Journal of Environmental Economics and Management 63 (2012) 1–18

Contents lists available at SciVerse ScienceDirect

Journal of Environmental Economics and Management

journal homepage: www.elsevier.com/locate/jeem

Wildlife corridors as a connected subgraph problem ☆

Jon M. Conrad [a], Carla P. Gomes [a,b], Willem-Jan van Hoeve [c], Ashish Sabharwal [d], Jordan F. Suter [e,*]

[a] Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY, USA
[b] Department of Computer Science, Cornell University, Ithaca, NY, USA
[c] Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, USA
[d] IBM Watson Research Center, Yorktown Heights, NY, USA
[e] Department of Economics, Oberlin College, 10 N. Professor St., Oberlin, OH 44074, USA

ARTICLE INFO

Article history:
Received 28 September 2010
Available online 18 August 2011

Keywords:
Wildlife corridors
Habitat fragmentation
Optimization
Mixed integer programming
Computational sustainability

ABSTRACT

Wildlife corridors connect areas of biological significance to mitigate the negative ecological impacts of habitat fragmentation. In this article we formalize the optimal corridor design as a *connected subgraph problem*, which maximizes the amount of suitable habitat in a fully connected parcel network linking core habitat areas, subject to a constraint on the funds available for land acquisition. To solve this challenging computational problem, we propose a hybrid approach that combines graph algorithms with Mixed Integer Programming-based optimization. We apply this technique to the design of corridors for grizzly bears in the U.S. Northern Rockies, illustrating the underlying computational complexities by varying the granularity of the parcels available for acquisition. The approach that is introduced is general and can be applied to other species or other similar problems, such as those occurring in social networks.

1. Introduction and overview

In many parts of the world land development has resulted in a reduction and fragmentation of natural habitat, leading to increased rates of species decline and extinction. To combat the negative consequences of anthropogenic habitat fragmentation, the procurement of biologically valuable conservation land has been promoted as a way to ensure species viability. A large number of models for optimally selecting land parcels for conservation, formally referred to as the reserve site selection problem (RSSP), have been proposed in the conservation biology literature. These models select parcels to ensure that all targeted species in a given region are protected, as in the Set Covering Problem (SCP) (e.g. [20,34]), or they select a constrained number of parcels that maximize species richness, as in the Maximal Covering Problem (MCP) (e.g. [4,6]).

A number of subsequent studies have added to the conservation biology literature by incorporating economic variables into the RSSP. These studies seek to procure conservation parcels, given a budget constraint, that maximize the number of species protected (e.g. [1,7,27]) or maximize the environmental benefits of the sites selected (e.g. [13,22,24]). The results of

☆ We thank the associate editor and two anonymous referees for their helpful comments and suggestions. This work was supported by NSF Expeditions in Computing award for Computational Sustainability (Grant 0832782), NSF IIS award (Grant 0514429), and the Intelligent Information Systems Institute, Cornell University (AFOSR Grant FA9550-04-1-0151).

* Corresponding author. Fax: +1 440 775 6978.
E-mail address: jordan.suter@oberlin.edu (Jordan F. Suter).

0095-0696/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jeem.2011.08.001

these economic-based studies show that incorporating spatially heterogeneous financial costs into reserve site selection models leads to a substantially different set of priority parcels than standard SCP or MCP models that ignore parcel costs. Importantly, the parcels selected based on budget constrained optimization obtain considerably greater environmental benefits for the same conservation budget than traditional site selection models [23].

In recent years, researchers have recognized that a parcel's spatial location relative to other protected parcels is also an essential attribute to consider in reserve site selection. Reflecting this, a variety of models that seek to increase the degree of spatial coherence in the set of parcels selected for conservation have been developed ([40] provide a thorough review). One primary way in which spatial attributes have been incorporated into site selection models is through the optimal selection of a connected reserve network, which we refer to as a wildlife corridor.[1] The focus on developing models for the design of optimal wildlife corridors has come as biologists have highlighted the environmental imperative of connecting core areas of biological significance [25].

Properly implemented wildlife corridors provide numerous ecological benefits by returning the landscape to its natural connected state. By allowing species the ability to migrate between core areas of biological significance, corridors increase gene flow and reduce rates of inbreeding, thereby improving species fitness and survival [29]. Corridors also allow for greater mobility [2], thus allowing species to escape predation and respond to stochastic events such as fire. In addition, corridors allow species to respond more easily to long-term climatic changes [21].

Responding to the ecological benefits of connected ecosystems, a wide range of corridor projects have been proposed or are currently being implemented worldwide. Yet despite the increasing number of corridors being implemented and several studies documenting the positive ecological benefits of existing corridors (e.g. [11,16,31]), *designing models for the optimal selection of corridor parcels* has received comparatively little attention.

Research on the optimal design of wildlife corridors commenced with the work of Sessions [30], who models the selection of a hypothetical corridor as a network Steiner tree (NST) problem. Subsequent models of optimal corridor design have focused on minimizing the number of parcels selected in a corridor such that a specific number of species are preserved [5,14,26]. Of particular note is research by Önal and Briers [26], who utilize graph theory and linear integer programming to find a species-covering connected subgraph within a larger geographic area. Specifically, their model seeks to identify the minimum number of connected parcels that fully cover 118 separate bird species, without including predefined reserves or parcel-level costs. The model incorporates a tail function approach, which amounts to a set of creative linear constraints that prevent cycling. The solution procedure that the authors implement involves solving the problem at a more aggregate scale and then selecting the optimal set of small disaggregate sites within the aggregate solution. This selection algorithm is found to out-perform a heuristic procedure that is an extension of the greedy algorithm.

Other research has dealt more explicitly with heterogeneous parcel cost in corridor design. For example, Williams [38] formulates a NST model with the bi-objectives of minimizing corridor cost and the amount of unsuitable area included in the corridor. Using binary linear programming, the model implements a multi-objective weighting method to generate a set of non-inferior solutions to a hypothetical corridor example. These solutions then allow for a comparison of the tradeoffs between aggregate corridor cost and habitat suitability. In subsequent work, Williams modifies his original model to consider cases where there are no predefined reserves and the planner is simply trying to form a connected reserve network [39,41].

More recent work by Tóth et al. [32] introduces a bi-objective model that seeks to minimize the cost of parcel acquisition and maximize the total area of protected habitat for bird populations in suburban Chicago. They do not deal with corridor design directly, however, as their formulation includes a constraint on the minimum area of contiguous habitat for a parcel to be included rather than a connectivity constraint. The contiguity constraints are realized through a cluster enumeration algorithm that limits the number of parcel clusters under consideration before solving the binary parcel selection problem.

In this article we develop a model that seeks the optimal construction of a wildlife corridor between multiple areas of biological significance. We propose a budget constrained optimization model and a corresponding hybrid solution methodology that efficiently incorporates both economic and ecological information in the design of optimal corridors. These techniques are then applied to the design of a wildlife corridor for grizzly bears connecting the Yellowstone, Salmon–Selway[2] and Northern Continental Divide Ecosystems in Idaho, Wyoming and Montana.

The primary contribution of our work is the formalization of the corridor design problem as a graph theory problem that is referred to as a connected subgraph problem. Expanding on earlier research, this approach allows us to focus on the computational issues of the problem, independently of the particular domain. It also highlights the fact that other problems, with the same structure as the corridor design problem, as they occur, for example, in social network applications, can be modeled as a connected subgraph problem. We formally characterize the worst-case computational complexity of the connected subgraph problem as a so-called NP-Hard problem. In order to further understand its *typical case* complexity, beyond the standard NP-Hard worst-case notion used in computer science, we developed a randomized

[1] Wildlife corridors are also referred to more or less interchangeably as conservation, habitat, and movement corridors.

[2] The Salmon–Selway Ecosystem is also referred to as the Bitterroot Ecosystem.

generator of "synthetic" instances of the connected subgraph problem, using semi-structured graphs.[3] By studying the behavior of different algorithms, combined with different model formulations, on synthetic instances, we gained insights into the structure of the problem. Our analysis led to the discovery of an interesting "easy-hard-easy" pattern in the typical computational complexity of proving optimality for instances of this problem. Such insights led to the development of a hybrid algorithm that exploits the structure of the problem.

Our hybrid algorithm for the connected subgraph problem allows us to dramatically scale up solutions. The algorithm incorporates a *provably* efficient procedure (i.e., it runs in polynomial time) for computing the optimal minimum cost corridor. The procedure involves using the minimum cost solution to initialize a general procedure that allows the algorithm to converge to the overall optimal solution much faster. The algorithm also incorporates so-called propagation and pruning techniques that considerably speed up the solution procedure by ruling out early candidate solutions that are guaranteed to be sub-optimal. The resulting hybrid algorithm, described in detail in this article, performs remarkably well, with *strong optimality guarantees*, both when considering the synthetic instances and instances of the real-world wildlife corridor for grizzly bears connecting the Yellowstone, Salmon–Selway and Northern Continental Divide Ecosystems. For the real-world wildlife corridor problem we analyze instances with over 12,000 parcels. Although we are not able to solve the largest instances to optimality at budgets higher than the minimum cost corridor, the solutions that we do find are provably very close to the optimum. For example, when considering a budget of $8 M (the minimum cost of a corridor is $7.2 M) our procedure provides a solution that is guaranteed to be within 1% of the optimal solution. For instances where computational constraints prevent us from identifying the optimal solution we also test a heuristic, described in more detail below, that uses the minimum cost corridor as a baseline for the optimization routine on the remaining parcels.

Our approach diverges from previous corridor design studies in that we do not require the corridor to have a "tree" structure, which is what one obtains in models that seek the minimum cost Steiner tree as the "best" wildlife corridor. For example, the models of Sessions [30] and Williams [5,14,26] limit their attention to corridors in which the path can have a width of at most one parcel. Although utilizing a tree structure decreases the computational complexity of the problem, we feel that the model that we present is more general because it means that the benefits of the corridor can be improved either by selecting an alternative route or by making the corridor wider so that it cost-effectively includes adjacent parcels.

Another contribution of this study is that we incorporate estimated parcel costs from a naturally occurring landscape. In addition, by changing the granularity of the parcels available for selection we gain a greater understanding of the relationship between computational complexity and the number of parcels in the landscape (see Fig. 4). We also gain insight into the tradeoffs between parcel size and model specificity.

We formally describe the corridor problem and characterize its computational complexity in Section 2. Section 3 formulates the problem as a mixed integer programming problem and describes our solution procedure in detail. Section 4 provides experimental results and describes the application of corridor design for the grizzly bear in the U.S. Northern Rockies. Section 5 concludes.

