# Optimization Models and Algorithms in Natural Resource Planning

David Shmoys
Cornell University

# Institute for Computational Sustainability


**A Multi-institutional, Multidisciplinary Research Team**

*6 Institutions, 7 colleges, 13 departments*

29 graduate students
24 undergrad. students

# Vision

*Computer scientists can — and should — play a key role in increasing the efficiency and effectiveness of the way we manage and allocate our natural resources, while enriching and transforming Computer Science.*

Why we proposed this expedition in
Computational Sustainability!!!

# Conservation and Biodiversity: Reserve Design for Bird Conservation

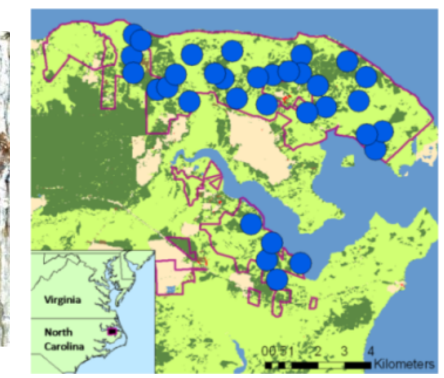
Red Cockaded Woodpecker (RCW) is a federally endangered species

Current population is estimated to be about
1% of original stable population (~12,000 birds)

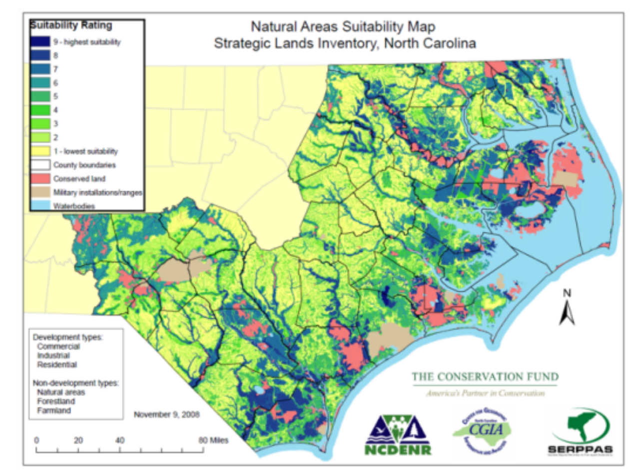
Conservation Funds manages

Palmetto Peartree Preserve (North Carolina)

32 active RCW territories (as of Sept 2008)









