Reminder

- Confirm course project group members
  - Due 1/23, 10pm
- Online Homework 0 (HW0)
  - Required, but worth zero points, Due 1/23, 10 pm
- Paper Reading Assignment 1 (PRA1)
  - Due 1/25, 10 pm
- Project proposal
  - Due 1/30, 10pm

Artificial Intelligence Methods for Social Good

Lecture 3:

Case Study: AI for Wildlife Corridor Design

17-537 (9-unit) and 17-737 (12-unit)

Fei Fang

feifang@cmu.edu

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study

- Discussion

Learning Objectives

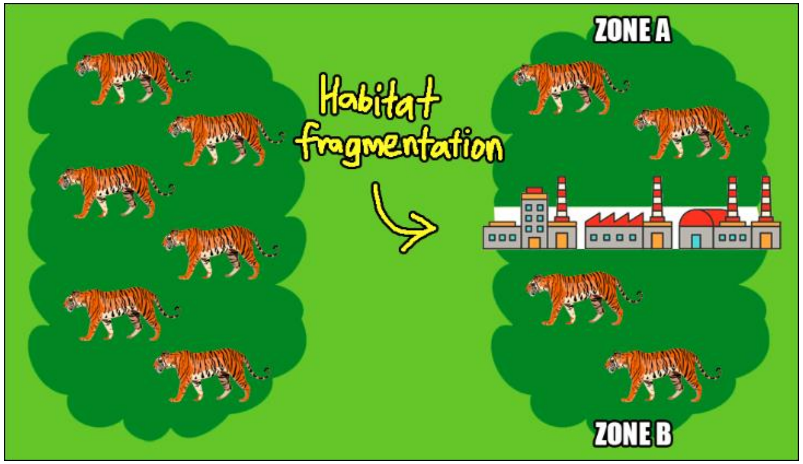
- Briefly describe
  - Challenges in wildlife corridor design
  - MILP-based solution for wildlife corridor design
  - Methodology of applying the solution to a specific case and evaluation criteria
- Write down general constraints for network flow problems

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

Motivation

- Wildlife habitat diminished and fragmented



https://cilisos.my/malaysia-is-building-a-love-tunnel-that-will-help-animals-find-their-soulmate/tiger-habitat-fragmentation/

https://commons.wikimedia.org/wiki/File:Indiana_Dunes_Habitat_Fragmentation.jpg

Motivation

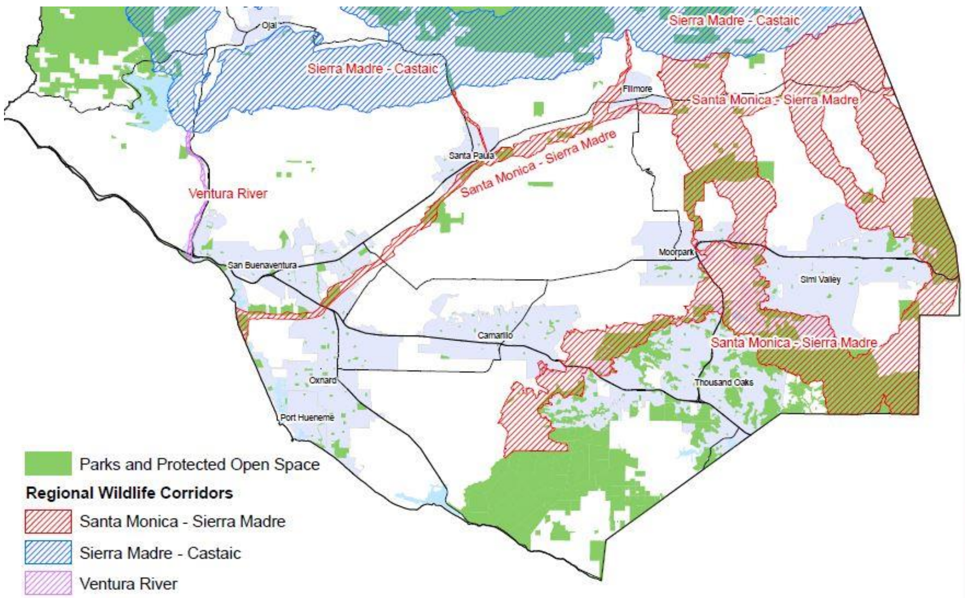
- Isolated protected areas are not enough for long-term maintenance of biodiversity
- To create/enhance connectivity: build and protect wildlife corridors

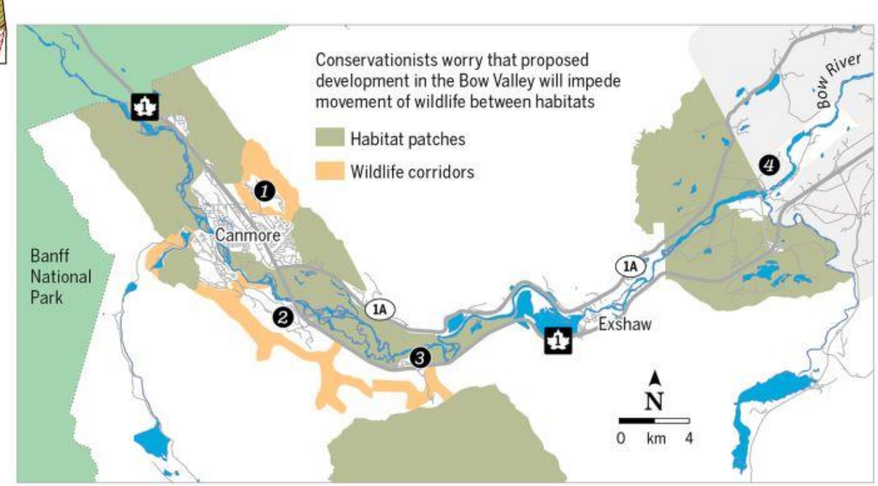
https://envirothink.wordpress.com/2015/06/09/new-wildlife-corridor-to-be-built-in-washington-state/

https://www.lifegate.com/people/lifestyle/5-important-wildlife-corridors

Motivation

- Question: Where to build wildlife corridor?