2. Problem description: wildlife corridors as connected subgraphs

We begin by mathematically defining the wildlife corridor design problem as a problem of finding a connected subgraph of a given graph with costs and utilities associated with its edges. We then give a brief analysis of this problem from the traditional worst-case complexity perspective, proving that the corresponding decision problem is NP-complete and the cost optimization variant of the problem is NP-hard to approximate within a certain constant factor.

2.1. The connected subgraph problem

Let $\mathbf{Z}^+$ denote the set $\{0,1,2,\ldots\}$ of non-negative integers. The decision version of the connected subgraph problem is defined on an undirected graph as follows:

Definition 1 (*Connected subgraph problem*). Given an undirected graph $G=(V,E)$ with terminal vertices $T \subseteq V$, vertex costs $c: V \rightarrow \mathbf{Z}^+$, vertex utilities $u: V \rightarrow \mathbf{Z}^+$, a cost bound $C \in \mathbf{Z}^+$, and a desired utility $U \in \mathbf{Z}^+$, does there exist a vertex-induced subgraph H of G such that

1. H is connected,
2. $T \subseteq V(H)$, i.e., H contains all terminal vertices,
3. $\sum_{v \in V(H)} c(v) \leq C$, i.e., H has cost at most C, and
4. $\sum_{v \in V(H)} u(v) \geq U$, i.e., H has utility at least U?

[3] An instance of a problem results from assigning concrete values to its parameters. For example, in the corridor design problem, we assign concrete values to the parcel layout, parcel utilities, parcel costs, and the budget. One can generate multiple instances by randomly assigning parameter values. The value of analyzing multiple problem instances is that it provides a much better understanding of the problem's computational complexity, as opposed to focusing on a single instance that may, or may not, be representative.

In this decision problem,[4] we can relax one of the last two conditions to obtain two natural optimization problems: (1) *Utility Maximization*: given a cost bound C, maximize the utility of H; (2) *Cost Minimization*: given a desired utility U, minimize the cost of H.

The connected subgraph problem captures the key mathematical aspects of the corridor design problem if we think of the available land parcels as vertices of a graph, reserves as terminal vertices, parcel cost (or utility) as the cost (or utility, resp.) associated with the corresponding vertex, and two land parcels sharing a boundary being equivalent to having an edge between the two corresponding vertices in the graph. A connected subgraph of this graph containing the designated terminal vertices corresponds to a wildlife corridor connecting the given reserves.

In the context of social networks, a similar problem has been investigated by Faloutsos et al. [12]. Here, one is interested, for example, in identifying the few people most likely to have been infected with a disease, or individuals with unexpected ties to any members of a list of other individuals. This relationship is captured through links in an associated social network graph with people forming the nodes. Faloutsos et al. [12] consider networks containing two special nodes (the "terminals") and explore practically useful utility functions that capture the connections between these two terminal nodes. Our interest, on the other hand, is in studying this problem with the sum-of-weights utility function but with several terminals. In either case, the problem has a bounded-cost aspect that competes with the utility one is trying to maximize.

2.2. Worst-case complexity analysis

From a computer science perspective, the first question one typically asks is how hard the problem under consideration is, in terms of the traditional computational complexity hierarchy. Broadly speaking, computer scientists consider a problem to be "easy" or efficiently solvable if there is a polynomial time algorithm (polynomial in the size of the input) that solves the problem. A large set of real-world problems belong to the so-called NP-complete class for which only exponential time algorithms are known and for which it is believed by many that no polynomial time algorithm exists.[5]

We next discuss the computational complexity of the connected subgraph problem, before moving on to our solution methodology and experimental evaluation. In order to maintain the focus of the paper on effective solution methods, this section is kept brief and all proofs are deferred to Appendix B.

The connected subgraph problem is a generalized variant of the standard *Steiner tree problem*[6] (cf. [28]) on undirected graphs, with the difference being that the costs are on vertices rather than on edges and that we have utilities in addition to costs. The utilities add a new dimension of hardness to the problem. In fact, while the Steiner tree problem is polynomial time solvable when $|T|$ is any fixed constant (cf. [28]), we will show that the connected subgraph problem remains NP-complete even when $|T|=0$. We prove this by a reduction from the Steiner tree problem. This reduction also applies to planar graphs, for which the Steiner tree problem is still NP-complete (cf. [28]).

Theorem 1 (*NP-completeness*). *The decision version of the connected subgraph problem, even on planar graphs and without any terminals, is NP-complete.*

The reader is referred to Appendix B for the relatively short proof of this theorem. The theorem immediately implies the following:

Corollary 1 (*NP-hardness of optimization*). *The cost and utility optimization versions of the connected subgraph problem, even on planar graphs and without any terminals, are both NP-hard.*

It turns out that in the NP-hardness reduction used in the proof of Theorem 1, the graph $\widehat{G}$ in the given Steiner tree instance has a Steiner tree with cost C' if and only if the graph G constructed for the connected subgraph problem has a connected subgraph with cost C'. Consequently, if the cost optimization version of the connected subgraph instance (i.e., cost minimization) can be approximated within some factor $\alpha \geq 1$ (i.e., if one can find a solution of cost at most α times the optimal), then the original Steiner tree problem can also be approximated within factor α. It is, however, known that there exists a factor α_0 such that the Steiner tree problem *cannot* be approximated within factor α_0, unless P=NP. This immediately gives us a hardness of approximation result for the utility optimization version of the connected subgraph problem. Unfortunately, the best known value of α_0 is roughly only $1+10^{-7}$ (cf. [28]).

[4] A *decision problem* is a problem with a yes–no answer. Typically, given an algorithm for the yes–no version of a problem, it is easy to produce an equally efficient algorithm that actually produces a solution if the answer is yes.

[5] NP stands for Non-deterministic Polynomial time. This captures the idea that, given a candidate solution, one can *verify* its validity as a solution in polynomial time. Note that this does not mean that one can *generate* the solution in polynomial time—being able to do that would make the problem polynomial time solvable, i.e., "easy". NP-complete problems are the *hardest* problems within the class NP and all known algorithms for them take exponential time (in the input size) in the worst case. Roughly speaking, the notion of being complete for a class means that all other problems in the class can be translated to this problem in polynomial time; therefore, if one could find a polynomial time algorithm to solve any one of the complete problems in a class such as NP, then all the problems in the class would be solved in polynomial time as well.

[6] A Steiner tree connecting a set of T vertices of a graph $G=(V,E)$ is a set U of vertices such that $T \subseteq U \subseteq V$ and the subgraph of G induced by U forms a tree, whereby any two vertices in U are connected by exactly one path.

Fortunately, we can use a different reduction – from the NP-complete Vertex Cover problem – which will enable us to derive as a corollary a much stronger approximation hardness result for the connected subgraph problem.

Lemma 1. *There is a polynomial time reduction from Vertex Cover to the connected subgraph problem, even without any terminals, such that the size of the vertex cover in a solution to the former equals the cost of the subgraph in a solution to the latter.*

We again refer the reader to Appendix B for a proof of this statement. Combining Lemma 1 with the fact that the vertex cover problem is known to be NP-hard to approximate within a factor of 1.36 [10] immediately gives us the following:

Theorem 2 (*APX-hardness of cost optimization*). *The cost optimization version of the connected subgraph problem, even without any terminals, is NP-hard to approximate within a factor of* 1.36.

We refer the reader to Appendix A for a simple example that highlights some of the combinatorial issues of the connected subgraph problem that make it computationally hard.

3. Solving the connected subgraph problem

Next we present the mixed integer linear programming model (MIP model) for the connected subgraph problem that was used in our experiments. We will then discuss a hybrid solution method for efficiently solving this MIP model.

3.1. Mixed integer linear programming formulation

Let $G=(V,E)$ be the graph under consideration, with $V=\{1,\ldots,n\}$ and budget C. We introduce a binary variable x_j for each vertex $j \in V$, representing whether or not j is in the connected subgraph. The cost associated with including vertex j in the subgraph is given by c_j while the utility is given by u_j. In addition, for each edge we introduce a non-negative variable y_{ij} to indicate the amount of flow (as described in detail below) from vertex i to vertex j. The corresponding MIP formulation is

$$\text{maximize} \quad \sum_{j\in V} u_j x_j \tag{1}$$

$$\text{subject to} \quad \sum_{j\in V} c_j x_j \le C \tag{2}$$

$$x_j \in \{0,1\} \quad \forall j \in V \tag{3}$$

$$x_t = 1 \quad \forall t \in T \tag{4}$$

$$z_0 + y_{0,\hat{t}} = n \tag{5}$$

$$y_{ij} \le n x_j \quad \forall (i,j) \in E' \tag{6}$$

$$\sum_{i:(i,j)\in E'} y_{ij} = x_j + \sum_{\ell:(j,\ell)\in E'} y_{j\ell} \quad \forall j \in V \tag{7}$$

$$\sum_{j\in V} x_j = y_{0,\hat{t}}. \tag{8}$$

The budget constraint is given by inequality (2). To ensure the connectivity of the subgraph, we apply a particular *network flow model*, where the network is obtained by replacing all undirected edges $\{i,j\} \in E$ by two directed edges (i,j) and (j,i). Call the set of directed edges E'. First, we introduce a source vertex 0, with maximum total outgoing flow n. We arbitrarily choose one terminal vertex $\hat{t} \in T$ as the "root" node, and define a directed edge $(0,\hat{t})$ to insert the flow into the network, assuming that there exists at least one terminal vertex.[7] Then, by demanding that the flow reaches all terminal vertices, the edges carrying flow (together with the corresponding vertices) represent a connected subgraph. To this end, each of the vertices with a positive incoming flow will act as a "sink", by "consuming" one unit of flow. In particular, all terminal vertices will act as sinks, and any other vertex that is part of the eventual connected subgraph will also be a sink (in other words, $x_j = 1$ will correspond simultaneously to vertex j being in the connected subgraph solution and to it acting as a sink for the network flow). Finally, we will add constraints to enforce flow conservation: for every vertex the amount of incoming flow equals the amount of outgoing flow plus the amount of consumed flow.

[7] If there are no terminal vertices specified, we add edges from the source to all vertices in the graph and demand that at most one of these edges is used to carry flow.

