes for Steiner forest on planar graphs and graphs of bounded treewidth. In *STOC*, 211–220.

Bates, W., and Jones, A. 2007. Least-Cost Corridor Analysis for Evaluation of Lynx Habitat Connectivity in the Middle Rockies.

Beier, P.; Majka, D. R.; and Spencer, W. D. 2008. Forks in the road: Choices in procedures for designing wildland linkages. *Conservation Biology* 22(4):836–851.

Byrka, J.; Grandoni, F.; Rothvoß, T.; and Sanità, L. 2010. An improved LP-based approximation for Steiner tree. In *STOC*, 583–592.

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*, 16–28.

Copeland, J.; McKelvey, K.; Aubry, K.; Landa, a.; Persson, J.; Inman, R.; Krebs, J.; Lofroth, E.; Golden, H.; Squires, J.; Magoun, a.; Schwartz, M.; Wilmot, J.; Copeland, C.; Yates, R.; Kojola, I.; and May, R. 2010. The bioclimatic envelope of the wolverine (Gulo gulo): do climatic constraints limit its geographic distribution? *Canadian Journal of Zoology* 88(3):233–246.

Demaine, E.; Hajiaghayi, M.; and Klein, P. 2009. Node-Weighted Steiner Tree and Group Steiner Tree in Planar Graphs. In *ICALP*, 328–340.

Dilkina, B., and Gomes, C. P. 2010. Solving Connected Subgraph Problems in Wildlife Conservation. In *CPAIOR*, 102–116.

Dreyfus, S., and Wagner, R. 1972. The steiner problem in graphs. *Networks* 1:195–207.

Even, S.; Itai, A.; and Shamir, A. 1976. On the complexity of time table and multi-commodity flow problems. *SIAM J. Comput.* 5(4):691–703.

Fagan, W., and Holmes, E. 2006. Quantifying the extinction vortex. *Ecology Letters* 9(1):51–60.

Faloutsos, C.; McCurley, K. S.; and Tomkins, A. 2004. Fast discovery of connection subgraphs. In *KDD*, 118.

Gardner, C.; Ballard, W.; and Jessup, R. 1986. Long distance movement by an adult wolverine. *Journal of Mammalogy* 67(3).

Gilpin, M., and Soule, M. 1986. Minimum viable populations: processes of extinction. In Soule, M., ed., *Conservation Biology: The Science of Scarcity and Diversity*. Sinauer Associates. 19–34.

Gomes, C. P.; van Hoeve, W.-J.; and Sabharwal, A. 2008. Connections in networks: A hybrid approach. In *CPAIOR*, 303–307.

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

Haddad, N. M.; Bowne, D. R.; Cunningham, A.; Danielson, B. J.; Levey, D. J.; Sargent, S.; and Spira, T. 2003. Corridor Use By Diverse Taxa. *Ecology* 84(3):609–615.

Inman, R. M.; Wigglesworth, R.; Inman, K.; Schwartz, M. K.; Brock, B.; and Rieck, J. 2004. Wolverine makes extensive movements in Greater Yellowstone Ecosystem. *Northwest Science* 78:261–266.

Klein, P., and Ravi, R. 1995. A nearly best-possible approximation algorithm for node-weighted Steiner trees. *J. Algorithm* 19(1):104–115.

Lai, K. J.; Gomes, C. P.; Schwartz, M. K.; McKelvey, K. S.; Calkin, D. E.; and Montgomery, C. S. 2011. The Steiner Multigraph Problem: Wildlife Corridor Design for Multiple Species. Technical report, Cornell University.

Lichtenstein, D. 1982. Planar Formulae and Their Uses. *SIAM J. Comp.* 11(2):329–343.

McKelvey, K. S.; Copeland, J. P.; Schwartz, M. K.; Littell, J.; Aubry, K.; Squires, J. R.; Parks, S.; Elsner, M.; and Mauger, G. In press. Climate change predicted to shift wolverine distributions, connectivity, and dispersal corridors. *Ecological Applications*.

Moriarty, K. M.; Zielinski, W. J.; Gonzales, A. G.; Dawson, T. E.; Boatner, K. M.; Wilson, C. a.; Schlexer, F. V.; Pilgrim, K. L.; Copeland, J. P.; and Schwartz, M. K. 2009. Wolverine Confirmation in California after Nearly a Century: Native or Long-Distance Immigrant? *Northwest Science* 83(2):154–162.

Newman, M., and Girvan, M. 2004. Finding and evaluating community structure in networks. *Physical Review E* 69(2).

Promel, H. J., and Steger, A. 2002. *The Steiner Tree Problem: A Tour through Graphs, Algorithms, and Complexity*. Friedrick Vieweg & Son.

Rabani, Y., and Scalosub, G. 2008. Bicriteria approximation tradeoff for the node-cost budget problem. In *SWAT*, 90–101.

Salhieh, A.; Weinmann, J.; and Kochhal, M. 2001. Power efficient topologies for wireless sensor networks. In *ICPP*, 156–163.

Schwartz, M.; Mills, L.; McKelvey, K.; Ruggiero, L.; and Allendorf, F. 2002. DNA reveals high dispersal synchronizing the population dynamics of Canada lynx. *Nature* 415(6871):520–522.

Schwartz, M. K.; Copeland, J. P.; Anderson, N. J.; Squires, J. R.; Inman, R. M.; McKelvey, K. S.; Pilgrim, K. L.; Waits, L. P.; and Cushman, S. a. 2009. Wolverine gene flow across a narrow climatic niche. *Ecology* 90(11):3222–32.

Shepherd, B., and Whittington, J. 2006. Response of wolves to corridor restoration and human use management. *Ecol. Soc.* 11(2).

Simon, H. 1962. The architecture of complexity. *Proceedings of the American Philosophical Society* 106(6):467–482.

Squires, J. R., and Ruggiero, L. F. 2007. Winter Prey Selection of Canada Lynx in Northwestern Montana. *Journal of Wildlife Management* 71(2):310–315.