Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence

Large Landscape Conservation — Synthetic and Real-World Datasets

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

Ashish Sabharwal
IBM Research
Yorktown Heights, NY
ashish.sabharwal@us.ibm.com

Jordan Suter
Oberin College
Oberlin, OH
jordan.suter@oberlin.edu

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

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

Abstract

Biodiversity underpins ecosystem goods and services and hence protecting it is key to achieving sustainability. However, the persistence of many species is threatened by habitat loss and fragmentation due to human land use and climate change. Conservation efforts are implemented under very limited economic resources, and therefore designing scalable, cost-efficient and systematic approaches for conservation planning is an important and challenging computational task. In particular, preserving landscape connectivity between good habitat has become a key conservation priority in recent years. We give an overview of landscape connectivity conservation and some of the underlying graph-theoretic optimization problems. We present a synthetic generator capable of creating families of randomized structured problems, capturing the essential features of real-world instances but allowing for a thorough typical-case performance evaluation of different solution methods. We also present two large-scale real-world datasets, including economic data on land cost, and species data for grizzly bears, wolverines and lynx.

1 Landscape Connectivity

The resilience of animals to disturbance events and their accommodation to long-term ecosystem adaptations (e.g., climate change) depends heavily on their ability to move safely throughout the environment to find food, reproduce, and migrate between habitat patches (Taylor et al. 1993).Hence, loss of connectivity can lead to population declines, loss of genetic variation, and ultimately species extinction. Yet, the pressures of both human development and climate change have resulted in a drastic habitat loss and fragmentation. Therefore, preserving and restoring habitat connectivity has been identified as a key conservation priority for government agencies and conservation organizations, in order to offset the impacts of habitat loss and fragmentation on biodiversity conservation, and to increase the resilience of reserve networks to potential threats (Kareiva 2006). In particular, the long-term value of existing conservation areas relies on maintaining connections to other intact areas. The term *landscape connectivity*, used by ecologists, refers to the extent to which a landscape facilitates the movements of organisms and their genes, and is an important factor in evaluating the well-being of endangered species. The importance of landscape connectivity has been highlighted in the latest installment of *Issues in Ecology* published by the Ecological Society of America (Rudnick et al. 2012).

Considerable ecology research has concentrated on creating models of landscape permeability or conversely of landscape resistance. The landscape is represented as a set of small parcels or pixels, each of which has a resistance value that gives the species-specific cost of moving through particular landscape features. The resistance surface depends on both the focal species and the actual landscape characteristics. Resistance models are inferred by, e.g., relating landscape characteristics to genetic distance between individuals at different locations or to radio-collar movement data. The connectivity between two habitat patches is measured in terms of their resistance distance. Two widely-used connectivity frameworks are Least-Cost Path (LCP) modeling (Singleton, Gaines, and Lehmkuhl 2002) and Circuitscape (McRae et al. 2008). Under the LCP model, the connectivity of two habitat patches is measured as the length of the shortest resistance-weighted path between them. The Circuitscape model interprets the graph as a network of resistors and measures connectivity between two nodes as the effective resistance between the two using Kirchhoff's laws. The Circuitscape model is complementary to the LCP analysis since it incorporates both the minimum movement distance, as well as the availability of alternative pathways. Both the LCP and Circuitscape models have been used in numerous studies to measure the level of connectivity between core habitat areas in different study areas, and to identify the parts of the landscape that are most likely to contribute to that connectivity, i.e. to serve as paths or corridors for animal dispersal.

While conservation mandates and plans are often determined solely on the basis of biological and ecological needs of species, several recent studies have shown that it is crucial to incorporate both economic and biodiversity considerations from the outset of the planning process in order to design conservation strategies that are efficient and practical (Naidoo et al. 2006; Joseph, Maloney, and Possingham 2009). Decision-support tools to design efficient budget-constrained conservation strategies are therefore needed and yet are still largely lacking. The underlying computational problems to address landscape connectivity conservation fall into the challenging computational domains of combinatorial optimization and network design, with interesting research questions for computer and operations research scientists.

Developing models for landscape connectivity that incorporate economic considerations requires three types of inputs: the locations of habitat areas whose connectivity needs to conserved; the species-specific resistance or utility of the landscape; and the conservation cost of land. Recently, several different optimization models (Conrad et al. 2007; Dilkina and Gomes 2010; Conrad et al. 2012; Lai et al. 2011; Dilkina, Lai, and Gomes 2011; Le Bras et al. 2013) were proposed to study conservation planning for landscape connectivity in the context of the new field of Computational Sustainability (Gomes 2009). These computational approaches enable decision makers to perform a systematic study of the tradeoffs between economic costs and ecological benefits for the conservation settings addressed. However, further computational advances are possible and needed. To facilitate research on this topic, we present a synthetic generator that produces the three types of inputs describing a conservation planning instance, and captures some of the inherent spatial structure of real datasets. Due to its randomized nature, the synthetic generator allows for the creation of families of instances that can be used to characterize the typical-case behavior of different solution approaches—something that cannot be evaluated using only one or two real-world problem instances. We also publicly release both the numeric and geographical data of two case studies used in previous work, which provide a testbed with realistic spatial distribution of resistances and land costs. *The data associated with these instances were compiled strictly for evaluation of solution methodologies and should not be used for actual conservation recommendations*. The generator and datasets can be found online[1].