As mentioned above, for each (directed) edge $(i,j) \in E'$ the variable y_{ij} indicates the amount of flow from i to j. For the source, we introduce a variable $z_0 \in [0,n]$ that serves (1) to introduce flow to the network and (2) to absorb any residual flow. The residual flow plus the flow injected into the network corresponds to the total system flow, as given by Eq. (5), where $\hat{t} \in T$ is arbitrarily chosen. The constraint in Eq. (6) ensures that only vertices included in the solution retain positive incoming flow. In particular, each of the vertices with a positive incoming flow retains one unit of flow, i.e., $(y_{ij} > 0) \Rightarrow (x_j = 1), \forall (i,j) \in E'$. We convert this relation into the flow conservation constraint provided in Eq. (7). All terminal vertices are forced to retain one unit of flow and thus be in the connected subgraph constructed by this process, using Eq. (4). Finally, the overall flow absorbed by the network is set to equal the flow injected into the system, using Eq. (8).

Fig. 1 depicts an example of this network flow representation, where we omit the costs for clarity. Fig. 1(a) presents a graph on 9 vertices with terminal vertices 1 and 9. In Fig. 1(b), a feasible flow for this graph is depicted, originating from the source 0, with value 9. It visits all vertices, while each visited vertex consumes one unit of flow. The thus connected subgraph contains all vertices in this case, including all terminal vertices.

Remark 1. This network flow-based MIP formulation as well as the connected subgraph problem itself allow for the possibility of cycles and loops. We see this as a favorable option given that the overall utility of the parcels selected can be increased by widening the corridor or by incorporating paths to areas of high quality habitat. It may, however, be that the conservation planner wishes to eliminate the possibility of having "peninsulas" in the network, which could represent geographic dead ends to wildlife in the corridor. While this option is not explored empirically in this article, in practice peninsulas could be reduced through the institution of an additional constraint requiring that every vertex receiving flow must output flow to at least one other vertex that is different from the input vertex. Formally, the constraint is

$$y_{ij} \leq n \sum_{\ell \neq i}^{n} y_{j\ell} \quad \forall j \in V \backslash T. \tag{9}$$

Note the use of the multiplier n in the right-hand side of the above constraint, which is needed because the outgoing flow from j would, by design, be one unit less than the incoming flow (when the incoming flow is non-zero) as node j would absorb one unit of flow. While this constraint will eliminate all single parcel wide peninsulas, it is still possible for there to exist a multiple parcel wide peninsula.

3.2. Meeting the scalability challenge: a hybrid solution method

The MIP formulation presented above can be solved to optimality by state-of-the-art MIP solvers, such as IBM/Ilog's CPLEX, for relatively small size problems. Scalability quickly becomes a challenge, however, as one begins to handle real-life data. In order to address the scalability challenge, we use a two-phase solution method.

In Phase I, we compute a minimum cost Steiner tree for the terminal nodes of the graph (i.e., a Steiner tree such that the sum of the costs of the included vertices is the minimum possible), ignoring all utilities. While there are fixed parameter tractable (FPT) algorithms for computing a minimum cost Steiner tree, we implement a simpler "enumeration" method (see, e.g., [28]) based on computing all-pairs-shortest-paths (APSP) with respect to vertex costs, using the Floyd–Warshall Algorithm. The APSP matrix can be computed in time $O(n^3)$ for a graph with n vertices. The matrix is also used to "prune" away a large fraction of the vertices of the graph, as described below. The idea behind the enumeration-based Steiner tree algorithm, which runs in polynomial time for a constant number of terminal nodes, is to first compute a minimum Steiner tree $\tilde{T}$ for the "complete shortest distance graph" $\tilde{G}$ of the original graph G, where $\tilde{G}$ is a complete graph with as many vertices as G and where the weight assigned to an edge $\{u,v\}$ in $\tilde{G}$ equals the cost of a shortest path between the corresponding vertices u and v in G (provided by the APSP matrix). The algorithm uses the fact that in any complete shortest distance graph (such as $\tilde{G}$), there exists a minimum Steiner tree whose non-terminal nodes have degree at least three, thereby limiting the total number of nodes in the Steiner tree to be two fewer than the number of terminal nodes. A minimum Steiner tree $\tilde{T}$ of $\tilde{G}$ yields a minimum Steiner tree T for the original graph G by simply replacing edges $\{u,v\}$ in $\tilde{T}$ by paths in T corresponding to a shortest path between u and v in G.

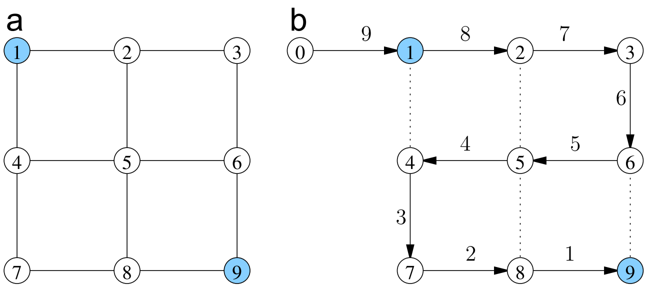

Fig. 1. Flow representation of the connected subgraph problem on a graph with 9 vertices. The terminal vertices, 1 and 9, are shaded. The special terminal node, $\hat{t}$, is vertex 1. (a) Original graph. (b) Feasible flow.

The computation of the Steiner tree in Phase I typically took a few minutes to a few hours on our problem instances, which was in fact an almost negligible amount of time compared to Phase II, which we describe next. The Steiner tree computation either classifies the problem instance as infeasible for the given budget or provides a feasible (but often suboptimal) "mincost" solution. In the latter case, Phase II of the computation translates the problem into a MIP instance using the encoding discussed in Section 3.1, and solves it using IBM/Ilog's CPLEX solver [17]. Solving the MIP formulation using CPLEX is the most computationally intensive part of the whole process. The mincost solution obtained from Phase I is passed on to CPLEX as the starting solution.[8]

Further, the APSP matrix computed in Phase I is also passed on to Phase II. It is used to statically (i.e., at the beginning) *prune* away all nodes that are easily deduced to be too far to be part of a solution (e.g., if the minimum Steiner tree containing that node and all of the terminal vertices already exceeds the budget). This significantly reduces the search space size, often in the range of 40–60%. We also experimented with dynamic pruning, performed during the branch-and-bound search of CPLEX, but found that this did not reduce solution times. The experimental results are reported with static pruning only.

Overall, Phase II is designed to compute an *optimal solution to the utility-maximization version* of the connected subgraph problem. In case it runs out of time, which happened on our large instances, it provides a feasible solution along with a conservative bound on how far this solution is from the optimum (i.e., the *optimality gap*).

As a comparison point we also test a heuristic method, which results in what we call the *extended-mincost solution*. The heuristic proceeds by "freezing" the vertices that form the minimum cost solution to be in the constructed solution. We then solve the MIP encoding of Phase II, described above, using the intended budget with these frozen parcels "forced" to be in the solution. Interestingly, this approach is similar to the two-stage approach used by Önal and Briers [26]. In the first stage the authors aggregate parcels into four-parcel squares and solve for the optimal corridor using the aggregated parcels. They then use only the parcels in the aggregated optimal solution in the second stage to find the optimal disaggregated path. Our two-stage approach is roughly the inverse of their procedure in that we first find the minimum cost solution and then optimally extend this solution. The procedure used by Önal and Briers [26] would not be suitable to our formulation given that our model includes a budget constraint. Therefore the more expensive aggregate set of parcels could never be improved upon in the second stage when the parcels are disaggregated.

We also tested a second heuristic that utilizes the minimum cost corridor as a baseline and simply uses any residual budget to acquire additional vertices in a "greedy" fashion. In other words, we consider those vertices in the graph that are adjacent to the current solution and have cost lower than the residual budget and identify one whose *gain*, defined as the utility-to-cost ratio, is the highest. If there is such a vertex, we add it to the current solution, appropriately reduce the residual budget, and repeat until no more vertices can be added.

In the results section we focus on the comparison of the extended-mincost solution to the optimal solution and do not provide results from application of the greedy heuristic. In general terms, the solution quality (i.e., the overall utility) of the greedy solution lies below that of the extended-mincost solution and the optimal solution for any given budget. The computation of this solution also follows the same trend: the greedy computation is faster than the extended-mincost solution, which itself is faster than the full MIP encoding for computing the optimal solution.

4. Experimental results and an application to a grizzly bear corridor in the U.S. Northern Rockies

While our main goal is to identify the optimal corridor for grizzly bears in the U.S. Northern Rockies, we are also interested in understanding properties of general instances of the connected subgraph problem. To that end, we conducted a series of experiments to study the typical case complexity of the problem. In particular, we investigate the *empirical* computational hardness of the problem with respect to computing the optimal solution or the extended-mincost solution mentioned in Section 3.2, as we vary the budget.

All of our experiments were conducted on a number of 3.8 GHz Intel Xeon machines with 2 GB memory, running Linux 2.6.9-22.ELsmp. We used the CPLEX 10.1 solver [17] to solve the mixed integer programming formulation of the problem instances considered. For the larger instances, which would not fit in the 2 GB RAM of the computers used, we relied on the built-in disk use capabilities of CPLEX (rather than the computer's virtual memory mechanism) to store and manage very large search trees.[9]

4.1. Scaling behavior: semi-structured instances and easy-hard-easy pattern

For the experiments in this and later sections, our main parameter is the feasibility component of the problem, i.e., the budget. Here, for a varying budget level, we investigate the computational hardness of the problem with respect to computing the optimal solution or the extended-mincost solution. In this section, we make use of semi-structured graphs, with uniform random utility and cost functions. The graphs are composed of an $m \times m$ rectangular lattice or grid, in which

[8] In reality, we actually pass on to CPLEX the *greedy solution* to be described shortly. This provides a major boost to the efficiency of CPLEX in solving the MIP encoding.

[9] Specifically, we used the following parameter settings: `cplex.setParam(IloCplex::WorkMem, 1024)` and `cplex.setParam(IloCplex::NodeFileInd, 3)`.