1 Silvertip Resort plans to build a $1-billion resort with up to 13 boutique hotels, conference centre and a gondola.

2 Three Sisters Mountain Village wants to develop its remaining land and proposes fencing a wildlife corridor.

3 Municipal District of Bighorn proposes future light industrial development at Dead Man's Flats.

4 Stoney Nakoda Developments wants to build a community with housing, business and amenities on 235 hectares along the Bow River.

SOURCE: ALBERTA PARKS DARREN FRANCEY / POSTMEDIA NEWS

https://twitter.com/Y2Y_Initiative/status/841314661039460352

http://colabvc.org/wildlife-corridors-already-wild/

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

Problem Statement

- Wildlife distribution: High density in core areas
  - Core areas of different species may overlap
- Wildlife movement:
  - May move in any direction, heterogeneous difficulty
  - Each pixel associated with a resistance cost
  - Path of higher total resistance cost is more difficult to walk through
- Build a corridor: purchase parcels of land to connect protected areas
  - Parcels purchased + existing protected area = conservation network

Problem Statement

- Single-minded goal: build corridors to connect core areas of a species and minimize total resistance cost
  - Connect core areas: exist a path that falls entirely within the conservation network

- Limitations
  - Economic cost is not considered
  - Multiple species are not considered

- Ideally:
  - Connect core areas for all species
  - Low total resistance cost (cumulative resistance)
  - Low expenditure on purchasing the parcels (expenditure)

Problem Statement

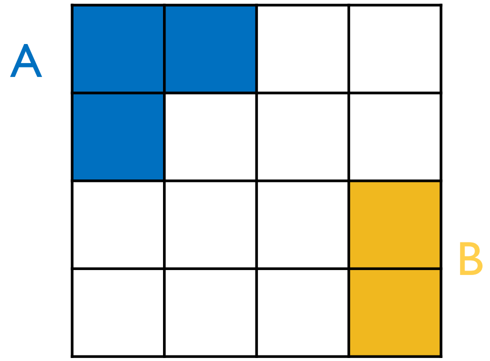
- Problem Statement: Budget constrained corridor design for multiple species
  - Set limit on expenditure
  - Minimize cumulative resistance
  - Ensure connectivity between each pair of core areas of each species

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

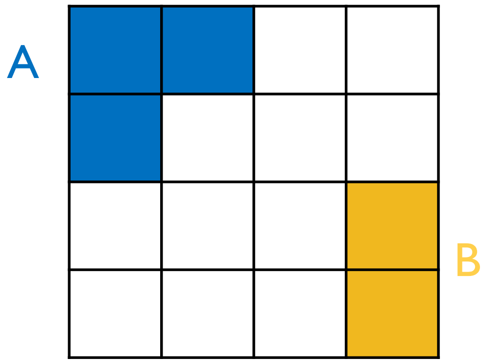
Model

- A raster of grid cells
- A core area: a set of contiguous raster cells



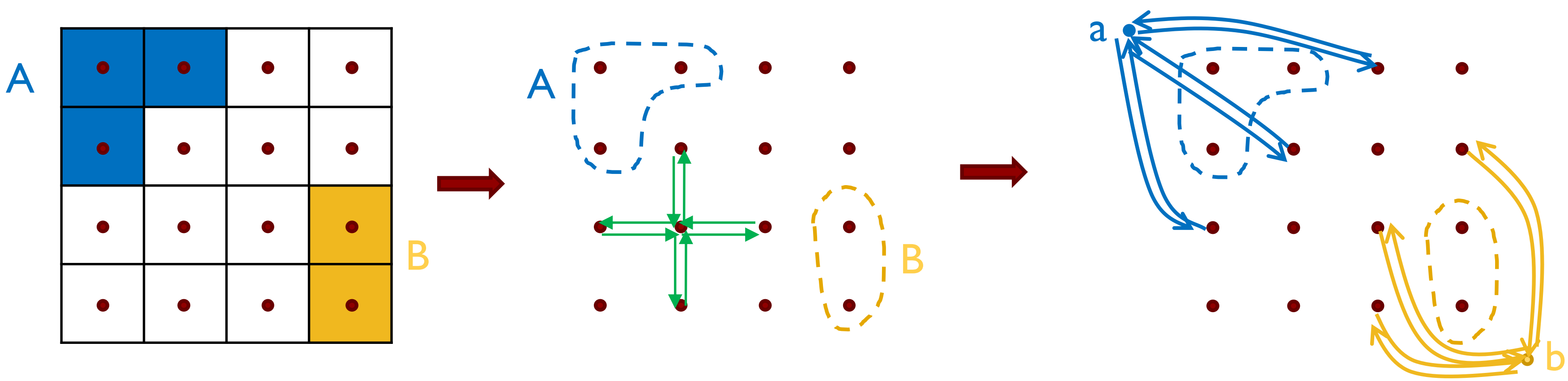
Graph Model for Corridor Design Problem

- Nodes: a cell that can be purchased (not in core areas)
- Edges: connecting neighboring cells
- Additional nodes: virtual nodes for core areas
- Additional edges: core areas and their neighboring cells

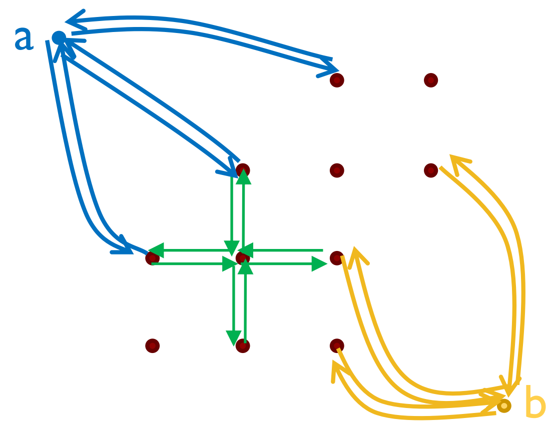


Graph Model for Corridor Design Problem

- Nodes: a cell that can be purchased (not in core areas)
- Edges: connecting neighboring cells
- Additional nodes: virtual nodes for core areas
- Additional edges: core areas and their neighboring cells