2 Optimization Problems

One method to mitigate fragmentation and maintain landscape connectivity is to set aside so-called wildlife corridors, or swaths of preserved land that connect important patches of habitat for the endangered species.

(Conrad et al. 2007; Dilkina and Gomes 2010; Conrad et al. 2012) studied the following conservation planning problem. Given a landscape divided into spatially-explicit planning units or pixels associated with economic costs and species utility, as well as a set of core habitat areas, the goal is to design a conservation strategy that purchases a set of planning units with total cost within a specific budget constraint such that they form a connected network between the core habitat areas and maximize the total utility of the protected land. This planning setting was formalized as the *Connection Subgraph Problem* and several solution methodologies were developed.

When considering one species, finding the set of planning units which connect the core areas at a minimum cost corresponds to the well-known *Minimum Steiner Tree* problem, solvable in polynomial time for a fixed number of core areas. (Lai et al. 2011) studied the minimum cost corridor design problem for multiple species, where the landscape is differentially permeable for each specific species.The authors show that in this more complex setting the problem is computationally harder.

Most recently, (Le Bras et al. 2013) consider another important aspect of conservation planning for landscape connectivity, addressing the fact that conservation plans need to be robust to changes in the landscape, for example due to climate change (McKelvey et al. 2011) or unexpected events such as wildfires. One way to achieve robustness is by protecting multiple mutually non-overalapping corridors between pairs of core areas. The authors provide both optimal as well as fast heuristic approaches to find budget-constrained robust plans that minimize resistance.

(Dilkina, Lai, and Gomes 2011) addressed a different conservation planning problem concerning landscape connectivity. Instead of conserving explicit whole corridors, in this setting each land parcel may contribute to the connectivity of the core habitat areas, whether or not it has been bought (measured using the LCP model). Each parcel is associated with two resistance values, one lower and one higher, corresponding to the state of the parcel with and without conservation respectively. One can interpret the benefit of buying a piece of land as decreasing the lands effective resistance through management, or alternatively protecting the current resistance of the parcel which otherwise increases due to human development and land use change. Given a set of pairs of core areas, the resistance values with and without conservation, land costs and a budget, the goal is to choose which parcels to protect so that the final effective resistance-weighted distance between the pairs is minimized.

A related but very challenging task is to design techniques that allow for conservation planning over time, or adaptive management, in order to consider, e.g. strategies adapting to landscape changes over time, as well as limited budgets potentially not all available upfront but in different time periods. Additionally, one can consider optimization strategies for conservation planning with a different measure of connectivity, e.g. using Circuitscape.

3 Synthetic Dataset Generator

In order to perform typical-case analysis of optimization approaches for landscape connectivity planning, we created a synthetic dataset generator that captures the characteristics of landscape connectivity conservation planning problems, including multi-species conservation. The generator creates a landscape consisting of an NxN grid or matrix of cells. To generate the core habitat areas, it first splits the NxN grid of cells into SxS grid of blocks (each block is $\lfloor$N/S$\rfloor$-by-$\lfloor$N/S$\rfloor$ square of cells). The *outskirt* blocks are defined as those in the out-most laterals of the square.

Independently for each species, we first generate core areas and then a resistance matrix. Two core areas are created by selecting one pair of symmetric S-by-S blocks, by picking a random SxS block (x,y) among the upper outskirt blocks and then also picking the block (S-x, S-y), in order to make their locations as far as possible. In each selected

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

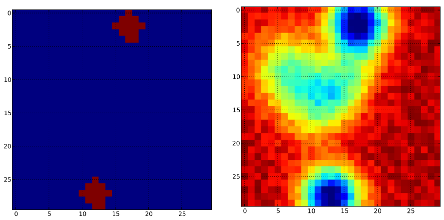

Figure 1: (Left) Core areas in dark red (Right) Resistance Matrix: Blue is low resistance, red is high resistance. The areas corresponding to core areas have low resistance, while the additional low resistance areas correspond to two extra randomly placed Gaussian functions.