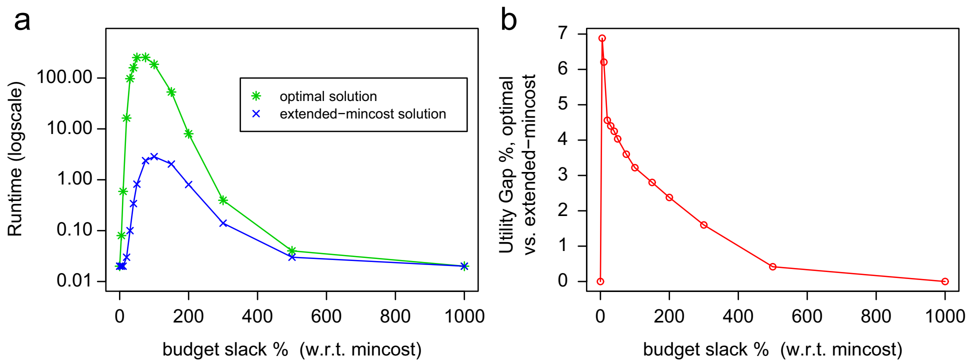

Fig. 2. Results for lattices of order 10 with 3 terminal vertices. (a) Hardness profile (runtime in log-scale). Upper curve: optimal solution. Lower curve: extended-mincost solution. (b) Percentage gap in the utility of optimal and extended-mincost solutions.

we place up to 3 terminal vertices. The details of this semi-structured graph model, as well as more empirical data on them, may be found in Appendix C.

In the figures below, each data point is based on statistics from 100 to 500 randomly generated instances. The hardness curves are represented by median running times over all instances per data point, normalized for the small but non-negligible variation in the characteristics of various randomly generated instances with the same parameters. On the x-axis of the plots is the *budget slack percentage*, rather than simply the budget, computed as follows. For every instance, we consider the *mincost*, i.e., the minimum budget needed to obtain a valid connected subgraph. The *budget slack %* with respect to mincost is defined as: $100 \times (budget - mincost)/mincost$. In other words, we consider computational hardness and other measured quantities as a function of the extra budget available for the problem beyond the minimum required to guarantee a feasible solution. The results are shown in Fig. 2(a) and (b).

In Fig. 2(a), we show the hardness profile of lattices of order 10 with 3 terminals.[10] These optimization problems exhibit an easy-hard-easy pattern, the peak of which is to the right of the mincost point (shown as 0 on the relative x-scale). As one might expect, computing the extended-mincost solution (the lower curve) is significantly easier than computing the true optimal solution (the upper curve; note that the y-axis is in log-scale).

This naturally raises the question: how much "better" are the true optimal solutions compared to the easier-to-find extended-mincost solutions? Fig. 2(b) shows the *relative gap %* between the solution qualities (i.e., attained utilities) in the two cases, defined as $100 \times (optimal - extended)/optimal$. We see that at mincost, both optimal and extended-mincost solutions have similar quality, which is not surprising. The gap between the qualities reaches its maximum shortly thereafter, and then starts to decrease rapidly, so that the extended-mincost solution at 100% budget slack is roughly 3.2% worse than the optimal solution, and at 500% budget slack, only around 0.4% worse. This suggests that for much larger, real-world problems representable as the connected subgraph problem, where computing the optimal solution is out of the question due to limited computational resources, it may suffice for practical purposes to only compute the extended-mincost solution.

4.2. Application to corridor design for U.S. Northern Rockies

4.2.1. Data collection

Study area: The study area for our analysis is comprised of 64 counties in Idaho and western Montana, located in the Northern Continental Divide region. At the most aggregate level, the parcels that we consider for inclusion in the corridor are the 64 counties themselves. While securing an entire county to be included in the reserve may seem infeasible, the county-level analysis provides an illustrative example for a case where the optimization problem is relatively simple from a computational perspective. The county level model allows us to identify general corridor areas that contain low cost, suitable habitat, similar to Ando et al. [1]. The county model also provides a means of comparing the results of an aggregate model with relatively few sites, to more granular models with greater numbers of parcels. A map of the study area is included as Fig. 3.

To investigate the impact of increasing the granularity of the available parcels, we segment the study area into contiguous sets of square grid cells. The largest grid cells that we analyze are 60 km on each side and segment the study area into 118 parcels. The parcel size is then incrementally reduced to square grids with sides of 50 km, 40 km, 25 km, 10 km and 5 km. With the most granular grid size of 5 km, the study area is segmented into 12,788 cells. Given the relatively large range of an adult grizzly (the home range of an adult female grizzly bear is approximately 125 square km), grid sizes smaller than 5 km are unlikely to be suitable for grizzly bear movement [19]. Increasing the granularity of the grid cells allows for much more precision in defining parcel habitat suitability and acquisition costs and it also increases the number of parcels in the landscape. Given the greater number of parcels available for inclusion in the corridor,

[10] We obtained similar results with 10 and 20 terminals as well.

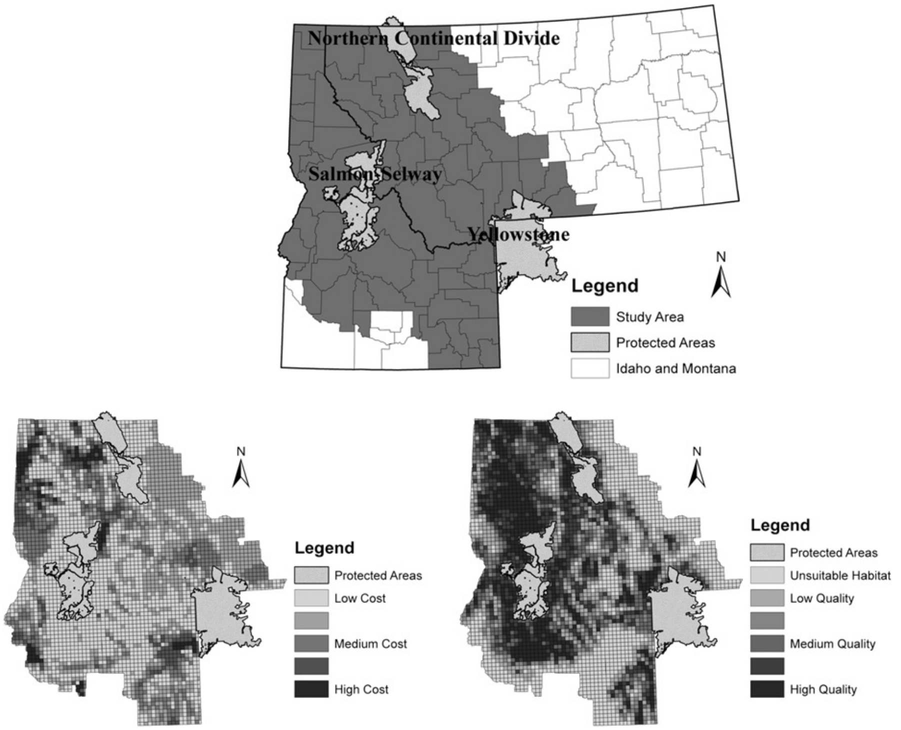

Fig. 3. Top: the study area for the corridor problem. Bottom: cost and utility landscape for 10 km parcels.

increasing the granularity also increases the complexity of the optimization problem. Thus, by comparing results across the continuum of cell sizes, we are able to investigate the tradeoffs inherent in the granularity of the model that allows for increased specificity at the cost of greater computational complexity.

In addition to square grid cells, we also consider a grid composed of 25 square km hexagonal parcels. The hexagonal grid allows parcel connections to occur diagonally and therefore generates more direct pathways between reserves that result in significantly lower costs than comparably sized square grid cells. Hexagonal grids are utilized by the Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) and have been used in reserve site selection research by Polasky et al. [27] as well as Csuti et al. [8].

Utility computation for parcels: To measure the utility of each parcel, we utilize grizzly bear habitat suitability data developed and provided by the Craighead Environmental Research Institute (CERI). These data spatially define habitat that is considered to be suitable for grizzlies. The suitable habitat is measured on a 30 m grid and given a score from 2 to 4, with 4 being the highest quality habitat. We then aggregate the habitat suitability data to the grid and county levels by summing the habitat scores within each parcel boundary. This method of aggregation implicitly assumes, for example, that a cell with a habitat suitability value of 4 is twice as beneficial as a cell with a habitat suitability value of 2.

Cost computation of parcels: We next discuss the process by which cost values are assigned to each land parcel under consideration. The estimate of parcel cost is calculated in three steps. First, spatial data on land stewardship, available for the states of Montana and Idaho from the Gap Analysis Project (GAP) [37], are used to classify privately and publicly owned land in the study area. Next, the amount of private land acreage within each parcel is calculated. The private land acreage is then multiplied by the county-specific average value of farm real estate per acre, available from the United States Department of Agriculture [35]. For grid cells with land acreage in multiple counties, the county-specific real estate value per acre is multiplied by the amount of private acreage in each county and then summed. Following Ando et al. [1], we use the value of farm real estate as a proxy for the cost of all private land, as it reflects the opportunity costs faced by private land owners. Although the value of forest land may in some cases be a more accurate measure of the costs associated with acquiring private land parcels, we do not have the data necessary to accurately estimate location-specific forest values.

In delineating the cost of each parcel, we assume that land already in the public domain is essentially freely available for inclusion in the corridor. One could, however, imagine incorporating the opportunity cost of lost timber or mining contracts as proxies for the cost of acquiring public land as done by Polasky et al. [27] and Sessions [30]. We have chosen not to incorporate costs on public land in the present analysis as there is insufficient data with which to accurately predict the heterogeneity in lost resource profitability associated with each parcel.

By calculating the cost of each parcel based on the real estate value of its privately owned acreage, we are essentially assuming that the parcels included in the corridor will be acquired with fee-simple purchases. For large projects, such as the corridor connecting the three large ecosystems in the Northern Rockies that we are considering, the funds necessary to purchase a viable corridor outright will be large. Yet our cost estimates should be put into perspective by comparison to the significant amount of both public and private funding currently being spent on land conservation.

For example, the federal government has an annual budget of $900 million through the Land and Water Conservation Fund (LWCF), which can support land conservation at the local, state, and federal level. In addition, the Trust for Public Land estimates that in the past decade more than $36 billion in public land conservation funding has been approved in over 1000 separate ballot initiatives across the U.S. [33]. This funding is in addition to federal conservation programs such as the Conservation Reserve Program (CRP), which have average annual expenditures exceeding $1.6 billion [36]. It should also be noted that parcels may not necessarily need to be purchased outright in order to be included in the corridor, as easements and other *voluntary agreements* may be sufficient to maintain habitat. This voluntary type of arrangement is being used, for example, in the ''Alps to Atherton'' project in Australia, where the Australian government is seeking agreements with private land owners to abstain from certain land use practices in exchange for annual payments.

While securing voluntary agreements for habitat protection may be a more viable strategy for cost-effectively targeting parcels to include in the corridor, we do not have accurate estimates of the incentives necessary to secure such voluntary arrangements. We therefore use real-estate value as an upper-bound on a parcel's cost, noting that the potential for voluntary habitat protection could significantly reduce the funds necessary to acquire the corridor. Future research on the incentives necessary for voluntary habitat protection could provide useful information for conservation planners.