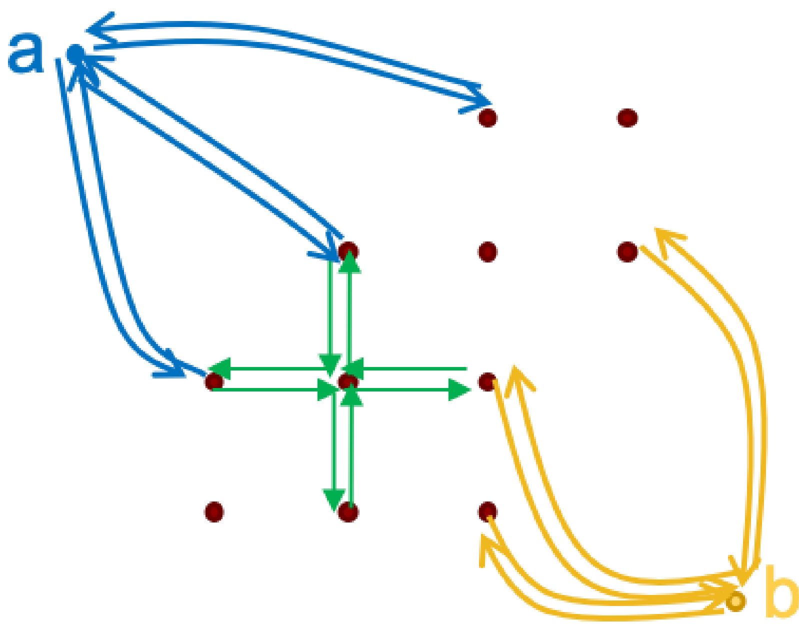
Graph Model for Corridor Design Problem



- For each node v
  - Acquisition cost $c(v)$
  - Resistance value $r^s(v)$ for species s
  - Special case: $c(a) = c(b) = 0, r^s(a) = r^s(b) = 0$
- Connectivity requirements: $P^s = \{(a_1, b_1); (a_2, b_2); \dots\}$
  - Pairs of (virtual) nodes of species s

Model

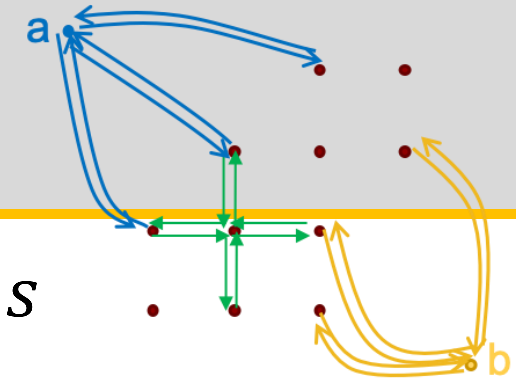
- Corridor design: select a subset of nodes on the graph to ensure connectivity between core areas



Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

Approach – Single Species



- Optimization Problem for Single Species s
  - MILP1

Objective function: Total cumulative resistance of best paths for all pairs for species s

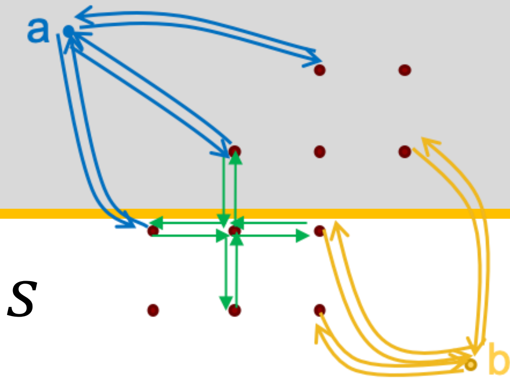
Variable x_v: Whether or not to select (i.e., purchase or acquire) node v

Budget constraint

Some constraints to ensure connectivity

Recall: P^S: the set of all pairs of core areas for species s
Acquisition cost for node v: $c(v)$

Approach – Single Species



- Optimization Problem for Single Species s
  - MILP1

R_p^s represents the cumulative resistance of best path linking core area pair p for species s

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^S} R_p^s$$

Objective function: Total cumulative resistance of best paths for all pairs for species s

$$\text{s.t. } x_v \in \{0,1\}, \forall v \in V$$

Variable x_v: Whether or not to select (i.e., purchase or acquire) node v

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

Budget constraint

$$\Pi_{sp}, \forall p \in P^s$$

Some constraints to ensure connectivity

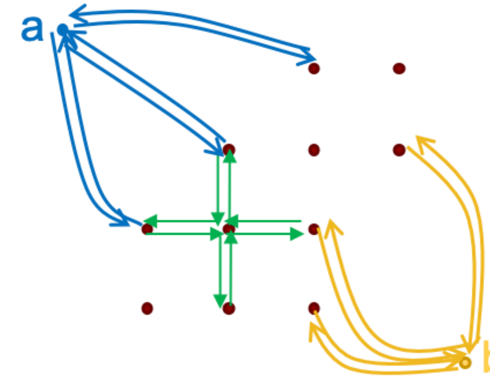
Π_{sp} represents a set of constraints that ensures there is a path linking p for species s

Recall: P^S: the set of all pairs of core areas for species s

Acquisition cost for node v: $c(v)$

Approach – Single Species

- Π_{sp} represents a set of constraints that ensures there is a path linking p for species s



Define new variable $y_e^{sp} \in \{0,1\}$ to represent whether edge e is on the best path that connects the pair p of species s