SxS block, we create a core area consisting of m contiguous parcels within the block. N, S, and m are parameters. The NxN resistance matrix is generated as a mixture of 2D Gaussian functions. A 2D Gaussian function is characterized by a 5-tuple, $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$, where μ_x, μ_y are the means and σ_x, σ_y are the standard deviations in the x and y dimensions and ρ is their correlation. First, to ensure low resistance values within core areas, we place a negated Gaussian function with mean vector at the center point of each core area, with $\rho = 0$ and $\sigma_x = \sigma_y = \sigma_{center}$, where σ_{center} is a parameter . After that, we create other $nExtra$ negated Gaussian functions with mean vectors located at randomly sampled cells from the non-outskirt blocks, and with randomized parameters $\hat{\sigma}_{center}$ and $\hat{\rho}$ in intervals around σ_{center} and ρ. The extra Gaussian functions are used to create "low resistance" areas outside core areas allowing animals to migrate between habitats. We sum the Gaussian functions placed at the core areas and the extra Gaussian functions (multiplied by 0.5 to reduce their influence) to obtain a resistance value for each cell. Finally, the resistance at each cell is perturbed with uniformly distributed noise, shifted, and normalized to fall within a desired range $[0, Rmax - 1]$. An example of core areas and a resistance matrix generated by this process is shown in Fig. 1. $Rmax$, $nExtra$, and σ_{center} are parameters. This generation process is repeated for each species.

Finally, the land cost values are generated in a way that is positively correlated with the resistance values. This is motivated by the fact that land cost is high in developed areas, which are usually associated with high species resistances, and vice versa. In addition, cells that belong to core areas are assigned 0 cost. An example of a cost matrix for an instance with two species is presented in Fig. 2. A full detailed description of the synthetic generator can be found online.

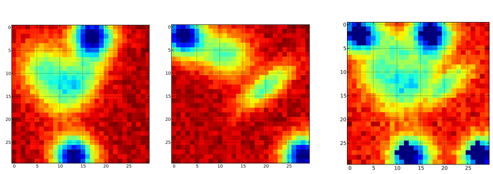

Figure 2: (Left and Center) Resistances for 2 species (Right) Cost matrix: Highly correlated with combined resistances.

4 Real-world datasets

In addition to the ability to create synthetic landscapes using the described generator, we have made available real world datasets for grizzly bears (Ursus arctos), Wolverine (Gulo gulo) and Canada Lynx (Lynx canadensis).

Grizzly Bear Dataset in the Rockies: This dataset concerns grizzly bear corridor design (see Fig. 3). The goal is to ensure connectivity between three major national conservation parks with existing grizzly populations, the Yellowstone, SalmonSelway and Northern Continental Divide Ecosystems in Idaho, Wyoming and Montana. The data include habitat suitability, or utility, values for grizzly bears and land costs for different parcels in the geographical area surrounding the three wildlife reserves. The utility of each parcel was computed using additive aggregation over grizzly bear habitat suitability data at 30m resolution provided by the Craighead Environmental Research Institute. Land costs were computed by multiplying the amount of privately owned acreage in each parcel by the county-specific average value of farm real estate per acre, available from USDA. This dataset was used in (Conrad et al. 2007; Dilkina and Gomes 2010) and a detailed description and study is provided in (Conrad et al. 2012). Full description of the dataset is provided with the online distribution. The data are provided as four datasets at resolutions of 40km (242 cells), 10km (3,299 cells) and 5km (12,788 cells) grid cells, and 25km^2 (12,889) hexagonal cells.

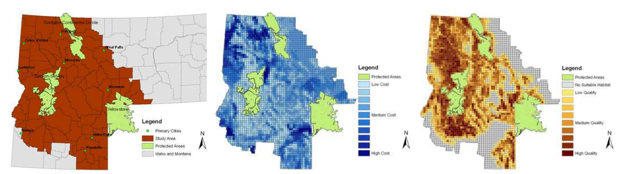

Figure 3: Study area for design of grizzly bear wildlife corridors (left) Land costs (center) Landscape resistance for grizzly bears (right) at 10km grid resolution.

One could use these data to generate landscape resistance values, since in many ecological studies habitat suitability (or permeability) and resistance are treated as complementary values (Coulon et al. 2004; Schwartz et al. 2009, e.g.). One could compute the resistance of a parcel v among all parcels V on the same scale as the utilities, using $resist(v) = \min_{u \in V} util(u) + \max_{u \in V} util(u) - util(v)$. The grizzly dataset transformed to resistances was used in (Dilkina, Lai, and Gomes 2011).

Lynx and Wolverine Dataset in Montana: The wolverine (*Gulo gulo*) and the Canada lynx (*Lynx canadensis*) are species that are proposed for listing as Threatened (Federal Register 77 FR 69993 70060) or Threatened (Federal Register 65 FR 16053 16086) under the U. S. Endangered Species act. In Montana, both suffer from habitat fragmentation and inhabit parts of the Northern Continental Divide Ecosystem and the Greater Yellowstone

Area (see Fig. 4). Therefore, preserving connectivity in this study area would be beneficial for both species. The dataset was compiled using publicly available GIS data[2]. This included elevation, land cover, roads, housing density, land value data (based on 2007 tax data), and persistent spring snow cover (queried from MODIS (Copeland et al. 2010)). The data are represented at a 6km resolution of square cells, with each cell having a cost and values for wolverine and lynx permeability. After pruning the study area to exclude eastern Montana, lakes, and barriers such as urban areas, the resulting dataset contains a total of 4514 planning cells. Versions of this dataset were used in (Lai et al. 2011; Le Bras et al. 2013). Full description of the dataset is provided online.