One additional consideration in terms of the overall cost of the corridor is the transaction and management costs associated with securing property rights and maintaining the parcels. Researchers have identified transaction and management costs as being an important consideration in reserve design (e.g. [23,24]), yet these costs are rarely included in optimal conservation models. One notable exception is Groeneveld [15] who looks at the theoretical implications of varying transactions costs on the number of sites included in a reserve. In the present analysis, we investigate the influence of transaction costs on corridor design for the 5 km grid parcels. Transaction costs are likely to play a more significant role when the cell granularity is small, as the transaction cost represents a greater proportion of the overall cost of the parcel and the number of potential paths is large. We include a fixed $5000 transaction cost for each parcel that is included, which would cover legal fees, signage and other fees associated with defining a particular land area as part of the corridor. The actual transaction and maintenance cost of a particular parcel is likely to be variable,[11] but we have chosen $5000 simply as an approximation that is in line with reported transaction costs for conservation lands [18].

Parcel adjacencies: Beyond defining the costs and utilities of parcel acquisition, it is also necessary to define the parcel adjacencies. The adjacencies are defined based on shared borders/edges. For the grid parcels this implies that interior parcels are adjacent to exactly four other parcels. This is referred to as a *rook pattern* of adjacency, which is differentiated from a *queen adjacency pattern* where adjacency is defined based on shared edges and corners. For the hexagonal grid parcels, each cell has 6 neighboring parcels regardless of the type of adjacency pattern.

4.2.2. Results for U.S. Northern Rockies

We begin with a study of the effect of parcel granularity on the cost and shape of the resulting wildlife corridor by presenting the results for *minimum cost corridors*, ignoring the utility maximization aspect for now. The minimum cost corridor for each granularity level was computed to optimality using Phase I of our solution methodology based on Steiner Tree computation (cf. Section 3.2).

Fig. 4 visually depicts the maps of the minimum cost wildlife corridors at various granularities. Overall, with the increase in the granularity of the parcels available for acquisition, the minimum cost of a corridor that connects the three ecosystems decreases considerably. For example, the cost of the cheapest corridor is $1.9 billion for the county level and drops to as low as $11.8 million for the 5 km grid, and further down to $7.3 million for the 25 square km hexagonal grid. It is, of course, not surprising that purchasing all of the private land in five counties is extremely expensive. Having the option of buying smaller parcels results in significant cost savings as the corridor is able to better incorporate low cost areas, which are composed primarily, and in some cases exclusively, of zero-cost national forest land.

Changing the parcel granularity not only influences the cost of the parcels selected, but it also influences the general path or *shape* that the corridor follows. For the county level, 60 km, and 50 km parcel granularities, the minimum cost corridor essentially forms the shape of an upside-down T, where the parcels selected are concentrated in the area in the middle of the three ecosystems. When the parcel size is reduced to 40 km and below, the minimum cost corridor traces a path connecting the three reserves that resembles the shape of a C, with the Salmon–Selway Ecosystem connecting directly to the Northern Continental Divide Ecosystem via a parcel path in the northwestern portion of the study area. By increasing the parcel granularity, the model avoids higher priced areas in southwestern Montana and instead chooses a slightly longer corridor that incorporates more national forest land.

In terms of *computational hardness*, as the granularity of the parcels is refined, the problem size and the corresponding search space grows rapidly. For grid parcels of 25 km or larger, the computation time necessary to prove the optimality of the minimum-cost corridor is less than 1 s. For smaller grid sizes, the solution time is no longer trivial, increasing to close to thirty minutes in the case of the 5 km grid and to a couple of hours for the 25 square km hexagonal grid. Thus, we begin

[11] As one referee suggests, transactions costs could also be calculated as a function of the number of distinct private landowners that occupy a given parcel, reflecting the challenges associated with contracting with all necessary parties.

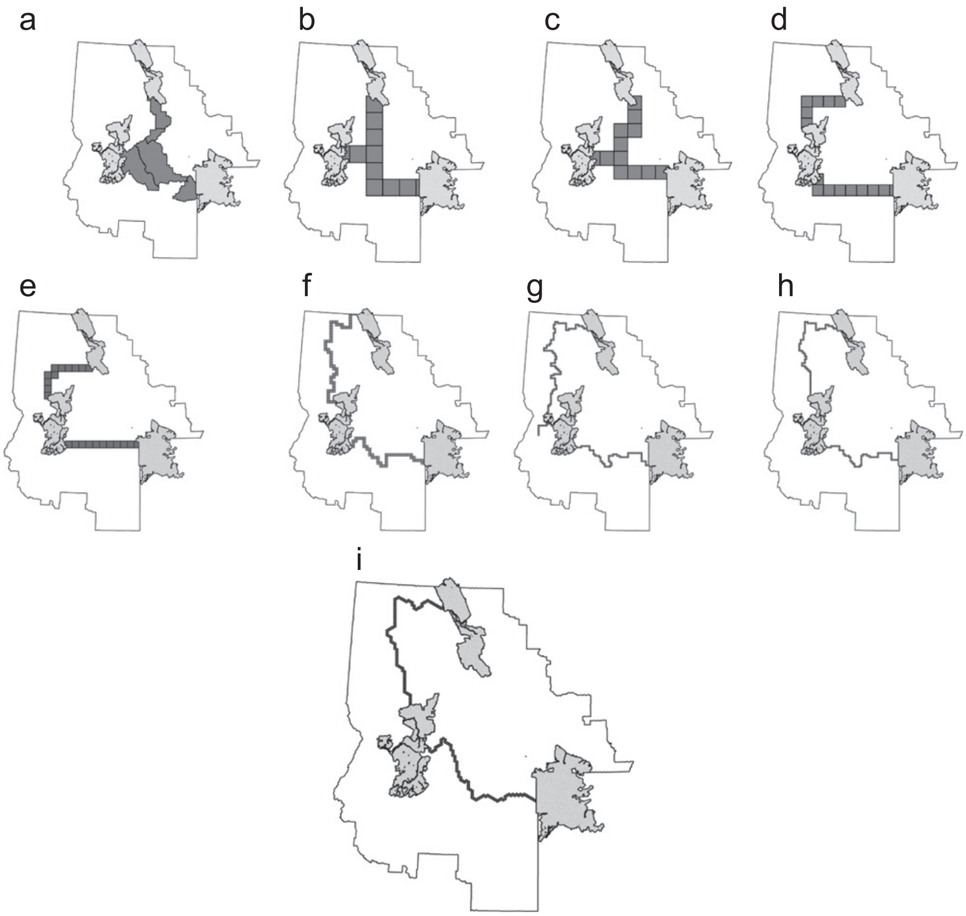

Fig. 4. Minimum cost solutions for the corridor problem at various granularities: (a) County level, (b) 60 km grid, (c) 50 km grid, (d) 40 km grid, (e) 25 km grid, (f) 10 km grid, (g) 5 km grid, (h) 5 km grid with transaction costs, and (i) 25 square km hexagonal grid with transaction costs.

to see the tradeoffs inherent in corridor design in terms of the model granularity, or alternatively the size of the study area, and the computational complexity of the problem.

The impact of transaction costs: The addition of transaction costs to the model also alters the structure of the selected corridor. We again use as an example the minimum cost corridors considered in the discussion above and depicted in Fig. 4. The most noticeable difference of the inclusion of a $5,000 transaction cost per parcel at the 5 km level is that the number of parcels selected is reduced from 265 to 196. Given that each additional parcel adds incrementally to the overall cost of the corridor, even if there is no private land on the parcel, the minimum cost corridor selects parcels that provide more of a direct link between the reserve sites, rather than following a slightly longer path that includes more zero cost, national forest parcels.

This difference is illustrated in panels (g) and (h) of Fig. 4, which show the chosen 5 km corridor both with and without transaction costs. The most noticeable difference between the two corridors is the portion of the corridor connecting the Salmon–Selway to the Northern Continental Divide Ecosystem. With the inclusion of transaction costs, the parcels selected link directly to the northern portion of the Salmon–Selway, rather than the longer path selected without transaction costs that connects to the western edge of the ecosystem. With transaction costs the model also does not select the zero cost parcels that form a peninsula starting from the western edge of the Salmon–Selway. Thus, incorporating transactions costs has a significant influence on both the number and shape of the resulting corridor that is selected and represents an important consideration for land use planners.

Budgets larger than the minimum cost. It is reasonable to expect land use planners to have budgets that are somewhat larger than the precise minimum cost needed for the cheapest wildlife corridor. In such cases a natural question to ask is: *Can neighboring land parcels be acquired which will significantly increase the net utility of the corridor?* Using results from the 25 square km hexagonal grid, we show that a relatively small increase in the budget beyond the minimum cost can often lead to solutions of much higher utility.

Solving the connected subgraph model for budgets larger than the cost minimum in a naïve manner using the CPLEX solver quickly becomes infeasible, especially with various budget levels beyond the basic minimum. Hence, we used the

Table 1
Wildlife corridors with budgets beyond the minimum cost, in the context of 25 square km hexagonal grid with minimum cost = \$7.29 million.

Budget value (unit: \$1 M)	25 square km hex grid corridor				
	Cost (unit: \$1 M)	Parcels		Utility	
		Number	Fractional increase	(1000 ×)	Fractional increase
min	7.29	169	–	1362	–
8.00	7.99	311	1.84 ×	2756	2.02 ×
9.00	9.00	511	3.02 ×	4599	3.38 ×
10.00	10.00	711	4.21 ×	6498	4.77 ×
11.00	11.00	911	5.39 ×	8270	6.07 ×
12.00	12.00	1111	6.57 ×	9973	7.32 ×
15.00	15.00	1708	10.11 ×	14,371	10.55 ×
20.00	20.00	2205	13.05 ×	17,477	12.83 ×
25.00	25.00	2421	14.33 ×	19,068	14.00 ×
50.00	50.00	2837	16.79 ×	22,229	16.32 ×

two-phase solution procedure discussed in Section 3.2 in order to scale up our method. As mentioned earlier, the minimum cost solution obtained from a polynomial time Steiner tree implementation for three reserves was first extended greedily up to the available budget, and this solution was passed on to CPLEX as the starting solution.