$\delta^-(v) \triangleq$ incoming edges for v

$\delta^+(v) \triangleq$ outgoing edges for v

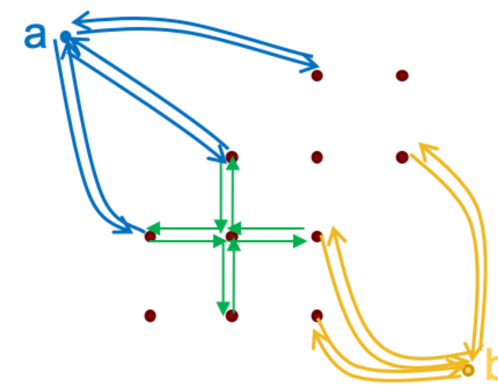
Let $p = (a, b)$, then the constraints are

Connecting path can only traverse selected nodes

Flow conservation

Approach – Single Species

- Π_{sp} represents a set of constraints that ensures there is a path linking p for species s



Define new variable $y_e^{sp} \in \{0,1\}$ to represent whether edge e is on the best path that connects the pair p of species s

$\delta^-(v) \triangleq$ incoming edges for v

$\delta^+(v) \triangleq$ outgoing edges for v

Let $p = (a, b)$, then the constraints are

$$\sum_{e \in \delta^-(v)} y_e^{sp} \leq x_v, \forall v \in V \backslash \{a, b\}$$

Connecting path can only traverse selected nodes

$$\sum_{e \in \delta^+(a)} y_e^{sp} = \sum_{e \in \delta^-(b)} y_e^{sp} = 1$$

$$\sum_{e \in \delta^+(v)} y_e^{sp} = \sum_{e \in \delta^-(v)} y_e^{sp}, \forall v \in V \backslash \{a, b\}$$

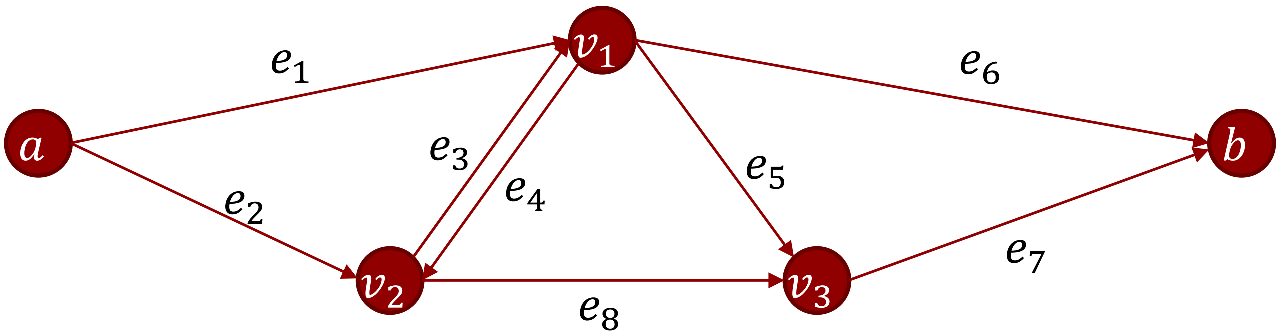
Flow conservation

$$y_e^{sp} \in \{0,1\}$$

Example for Flow Constraints

$$\sum_{e\in\delta^+(a)} y_e^{sp} = \sum_{e\in\delta^-(b)} y_e^{sp} = 1$$

$$\sum_{e\in\delta^+(v)} y_e^{sp} = \sum_{e\in\delta^-(v)} y_e^{sp}, \forall v \in V\backslash\{a,b\}$$

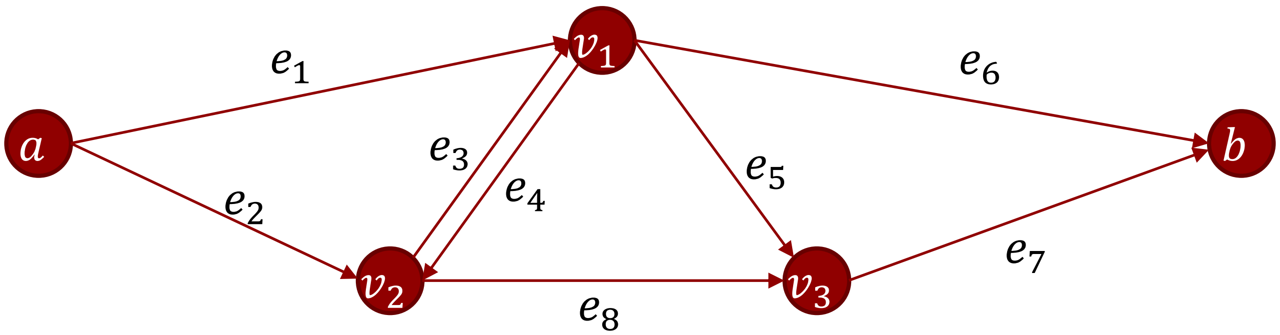


- $y_e \in \{0,1\}$: whether or not e is on the selected path
- For any path connecting a, b that does not traverse the same edge twice, the corresponding y_e satisfy:
- For any y that satisfy these constraints, it corresponds to a path connecting a, b

Example for Flow Constraints

$$\sum_{e \in \delta^+(a)} y_e^{sp} = \sum_{e \in \delta^-(b)} y_e^{sp} = 1$$

$$\sum_{e \in \delta^+(v)} y_e^{sp} = \sum_{e \in \delta^-(v)} y_e^{sp}, \forall v \in V \backslash \{a, b\}$$



- $y_e \in \{0,1\}$: whether or not e is on the selected path
- For any path connecting a, b that does not traverse the same edge twice, the corresponding y_e satisfy:

$y_{e_1} + y_{e_2} = 1$ (one edge goes out of node a)
$y_{e_6} + y_{e_7} = 1$ (one edge goes into node b)
$y_{e_1} + y_{e_3} = y_{e_4} + y_{e_5} + y_{e_6}$ (if there is an edge (or 2) goes into node v_1, there must be an edge (or 2) goes out of node v_1)
$y_{e_2} + y_{e_4} = y_{e_3} + y_{e_8}, y_{e_5} + y_{e_8} = y_{e_6} + y_{e_7}$

- For any y that satisfy these constraints, it corresponds to a path connecting a, b