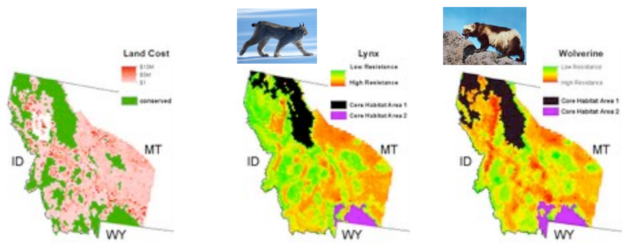

Figure 4: Western Montana case study: land cost (left), landscape resistance for lynx (center) and wolverines (right).

Core areas for wolverines were delineated based on conserved land with persistent spring snow cover (Copeland et al. 2010), and after additional filtering and consolidation resulted in 13 core areas. Landscape permeability for wolverines was calculated according to (Singleton, Gaines, and Lehmkuhl 2002). The core areas for lynx were based on conserved land that is part of lynx critical habitat as designated by the US Fish and Wildlife Service, and resulted in 4 consolidated core areas. Landscape permeability for the lynx was calculated according to (Bates and Jones 2007). To generate the costs, we regarded already-conserved land as free and otherwise used the taxable land value data from 2007. For cells that overlap some part of a primary road, we added a cost estimate for installing a wildlife bypass.

5 Conclusion

We present a synthetic dataset generator and real-world datasets to facilitate computational research on landscape connectivity problems in conservation planning.

Acknowledgments

This work was supported by the NSF Expeditions in Computing award for Computational Sustainability, grant 0832782), the NSF Computing research infrastructure for Computational Sustainability grant (grant 1059284), and the USDA Forest Service, Rocky Mountain Research Station (grant 10-JV-11221635-24).

[2]http://geoinfo.montanastatelibrary.org

References

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

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.

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

Copeland, J. P.; McKelvey, K. S.; Aubry, K. B.; Landa, A.; Persson, J.; et al. 2010. The bioclimatic envelope of the wolverine (gulo gulo): do climatic constraints limit its geographic distribution? *Canadian Journal of Zoology* 88(3):233–246.

Coulon, A.; Cosson, J.; Angibault, J.; Cargnelutti, B.; Galan, M.; et al. 2004. Landscape connectivity influences gene flow in a roe deer population inhabiting a fragmented landscape: an individual–based approach. *Molecular ecology* 13(9):2841–2850.

Dilkina, B., and Gomes, C. P. 2010. Solving connected subgraph problems in wildlife conservation. In *CPAIOR*, 102–116.

Dilkina, B.; Lai, K. J.; and Gomes, C. P. 2011. Upgrading shortest paths in networks. In *CPAIOR*, 76–91.

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

Joseph, L. N.; Maloney, R. F.; and Possingham, H. P. 2009. Optimal allocation of resources among threatened species: a project prioritization protocol. *Conservation biology* 23(2):328–38.

Kareiva, P. 2006. Introduction: evaluating and quantifying the conservation dividends of connectivity.

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

Le Bras, R.; Dilkina, B.; Xue, Y.; Gomes, C. P.; McKelvey, K. S.; Montgomery, C.; and Schwartz, M. K. 2013. Robust network design for multispecies conservation. In *AAAI*.

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

McRae, B. H.; Dickson, B. G.; Keitt, T. H.; and Shah, V. B. 2008. Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology* 89(10):2712–24.

Naidoo, R.; Balmford, A.; Ferraro, P.; Polasky, S.; Ricketts, T.; and Rouget, M. 2006. Integrating economic costs into conservation planning. *Trends in Ecology & Evolution* 21(12):681–687.

Rudnick, D. A.; Beier, P.; Cushman, S.; Dieffenbach, F.; Epps, C. W.; et al. 2012. The role of landscape connectivity in planning and implementing conservation and re storation priorities. *Issues in Ecology* 16.

Schwartz, M. K.; Copeland, J. P.; Anderson, N. J.; Squires, J. R.; Inman, R. M.; et al. 2009. Wolverine gene flow across a narrow climatic niche. *Ecology* 90(11):3222–3232.

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

Taylor, P. D.; Fahrig, L.; Henein, K.; and Merriam, G. 1993. Connectivity is a vital element of landscape structure. *Oikos* 73:43–48.