Table 1 shows the impact of larger budgets on the utility of the resulting wildlife corridor. Specifically, it reports the utility levels of our *best found solutions* for various budgets.[12] In particular, while the minimum cost corridor (costing \$7.29 million) results in 169 parcels with an overall utility of 1.36 million units, increasing the budget slightly—to \$8 million—results in acquiring 311 parcels with an overall utility of 2.76 million, a more than two-fold increase. By doubling the budget, i.e., going for a budget of \$15 million rather than \$7.29 million, we have an over 10-fold increase in both the number of parcels purchased and the overall utility.

These results suggest an interesting tradeoff between additional budget resources and expanding the wildlife corridor beyond the basic minimum required for achieving connectivity. Instead of focusing all resources on constructing a cost minimizing corridor, conservation planners may be better served by generating a feasible corridor of good quality (i.e., adequate width, suitable habitat, etc.) and then using the remaining budget to acquire nearby land with high net benefits.

Fig. 5 provides a graphical depiction of the results and highlights how the resulting utility rapidly increases as the budget level is initially increased. It also illustrates that the best found solutions for this challenging problem have utility levels *provably* very close to the respective optimal solutions.[13] Specifically, the general trend depicted in the figure illustrates that the benefit-to-cost ratio of the best found corridor slowly flattens out as the budget is increased beyond \$20 million. In other words, while there is a near-linear increase in the attainable utility when increasing the budget from nearly \$7.3 million to \$15 million, there is a significant decline in the rate at which the utility increases when further increasing the budget level.[14] For example, it could perhaps be argued that it is justifiable to invest \$15 million in order to obtain a 10-fold increase in utility and land over the minimum cost solution. At the same time, we also see that even with a budget as large as \$50 million, the best solutions found had a utility of "only" 16-fold that of the minimum cost corridor. Thus, depending on the available budget, land use planners can gain substantial insights from similar cost–benefit tradeoffs of going beyond the cost of the cheapest corridor.

Fig. 6 provides map realizations of model runs with four different budget levels for the 25 square km hex grid. Consider, for example, the solution for a budget of \$15 million and compare it with the minimum cost solution. The \$15 million solution suggests a general trend that by investing roughly twice the money needed for a minimally thin corridor, conservation agencies can greatly expand the amount of natural area acquired. In this particular case, the map suggests that a substantial amount of land near Salmon–Selway (the reserve depicted towards the bottom-left) can be acquired to significantly enhance the overall value of the investment. In addition, the path leading north from this reserve also happens to be very "thick", increasing the value of the corridor itself.

Streamlining as an aid for very large instances: As noted earlier, we employed both a full MIP formulation as well as a restricted or "streamlined" version of it—where we restrict the search to only those corridors that include all of the parcels

[12] As mentioned earlier, we consider two related solution strategies—aiming for the full optimal solution or for the best possible solution that includes all parcels belonging to the minimum cost solution as part of the corridor. The table reports the best solution we found (not necessarily optimal but often provably close to optimal) with either approach for each budget level. A relative comparison of the two approaches is included later in this section.

[13] It should be noted that Pareto-optimality cannot be ensured in this case even with a 0% optimality gap. The intuition for this observation is that given the discrete nature of the problem, there may be a different corridor of lower cost that achieves the same utility as the utility maximizing corridor for a given budget.

[14] As a referee points out, decreasing returns to investment such as what we observe here with corridor design are similar to the textbook example of pollution abatement costs, where firms can initially reduce emissions at fairly low marginal cost but above certain thresholds further abatement becomes increasingly costly.

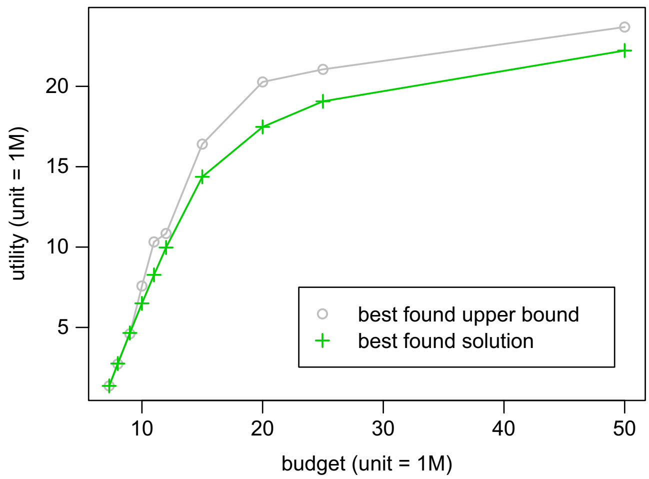

Fig. 5. The best found solutions, along with provable optimality upper bounds, for the 25 square km hexagonal grid, with a 30 day computation cutoff. The upper bound in each case is computed based on the optimality gap reported by CPLEX. The plot shows utility values on the *y*-axis for various budgets on the *x*-axis.

Fig. 6. Corridor maps for the 25 square km hexagonal grid with varying budget levels: minimum cost ($7.289 million), $10 million, $15 million, $20 million.

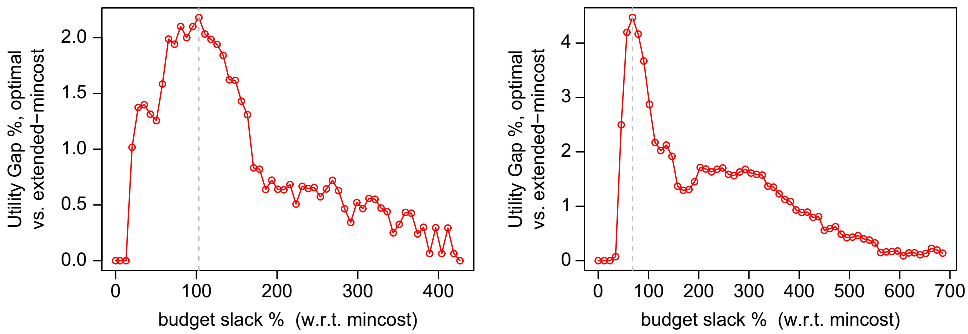

Fig. 7. Percentage gap in the utility of optimal and extended-mincost solutions for the 50 km corridor grid (left) and the 40 km grid (right).

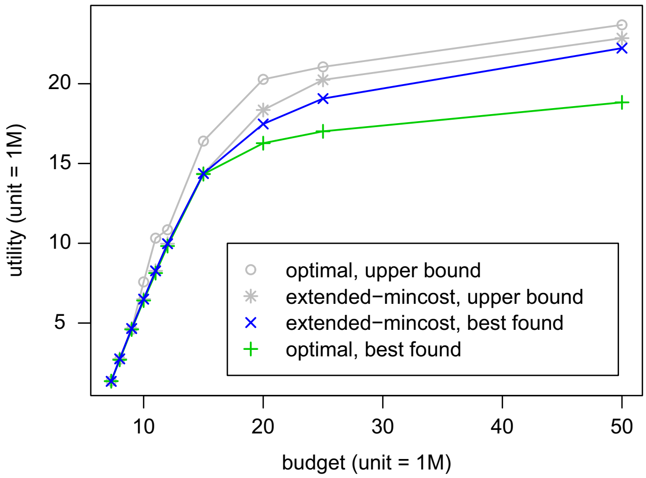

Fig. 8. A comparison between best found solutions for the optimality run vs. the extended-mincost solution runs, for the 25 square km hexagonal grid, with a 30 day cutoff. The upper bound, as before, is computed based on the optimality gap reported by CPLEX. The plot shows utility values on the *y*-axis for various budgets on *x*-axis.

that are in the minimum cost solution found in Phase I. The motivation behind using such "extended-mincost solutions" is computational feasibility—by restricting the search space, we hope to be able to attack larger problem instances than can be solved using the full MIP encoding. Fig. 8 shows that such techniques do often pay off once instances become very large.

For smaller size instances that we can still solve completely to optimality, Fig. 7 shows the relative gap between the utilities obtained with the optimal solution and the extended solution—for the 50 km and 40 km square grid abstractions of the corridor problem, corresponding to Fig. 2(b) discussed earlier for the connected subgraph problem on semi-structured lattice instances. The plots in Fig. 7 show that at its peak, the relative gap is on the order of 2–5%, and is usually within 2% of the optimum. Even a 2% utility gap, however, could imply that one could achieve a particular level of utility for a considerably lower budget when the slope of the relationship between budget and utility is relatively flat. For this problem the slope is relatively steep for budget levels close to the cost minimum and therefore one does not generally lose too much by solving only for the extended solution in this range.

The 25 square km hexagonal abstraction is, as one may expect, very challenging to solve, for both optimal and extended solutions. For instance, while the County level and the 50 km grid abstractions were solved to optimality within seconds with our two-phase solution process, and the 40 km grid took only a few minutes to half an hour depending on the budget, the extended solution itself for the hexagonal grid required several days of computation, and for many budgets, could not be solved optimally in over 10 days. Fortunately, the eventual optimality gap for the best extended solutions found was quite low for budgets up to $15 million, between 0 and 0.07%, meaning that the extended solutions were found to near optimality (the "best found" curve for extended solutions in Fig. 8 is visually right on top of the corresponding "upper bound" curve up to a budget of $15 million). The best found solutions for the true optimality runs, on the other hand, had a similarly low optimality gap for budgets under $10 million but up to a 26.9% gap for higher budgets,[15] as illustrated in Fig. 8.

[15] For a budget of $8 million, our full optimality runs reported an optimality gap of 57% for the best found solution within 30 days of computation. The upper bound for this budget shown in Fig. 8 is based on recent work by Dilkina and Gomes [9] that shows that this solution is actually within 1% of the optimal solution.

Interestingly, for higher budgets (especially $20 million and higher), the best extended-minimum cost solutions found for this challenging grid size were in fact of better quality than the best optimal solutions found, as seen from Fig. 8. This aligns well with the concept of streamlining—if done carefully, restricting the search space to a small but promising part of the full space can result in much better solutions for computationally challenging problems. In this case, restricting the problem to only the extended-minimum cost corridors allowed CPLEX to focus the search and compute better quality solutions than usual within the limited amount of computation time.

Note also in the figure that the best found upper bound was also lower for the extended-minimum cost solutions as compared to the full optimality runs. While this shows that the extended solutions were found to near optimality, it of course does not mean that for the full problem there is necessarily no solution of better quality than this upper bound.

5. Conclusion

Real-world conservation problems are computationally highly demanding. Designing effective conservation corridors is particularly complex due to the intricate combinatorial structure of the problem induced by the spatial connectivity requirement in addition to a strict budget constraint. Solving this problem challenges the scalability limits of current computational optimization methods. Unlike many other conservation problems for which a marginal change in the available budget affects the resulting solution only marginally, the corridor design problem and other spatially dependent selection problems are unique in the sense that a *marginal change in the available budget can result in the selection of very different, potentially mutually exclusive, sets of parcels*.