Poll 1

- Given directed graph $G = (V, E)$, each node representing a city. A company needs to send K cellphones from city s to city d. It may send the cellphones through multiple paths. Let y_e be the number of cellphones sent through edge $e \in E$. Let $\delta^-(v)$ and $\delta^+(v)$ denote the set of incoming and outgoing edges for $v \in V$. Which ones of the following are necessary constraints for y_e?
  - A: $\sum_{e \in \delta^+(s)} y_e = K$
  - B: $\sum_{e \in \delta^-(d)} y_e = K$
  - C: $\sum_{e \in \delta^+(v)} y_e = K, \forall v \in V$
  - D: $\sum_{e \in \delta^+(v)} y_e = \sum_{e \in \delta^-(v)} y_e, \forall v \in V$
  - E: $\sum_{e \in \delta^+(v)} y_e = \sum_{e \in \delta^-(v)} y_e, \forall v \in V \setminus \{s, d\}$
  - F: I don't know

Approach – Single Species

- Additional constraints and simplifications for Π_{sp}
  - If some nodes are not admissible for pair p of species s, e.g., slope is too high for species s to move:
  - Relax binary constraint on y_e^{sp} will not change the solution according to known results in network flow (ILP=LP)

Approach – Single Species

- Additional constraints and simplifications for Π_{sp}
  - If some nodes are not admissible for pair p of species s, e.g., slope is too high for species s to move:

$$y_e^{sp} = 0, \forall e = (u, v) \text{ where } adm_v^{sp} = 0 \text{ or } adm_u^{sp} = 0$$

  - Relax binary constraint on y_e^{sp} will not change the solution according to known results in network flow (ILP=LP)

$$y_e^{sp} \in [0,1]$$

Approach – Single Species

- R_p^s represents the cumulative resistance of best path linking core area pair p for species s

Recall:
$y_e^{sp} \in \{0,1\}$ represents whether edge e is on the best path that connects the pair p of species s
$r^s(v)$ represents resistance value of node v for species s

Q1: Is it equivalent to $R_p^s = \sum_{v \in V_p} r^s(v)$ where V_p is the set of nodes on the path connecting pair p?
Q2: Is it equivalent to $\sum_{v \in V} r^s(v) x_v$?

Approach – Single Species

- R_p^s represents the cumulative resistance of best path linking core area pair p for species s

Recall:
$y_e^{sp} \in \{0,1\}$ represents whether edge e is on the best path that connects the pair p of species s
$r^s(v)$ represents resistance value of node v for species s

$$R_p^s = \sum_{e=(u,v)\in E} \frac{r^s(u) + r^s(v)}{2} y_e^{sp}$$

Remark 1: This is equivalent to $R_p^s = \sum_{v\in V_p} r^s(v)$ where V_p is the set of nodes on the path connecting pair p. However, V_p is not known ahead of time. So we cannot write it in this form
Remark 2: This can be different from $\sum_{v\in V} r^s(v) x_v$. To ensure that they are equivalent, we need the assumption that a node not on the path of p will never be selected. This assumption holds if we are considering the optimization problem with single species and a single pair of core areas. When we consider more pairs or more species, they are not the same.

Approach – Single Species

- Putting everything together (MILP1)

$$R^s(B) \triangleq \min_{x,y} \sum_{p \in P^S} \sum_{e=(u,v) \in E} \frac{r^s(u) + r^s(v)}{2} y_e^{sp}$$

s.t.

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

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

$$\left.\begin{aligned} \sum_{e \in \delta^-(v)} y_e^{sp} \leq x_v, \forall v \in V \backslash \{a,b\} \\ \sum_{e \in \delta^+(a)} y_e^{sp} = \sum_{e \in \delta^-(b)} y_e^{sp} = 1 \\ \sum_{e \in \delta^+(v)} y_e^{sp} = \sum_{e \in \delta^-(v)} y_e^{sp}, \forall v \in V \backslash \{a,b\} \end{aligned}\right\} \forall p = (a,b) \in P^s$$

$y_e^{sp} = 0, \forall e = (u,v)$ where $adm_v^{sp} = 0$ or $adm_u^{sp} = 0$

$$y_e^{sp} \in [0,1], \forall e \in E, \forall p \in P^s$$

Approach – Two Species

- Optimization Problem for Two Species g and w
  - Updated objective function of MILP1
  - α controls the balance between the two species

Normalization used to avoid comparison in completely different scales

(Recall MILP1)

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^s} R_p^s$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

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

$$\Pi_{sp}, \forall p \in P^s$$

Pre-computed. Optimal value of optimization problem for single species w

Approach – Two Species

- Optimization Problem for Two Species g and w
  - Updated objective function of MILP1
  - α controls the balance between the two species

$$\min_{x,y,R^g,R^w} \alpha \frac{R^g}{R^g(B)} + (1-\alpha)\frac{R^w}{R^w(B)}$$

Normalization used to avoid comparison in completely different scales

$$R^g = \sum_{p \in P^g} R_p^g$$

$$R^w = \sum_{p \in P^w} R_p^w$$

Pre-computed. Optimal value of optimization problem for single species w

(Recall MILP1)

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^s} R_p^s$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

$$\sum_{v \in V} c(v) x_v \le B$$

$$\Pi_{sp}, \forall p \in P^s$$

Update to $\Pi_{sp}, \forall s, \forall p \in P^s$

Approach – Multiple Species

- Optimization Problem for Multiple Species
  - Extend the objective function for two species to multiple species
  - α controls the balance between the species

Approach – Multiple Species

- Optimization Problem for Multiple Species
  - Extend the objective function for two species to multiple species
  - α controls the balance between the species

$$\min_{x,y,R} \sum_i \alpha_i \frac{R^i}{R^i(B)}$$

Boundary Solutions

- Minimum budget to ensure connectivity
  - Slight modifications to MILP1

(Recall MILP1)

$$R^S(B) \triangleq \min_{x,\ldots} \sum_{p \in P^S} R_p^S$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