In this work, we have developed algorithms to solve large-scale corridor design problems, considering the minimum cost corridor but also going beyond the minimum cost solution to the best use of resources when a conservation planner has at his or her disposal a somewhat larger budget. Our empirical investigation into the general properties of the problem revealed the difficulty of solving it to optimality and an interesting easy-hard-easy pattern as a key problem parameter—the amount of extra budget beyond the minimum cost—is slowly increased. We presented a case study for a real-world instance of a corridor that would provide a link between the Yellowstone, Salmon–Selway, and the Northern Continental Divide ecosystems in Idaho, Wyoming, and Montana. Our study explores the implementation of such a corridor at various budget levels, showing, for example, that with a budget level of roughly twice the minimum cost, we can achieve over a 10-fold increase in both the number of parcels purchased and the overall utility.

Despite the evaluation of our method on a particular data set for the grizzly bear mentioned above, the methods and models developed here are general. They could be applied to other cost and utility distributions beyond the ones used in our experiments, or to designing effective corridors for other endangered species. More generally, the techniques introduced here are applicable to any problem domain that can be re-formulated as the connected subgraph problem, such as in the context of social networks.

Appendix A. Illustrative example

Example 1. We consider a simple example to illustrate some of the combinatorial issues of the connected subgraph problem that make it computationally hard. Consider the hypothetical 3×3 parcel map presented in Fig. A1. We use this map to illustrate, among other aspects, how both the optimal choice of parcels and the complexity of finding these parcels can vary dramatically when we formulate the problem as a cost constrained utility optimization problem rather than an unconstrained cost minimization problem. In this simple example, when ignoring utilities, the cost of the corridor is minimized with the selection of parcels B, E, and H, as shown in panel I. With this selection, the cost is 7 units and the utility of the parcels selected is 5. Now suppose that the conservation planner has available a budget of 10 units. Rather than simply selecting the least cost path consisting of parcels B, E, and H, the planner would now be interested in finding the corridor that yields the highest possible utility, with a cost of no more than 10 units. In panel II, we show that for a budget of 10 units, the planner maximizes utility by selecting parcels E, F, H, and I, for a total utility of 9 units. If the conservation planner's budget is increased to 11 units, as in panel III, the optimal selection of parcels is A, B, D, with a corresponding aggregate utility of 10 units.

It is not surprising that considering only parcel costs as in panel I results in a very different set of selected parcels from that in panels II and III, where both parcel cost and utility are considered. What is unique about the constrained corridor optimization problem is that a *marginal change in the available budget can result in the selection of very different, potentially mutually exclusive, sets of parcels*, as illustrated in panels II and III. Given the constraint that all of the selected parcels must be connected, the model outcomes can change drastically as budget levels are varied, which is different from typical reserve site selection models where additional budget levels generally only influence the selection of a small subset of the available parcels.

Fig. A1 also illustrates the computational challenges of the budget constrained utility maximization problem. If the goal is to find a least cost path, as has been done in most previous studies, only six possible paths in the 3×3 parcel grid need to be considered. The optimal selection will never include paths that are more than one parcel wide, as this can only add to the cost of the corridor. For the case of constrained utility maximization, however, the set of feasible corridors jumps from

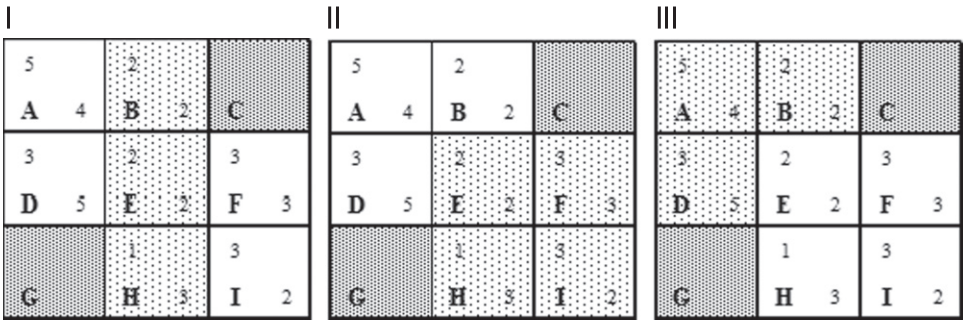

Fig. A1. Hypothetical corridor optimization. Parcel labels are provided in the lower left corner of each parcel, costs are in the lower right corner, and utilities are in the upper left corner. The reserves, G and C, are marked as dark gray. The optimal corridor in each case is shaded. I. Minimum cost corridor, cost = 7, utility = 5; II. Budget = 10, cost = 10, utility = 9 and III. Budget = 11, cost = 11, utility = 10.

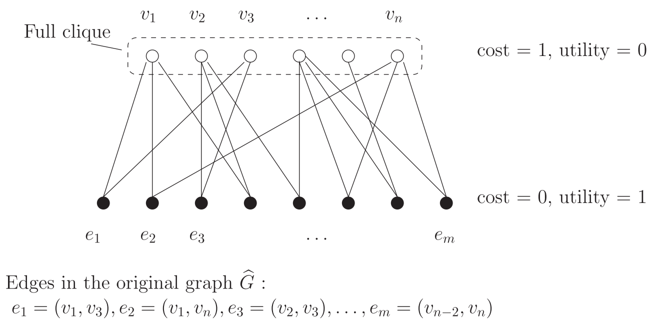

Fig. B1. Reduction from vertex cover.

6 to 30. Thus, even in this small hypothetical case, the challenge of maximizing utility given a budget constraint is considerably greater than simply finding the single-parcel-wide least cost path.

Appendix B. Proof details

Proof of Theorem 1. The problem is clearly in NP, because a certificate subgraph H can be easily verified to have the desired properties, namely, connectedness, low enough cost, and high enough utility. For NP-hardness, consider the Steiner tree problem on a graph $\widehat{G} = (\widehat{V}, \widehat{E})$ with terminal set $\widehat{T} \subseteq \widehat{V}$, edge cost function $\widehat{c} : \widehat{E} \to \mathbf{Z}^+$, and cost bound $\widehat{C}$.

An instance of the connected subgraph problem can be constructed from this as follows. Construct a graph $G = (V,E)$ with $V = \widehat{V} \cup \widehat{E}$ and edges defined as follows. For every edge $e = \{v,w\} \in \widehat{E}$, create edges $\{v,e\},\{w,e\} \in E$. The terminal set remains the same: $T = \widehat{T}$. Overall, $|V| = |\widehat{V}| + |\widehat{E}|$, $|E| = 2|\widehat{E}|$, and $|T| = |\widehat{T}|$. For costs, set $c(v) = 0$ for $v \in \widehat{V}$ and $c(e) = \widehat{c}(e)$. For utilities, set $u(v) = 1$ for $v \in T$ and $u(v) = 0$ for $v \notin T$. Finally, the cost bound for the connected subgraph is $C = \widehat{C}$ and the utility bound is $U = |E|$.

It is easy to verify that the Steiner tree problem on $\widehat{G}$ and $\widehat{T}$ has a solution with cost at most C iff the connected subgraph problem on G and T has a solution with cost at most C and utility at least U. This completes the reduction.

Note that if $\widehat{G}$ is planar, then so is G. Further, the reduction is oblivious to the number of terminals in G. Hence, NP-completeness holds even on planar graphs and without any terminals. □

Proof of Lemma 1. We give a reduction along the lines of the one given by Bern and Plassmann [3] for the Steiner tree problem. The reduction is oblivious to the number of terminals and holds in particular even when there are no terminals.

Recall that a vertex cover of a graph $\widehat{G} = (\widehat{V}, \widehat{E})$ is a set of vertices $V' \subseteq \widehat{V}$ such that for every edge $\{v,w\} \in \widehat{E}$, at least one of v and w is in V'. The vertex cover problem is to determine whether, given $\widehat{G}$ and $C \geq 0$, there exists a vertex cover V' of $\widehat{G}$ with $|V'| \leq C$. We convert this into an instance of the connected subgraph problem. An example of such a graph is depicted in Fig. B1.

Create a graph $G = (V,E)$ with $V = \widehat{V} \cup \widehat{E}$ and edges defined as follows. For every $v,w \in \widehat{V}, v \neq w$, create edge $\{v,w\} \in E$; for every $e = \{v,w\} \in \widehat{E}$, create edges $\{v,e\},\{w,e\} \in E$. Overall, G has $|\widehat{V}| + |\widehat{E}|$ vertices and $\binom{V}{2} + 2\widehat{E}$ edges. For costs, set $c(v)$ to be 1 if $v \in \widehat{V}$, and 0 otherwise. For utilities, set $u(e)$ to be 1 if $e \in \widehat{E}$, and 0 otherwise. Finally, fix the set of terminals to be an arbitrary subset of $\widehat{E}$.

We prove that solutions to the connected subgraph problem on G with costs and utilities as above, cost bound C, and desired utility $U = |\widehat{E}|$ are in one-to-one correspondence with vertex covers of $\widehat{G}$ of size at most C.

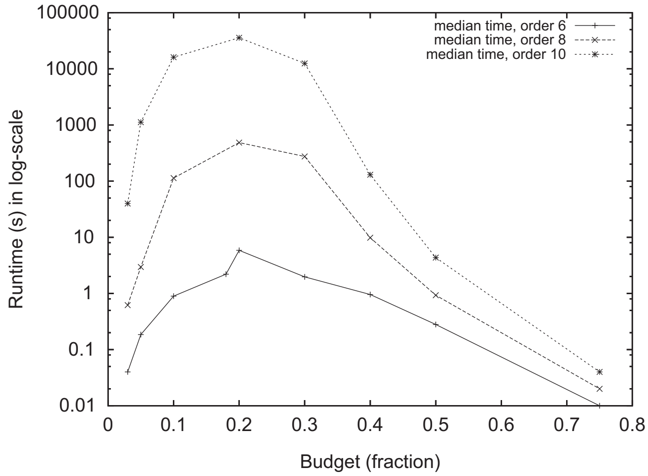

Fig. C1. Hardness profile for lattices of order 6, 8, and 10, without terminal vertices.

First, let vertex-induced subgraph H of G be a solution to the connected subgraph instance. Let $V' = V(H) \cap \widehat{V}$. We claim that V' is a vertex cover of $\widehat{G}$ of size at most C. Clearly, $|V'| \le C$ because of the cost constraint on H. To see that V' is indeed a vertex cover of $\widehat{G}$, note that (A) because of the utility constraint, V' must contain *all* of the vertices from $\widehat{E}$, and (B) because of the connectedness constraint, every such vertex must have at least one edge in $E(H)$, i.e., for each $e = \{v,w\} \in \widehat{E}$, V' must include at least one of v and w.

Conversely, let V' be a vertex cover of $\widehat{G}$ with at most C vertices. This directly yields a solution H of the connected subgraph problem: let H be the subgraph of G induced by vertices $V' \cup \widehat{E}$. By construction, H has the same cost as V' (in particular, at most C) and has utility exactly U. Since V' is a vertex cover, for every edge $e = \{v,w\} \in \widehat{E}$, at least one of v and w must be in V', which implies that H must have at least one edge involving e and a vertex in V'. From this, and the fact that all vertices of V' already form a clique in H, it follows that H itself is connected.

This settles our claim that solutions to the two problem instances are in one-to-one correspondence, and finishes the proof. □

Appendix C. Computational hardness profiles for synthetic data

We make use of semi-structured graphs, with uniform random utility and cost functions. The graphs are composed of an $m \times m$ rectangular lattice or grid, where the order m is either 6, 8, or 10. This lattice graph is motivated by the structure of the original conservation corridors problem. In this lattice, we place k terminal vertices; in the results reported here, k is 0 or 3. When $k \ge 2$, we place two terminal vertices in the "upper left" and "lower right" corners of the lattice, so as to maximize the distance between them and "cover" most of the graph. This is done to avoid the occurrence of too many pathological cases, where most of the graph does not play any role in constructing an optimal connected subgraph. The remaining $k-2$ terminal vertices are placed uniformly at random in the graph. To define the utility and cost functions, we assign uniformly and independently at random a utility and a cost from the set $\{1, 2, \ldots, 10\}$ to each vertex in the graph. The cost and utility functions are uncorrelated.

Each data point is based on statistics from over 100–500 randomly generated instances. Fig. C1 shows computational hardness results for the case of zero reserves. Note that these instances are always feasible, even with zero budget. Using the budget slack percentage relative to mincost as in Fig. 2(a) and (b) earlier, therefore, does not make sense in this case. We simply use for the x-axis the fraction *budget*/*total-budget*, where *total-budget* is the total cost of all vertices.

These connected subgraph instances, defined on graphs without terminal vertices, are always satisfiable and can thus be seen as instances of pure optimization problems. Fig. C1 shows the hardness profile (i.e., the running time) on lattices of order 6, 8, and 10. Notice that the median runtime (y-axis) is plotted in log-scale in this figure. The plots clearly indicate an easy-hard-easy pattern for these instances, even though they are all feasible with respect to the budget. Such patterns have been observed previously in some pure optimization problems, but only under specialized random distributions. For example, Zhang and Korf [42] identify a similar pattern for the Traveling Salesperson Problem, using a log-normal distribution of the distance function. In our case, the pattern naturally emerges from the model.

References

[1] A. Ando, J. Camm, S. Polasky, A. Solow, Spatial distributions, land values, and efficient conservation, Science 279 (1998) 2126–2128.

[2] H.P. Andreassen, S. Halle, R.A. Ims, Optimal width of movement corridors for root voles: not too narrow and not too wide, Journal of Applied Ecology 33 (1996) 63–70.

[3] M.W. Bern, P.E. Plassmann, The Steiner tree problem with edge lengths 1 and 2, Information Processing Letters 32 (4) (1989) 171–176.
[4] J.D. Camm, S. Polasky, A.R. Solow, B. Csuti, A note on optimal algorithms for reserve site selection, Biological Conservation 78 (3) (1996) 353–355.
[5] J.O. Cerdeira, K.J. Gaston, L.S. Pinto, Connectivity in priority area selection for conservation, Environmental Modeling and Assessment 10 (3) (2005) 183–192.
[6] R.L. Church, D.M. Stoms, F.W. Davis, Reserve selection as a maximal covering location problem, Biological Conservation 76 (2) (1996) 105–112.
[7] C. Costello, S. Polasky, Dynamic reserve site selection, Resource and Energy Economics 26 (2) (2004) 157–174.
[8] B. Csuti, S. Polasky, P. Williams, R. Pressey, J. Camm, M. Kershaw, R. Kiester, B. Downs, R. Hamilton, M. Huso, K. Sahr, A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon, Biological Conservation 80 (1) (1997) 83–97.
[9] B. Dilkina, C.P. Gomes, Solving connected subgraph problems in wildlife conservation, in: CPAIOR-10: 7th International Conference on Integration of AI and OR Techniques in Constraint Programming, Lecture Notes in Computer Science, vol. 6140, Bologna, Italy, June 2010, pp. 102–116.
[10] I. Dinur, S. Safra, On the hardness of approximating minimum vertex cover, Annals of Mathematics 162 (1) (2005) 439–486.
[11] J.D. Dixon, M.K. Oli, M.C. Wooten, T.H. Eason, J.W. McCown, D. Paetkau, Effectiveness of a regional corridor in connecting two Florida black bear populations, Conservation Biology 20 (1) (2006) 155–162.
[12] C. Faloutsos, K. McCurley, A. Tomkins, Fast discovery of connection subgraphs, in: Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, 2004, pp. 118–127.
[13] P.J. Ferraro, Conservation contracting in heterogeneous landscapes: an application to watershed protection with threshold constraints, Agricultural and Resource Economics Review 32 (1) (2003) 53–64.
[14] T. Fuller, M. Munguía, M. Mayfield, V. Sánchez-Cordero, S. Sarkar, Incorporating connectivity into conservation planning: a multi-criteria case study from central Mexico, Biological Conservation 133 (2) (2006) 131–142.
[15] R. Groeneveld, Economic considerations in the optimal size and number of reserve sites, Ecological Economics 52 (2005) 219–228.
[16] N.M. Haddad, D.R. Bowne, A. Cunningham, B.J. Danielson, D.J. Levey, S. Sargent, T. Spira, Corridor use by diverse taxa, Ecology 84 (3) (2003) 609–615.
[17] ILOG, SA. CPLEX 10.1 Reference Manual, 2006.
[18] Land Trust Alliance. New York State Conservation Partnership Program 2003 awards, 2004. Available at 〈http://www.lta.org/yourregion/ny/nyscpp_2003_awards.htm〉.
[19] R.K. Mace, J. Waller, Final Report: Grizzly Bear Ecology in the Swan Mountains. Technical Report, Montana Department of Fish, Wildlife and Parks, 1997.
[20] C.R. Margules, A.O. Nicholls, R.L. Pressey, Selecting networks of reserves to maximise biological diversity, Biological Conservation 43 (1) (1988) 63–76.
[21] A. McEuen, The wildlife corridor controversy: a review, Endangered Species Update (1993) 1–6.
[22] K.D. Messer, The conservation benefits of cost-effective land acquisition: a case study in Maryland, Journal of Environmental Management 79 (3) (2006) 305–315.
[23] R. Naidoo, A. Balmford, P.J. Ferraro, S. Polasky, T.H. Ricketts, M. Rouget, Integrating economic costs into conservation planning, Trends in Ecology and Evolution 21 (12) (2006) 681–687.
[24] D.A. Newburn, P. Berck, A.M. Merenlender, Habitat and open space at risk of land-use conversion: targeting strategies for land conservation, American Journal of Agricultural Economics 88 (1) (2006) 28–42.
[25] R.F. Noss, Corridors in real landscapes: a reply to Simberloff and Cox, Conservation Biology 1 (2) (1987) 159–164.
[26] H. Önal, R.A. Briers, Optimal selection of a connected reserve network, Operations Research 54 (2) (2006) 379–388.
[27] S. Polasky, J.D. Camm, B. Garber-Yonts, Selecting biological reserves cost-effectively: an application to terrestrial vertebrate conservation in Oregon, Land Economics 77 (1) (2001) 68–78.
[28] H.J. Prömel, A. Steger, The Steiner Tree Problem: A Tour Through Graphs, Algorithms, and Complexity. Vieweg, 2002.
[29] T. Schmitt, A. Seitz, Influence of habitat fragmentation on the genetic structure of *Polyommatus Coridon* (Lepidoptera: Lycaenidae): Implications for conservation, Biological Conservation 107 (3) (2002) 291–297.
[30] J. Sessions, Solving for habitat connections as a Steiner network problem, Forest Science 38 (1) (1992) 203–207.
[31] J.J. Tewksbury, D.J. Levey, N.M. Haddad, S. Sargent, J.L. Orrock, A. Weldon, B.J. Danielson, J. Brinkerhoff, E.I. Damschen, P. Townsend, Corridors affect plants, animals, and their interactions in fragmented landscapes, Proceedings of the National Academy of Science 99 (2) (2002) 12923–12926.
[32] S.F. Tóth, R.G. Haight, S.A. Snyder, S. George, J.R. Miller, M.S. Gregory, A.M. Skibbe, Reserve selection with minimum contiguous area restrictions: an application to open space protection planning in suburban chicago, Biological Conservation 142 (8) (2006) 1617–1627.
[33] Trust for Public Land. TPL landvote database, 2010. Available at 〈https://www.quickbase.com/db/bbqna2qct?a=dbpage&pageID=10〉.
[34] L.G. Underhill, Optimal and suboptimal reserve selection algorithms, Biological Conservation 70 (1) (1994) 85–87.
[35] United States Department of Agriculture. Census of Agriculture, 2002.
[36] United States Department of Agriculture. Conservation Reserve Program: Annual Summary, 2002.
[37] United States Geologic Survey. Gap Analysis—Land Stewardship Coverage, 1999. Available at 〈http://gapanalysis.nbii.gov〉.
[38] J.C. Williams, Delineating protected wildlife corridors with multi-objective programming, Environmental Modeling and Assessment 3 (1–2) (1998) 77–86.
[39] J.C. Williams, A zero-one programming model for contiguous land acquisition, Geographical Analysis 34 (4) (2002) 330–349.
[40] J.C. Williams, C.S. ReVelle, S.A. Levin, Spatial attributes and reserve design models: a review, Environmental Modeling and Assessment 10 (3) (2005) 163–181.
[41] J.C. Williams, S.A. Snyder, Restoring habitat corridors in fragmented landscapes using optimization and percolation models, Environmental Modeling and Assessment 10 (3) (2005) 239–250.
[42] W. Zhang, R.E. Korf, A study of complexity transitions on the asymmetric traveling salesman problem, Artificial Intelligence 81 (1996) 223–239.