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

$$\Pi_{sp}, \forall p \in P^S$$

Boundary Solutions

- Minimum budget to ensure connectivity
  - Slight modifications to MILP1

$$\min_{x,\dots} \sum_{v \in V} c(v) x_v$$

s.t.

$$x_v \in \{0,1\}, \forall v \in V$$
$$\Pi_{sp}, \forall p \in P^s$$

(Recall MILP1)

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^s} R_p^s$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

$$\sum_{v \in V} c(v) x_v \leq B$$
$$\Pi_{sp}, \forall p \in P^s$$

Boundary Solutions

- Minimum cumulative resistance if no budget constraint

Compute Minimum Cumulative Resistance

- For each s
  - For each $p \in P^s$
    - Compute R_p^s:
- Compute $\underline{R^s} = \sum_p R_p^s$

Recall: Special case: $c(a) = c(b) = 0, r^s(a) = r^s(b) = 0$

Boundary Solutions

- Minimum cumulative resistance if no budget constraint

Compute Minimum Cumulative Resistance

- For each s
  - For each $p \in P^s$
    - Compute R_p^s: If $p = (a, b)$, then find shortest path from a to b on constructed graph where distance is defined as $\frac{r^s(u)+r^s(v)}{2}$ for any edge $e = (u, v)$. R_p^s is the length of the shortest path.
- Compute $\underline{R^s} = \sum_p R_p^s$

Recall: Special case: $c(a) = c(b) = 0, r^s(a) = r^s(b) = 0$

Boundary Solutions

- Minimum budget solution among the ones with minimum cumulative resistance
  - First find minimum cumulative resistance $\underline{R^s}$
  - Then make slight modifications to MILP1

(Recall MILP1)

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^s} R_p^s$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

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

$$\Pi_{sp}, \forall p \in P^s$$

Boundary Solutions

- Minimum budget solution among the ones with minimum cumulative resistance
  - First find minimum cumulative resistance $\underline{R^s}$
  - Then make slight modifications to MILP1

$$\min_{x,\dots} \sum_{v \in V} c(v) x_v$$

s.t.

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

$$\Pi_{sp}, \forall p \in P^s$$

$$\sum_{p \in P^S} R_p^s \leq \underline{R^s}$$

(Recall MILP1)

$$R^s(B) \triangleq \min_{x,\dots} \sum_{p \in P^S} R_p^s$$

s.t. $x_v \in \{0,1\}, \forall v \in V$

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

$$\Pi_{sp}, \forall p \in P^s$$

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

Case Study

- Wolverines and Grizzly Bears in Western Montana
  - Low population, concentrated
    - Yellowstone National Park, Bob Marshall Wilderness Complex
    - 12.8 wolverines across 3 mountain ranges
    - 48 grizzly bears in 9900-km^2 zone

https://www.pinterest.com/pin/488429522063700417/

https://en.wikipedia.org/wiki/Grizzly_bear#/media/File:Grizzlybear55.jpg

Case Study

- Wolverines and Grizzly Bears in Western Montana
  - Different habitat requirements
  - Habitats partially overlap
  - Different capability of movement

Case Study

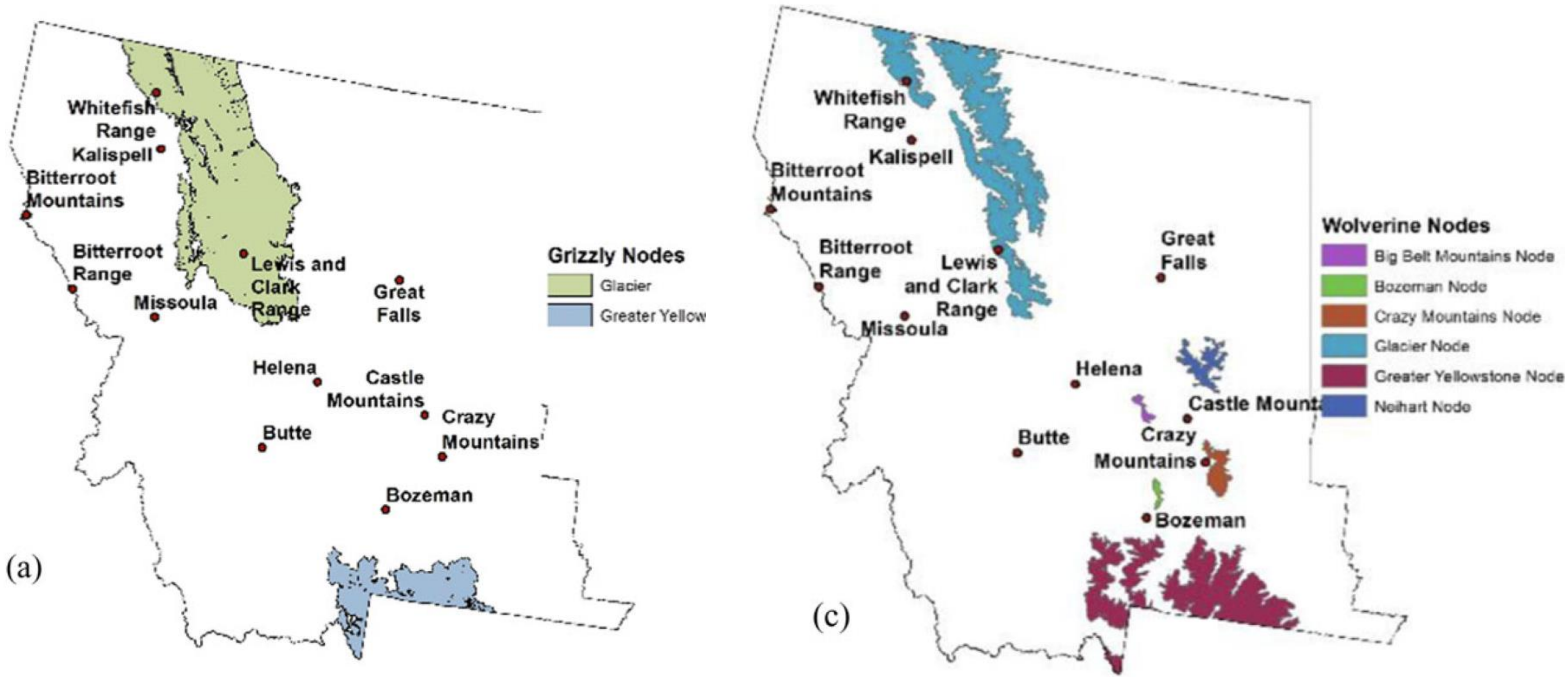
- Lands being considered
  - Public area (held by National Parks, U.S. Forest Service etc)
  - Tribal lands
  - Private lands (held by NGOs, timber companies, individuals etc)

Case Study

- Input for the Model / Data source
  - Western Montana, 1000m grid
  - Acquisition cost
    - Tax records
    - Information on conserved lands
    - Other information: water body, urban parcel, etc
    - Gap between model and practice: a parcel is not a set of cells
    - Estimated acquisition cost: area-weighted sum of all the parcel values in the cell (using ArcGIS)
  - Resistance
    - Geographical information and other landscape features
      - Grizzly bears: vegetation, human development, road density
      - Wolverines: snow cover, housing development, forest edge
    - Estimate resistance: Follow established method in conservation

Case Study

- Core areas
  - Grizzly bears: Northern Continental Divide Ecosystem and Greater Yellowstone Ecosystem
  - Wolverines: use habitat rule to identify core areas



Case Study

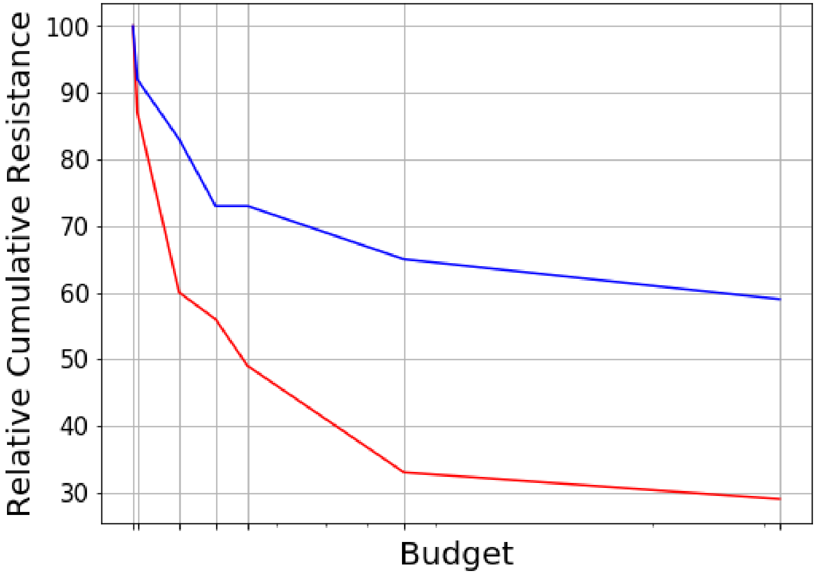
- Computation
  - Pruning (could be lossy), i.e., exclude cells that
    - Could not be made passable
    - Very far from any reasonable pathway
    - If included in the path, will lead to a high cumulative resistance
  - 42065 cells
  - Solve MILP using CPLEX, run on cluster
    - 5-40 hours of computer time

Case Study

- Results
  - Provide insights, suggestions, visualizations to assist decision makers
    - Boundary Solutions
      - Minimum budget to ensure connectivity: $2.9M (high cumulative resistance)
      - Least-resistance paths: $31.8M expenditure (cumulative resistance is 29% and 59% of the min-budget design for grizzly bear and wolverine separately)

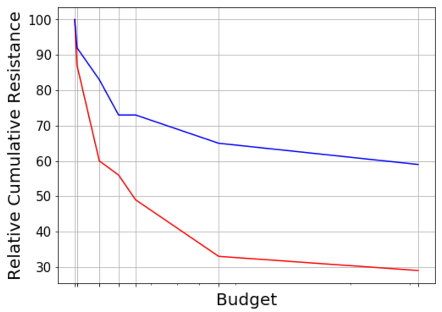
Case Study

- Results
  - Provide insights, suggestions, visualizations to assist decision makers
    - Fix $\alpha = 0.5$, examine tradeoff between budget and cumulative resistance
      - Find "Elbow" point



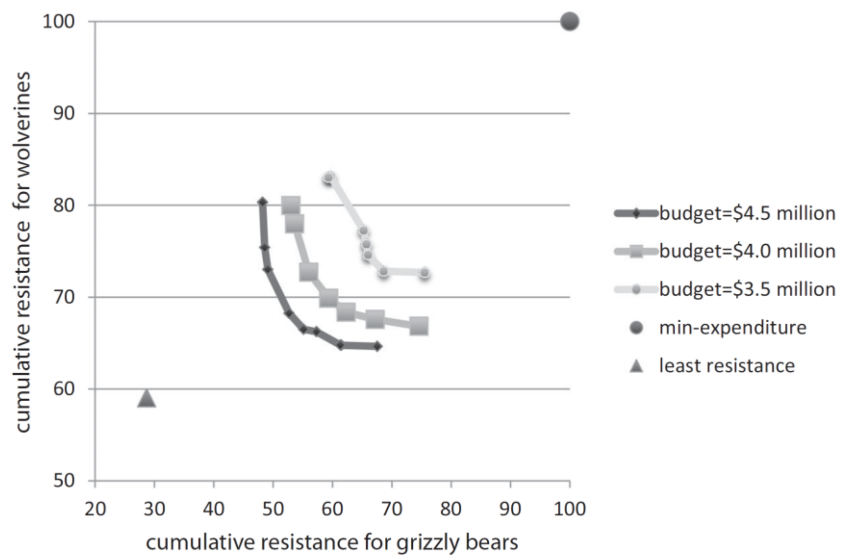
Poll 2

- Which ones of the following are true about the "elbow" point in the tradeoff plot of budget and cumulative resistance?
  - A: When budget is above this point, increase in budget does not lead to a significant reduction in cumulative resistance (compared to when budget is below this point)
  - B: Can be found by linking the first and last point to get a line, and check which point is farthest from this line
  - C: Is the ideal solution for wildlife corridor design problem
  - D: Can be a suggested solution to policy makers
  - E: I don't know



Case Study

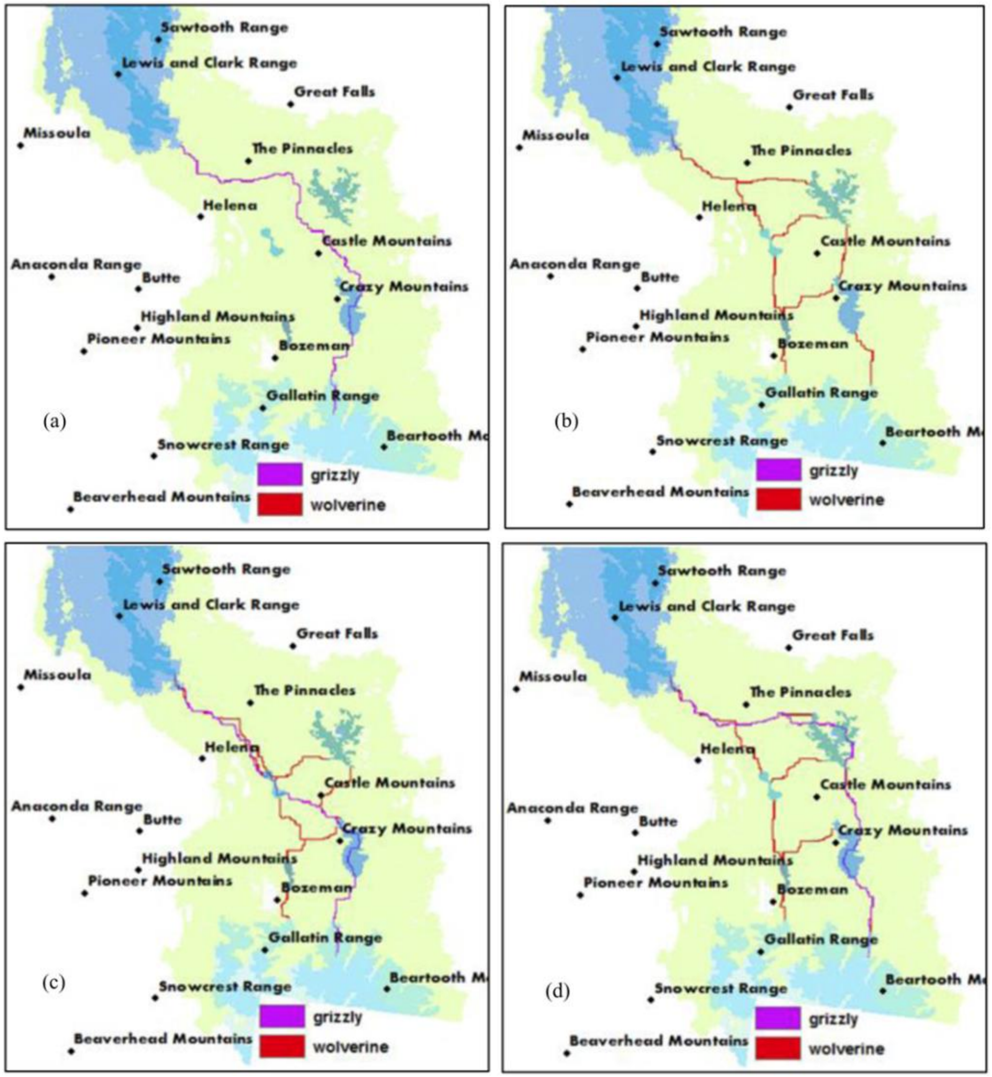
- Results
  - Provide insights, suggestions, visualizations to assist decision makers
    - Fix budget, plot cumulative resistance of two species with varying α
      - Find "Elbow" point
      - Difference across species: societal concerns and need for connectivity



Case Study

- Evaluation
  - Evaluate the advantage of optimizing jointly
  - Compare against separate single-species corridor design
  - Same total budget, compare cumulative resistance for both species
    - $4M for single-species corridor design for each species, get 67% and 40% of relative cumulative resistance for grizzly bear and wolverine
    - $8M for two-species corridor design with α= 0.5, get 65% and 33%
    - What's missing here?

Poll 3



- Compare the two results in the lower half. They correspond to different value of α (importance of grizzly bears). Which one corresponds to a higher value of α?
  - A: Lower Left
  - B: Lower Right
  - C: I don't know

Outline

- Wildlife Corridor Design
  - Motivation
  - Problem Statement
  - Model
  - Approach
  - Case Study
- Discussion

Discussion

- Heterogeneity: What if different core area pairs have different importance?
- Uncertainty in input: what if estimated resistance is not accurate?
- Uncertainty in acquisition: what if the purchase of a patch may fail?
- What if estimated resistance is not additive?
- How to reduce the runtime?
- Brainstorming: How can AI be used for protecting wildlife habitat?

Discussion

- What are the other potential ways to use AI for wildlife conservation?

References and Additional Resources

Reference and Related Work

- *Trade-offs and efficiencies in optimal budget-constrained multispecies corridor networks*
  - Bistra Dilkina, Rachel Houtman, Carla P. Gomes, Claire A. Montgomery, Kevin S. McKelvey, Katherine Kendall, Tabitha A. Graves, Richard Bernstein, Michael K. Schwartz
- *Solving Connected Subgraph Problems in Wildlife Conservation*
  - Bistra Dilkina & Carla P. Gomes
- *Robust Network Design for Multispecies Conservation*
  - Ronan Le Bras, Bistra Dilkina, Yexiang Xue, Carla P. Gomes, Kevin S. McKelvey, Michael K. Schwartz, Claire A. Montgomery

- Spatial conservation prioritization: quantitative methods and computational tools. Moilanen et al. 2009.