**Goal: Increase RCW population level**

**Management options:**

Prioritizing land acquisition adjacent
to current RCW populations

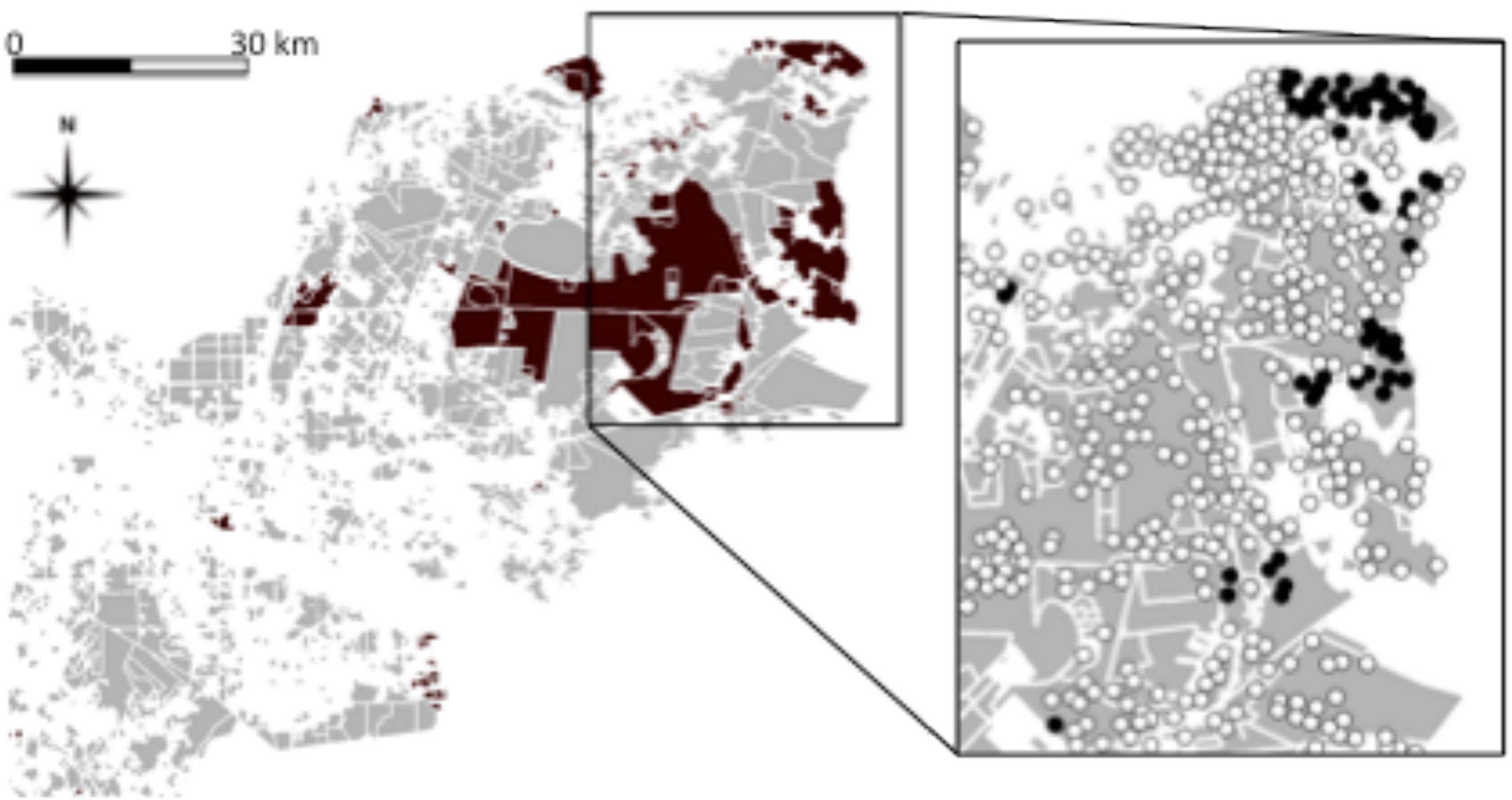
Building artificial cavities

Translocation of birds

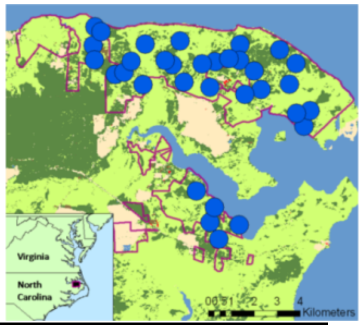
D. Sheldon, B. Dilkina, A. Elmachtoub, R. Finseth, A. Sabharwal, J. Conrad, C. Gomes, D. Shmoys, W. Allen, O. Amundsen, and B. Vaughn, *Maximizing the Spread of Cascades Using Network Design*

Problem Statement

- Given a limited conservation budget, which habitat patches should you put into conservation to maximize RCW population growth?



# Conservation and Biodiversity: Reserve Design for Bird Conservation



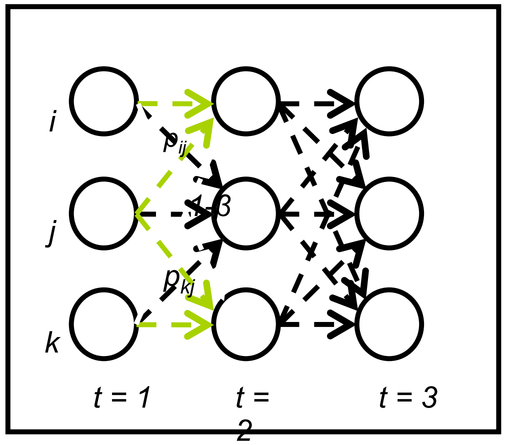

Red Cockaded W.

Biological and Ecological Model

Management Decisions:

Land acquisition
Artificial cavities
Translocation of birds

**Must explicitly consider interactions between biological/ecological patterns and management decisions**




**Stochastic diffusion model (movement and survival patterns) in RCW populations**

$$\text{maximize } (1/K)\sum_{k=1}^{K}\sum_{i\in\mathcal{R}} x^k(i,T)$$

subject to

$$\begin{aligned}
\sum_{i\in\mathcal{R}}\sum_{t=1}^{T} b(i,t)y(i,t) &\leq B; \\
\sum_{t=1}^{T} y(i,t) &\leq 1 \quad \forall \text{ territories } i \in \mathcal{R}; \\
x^k(i,t) &\leq \sum_{s=1}^{t} y(i,s), \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}, \text{ and periods } t; \\
x^k(i,1) &\leq a^k(i,i,1), \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}; \\
x^k(j,t) &\leq \sum_{i\in\mathcal{R}} a^k(i,j,t)x^k(i,t-1), \quad \forall \text{ scenarios } k, \text{ territories } j \in \mathcal{R}, \text{ and periods } 2\ldots T; \\
x^k(i,t) &\geq 0, \quad \forall \text{ scenarios } k, \text{ territories } i \in \mathcal{R}, \text{ and periods } t; \\
y(i,t) &\in \{0,1\}, \quad \forall \text{ territories } i \in \mathcal{R} \text{ and periods } t;
\end{aligned}$$

**Stochastic optimization model**
**Decisions: where and when to acquire land parcels**
**Goal: Maximize expected number of surviving RCW**

**Computational Challenge: scaling up solutions for considering a large number of years (100+)**
**→ decomposition methods and exploiting structure**

## PROBLEM: When do we buy territories and/or make them suitable?

- Suppose we want to maximize the expected total number of occupied regions at the end of time horizon (oversimplified objective)
- Decide to buy/improve certain territories in order to increase the potential number of future occupied territories
- Decision effects propagate across the space-time domain
- There is a budget constraint that limits the total spent on acquisition/improvement

# Simple Patch-based Diffusion Model

- There is a set $\mathcal{R}$ of regions and a time horizon of T periods
- For each region i $\in \mathcal{R}$ and for each t=1,…,T,
  if the region is occupied at that time, then the territory becomes unoccupied with probability $\beta$
- For each pair of regions i,j $\in \mathcal{R}$ and for each t=1,…,T,
  there is a given probablity $p_{ij}$, that, conditioned on the event that region i is occupied at time t-1, that region j is occupied at time t
- The transition probabilities were drawn based on the RCW DSS code provided to us by Jeff Walters.

- "True" Stochastic Optimization Model

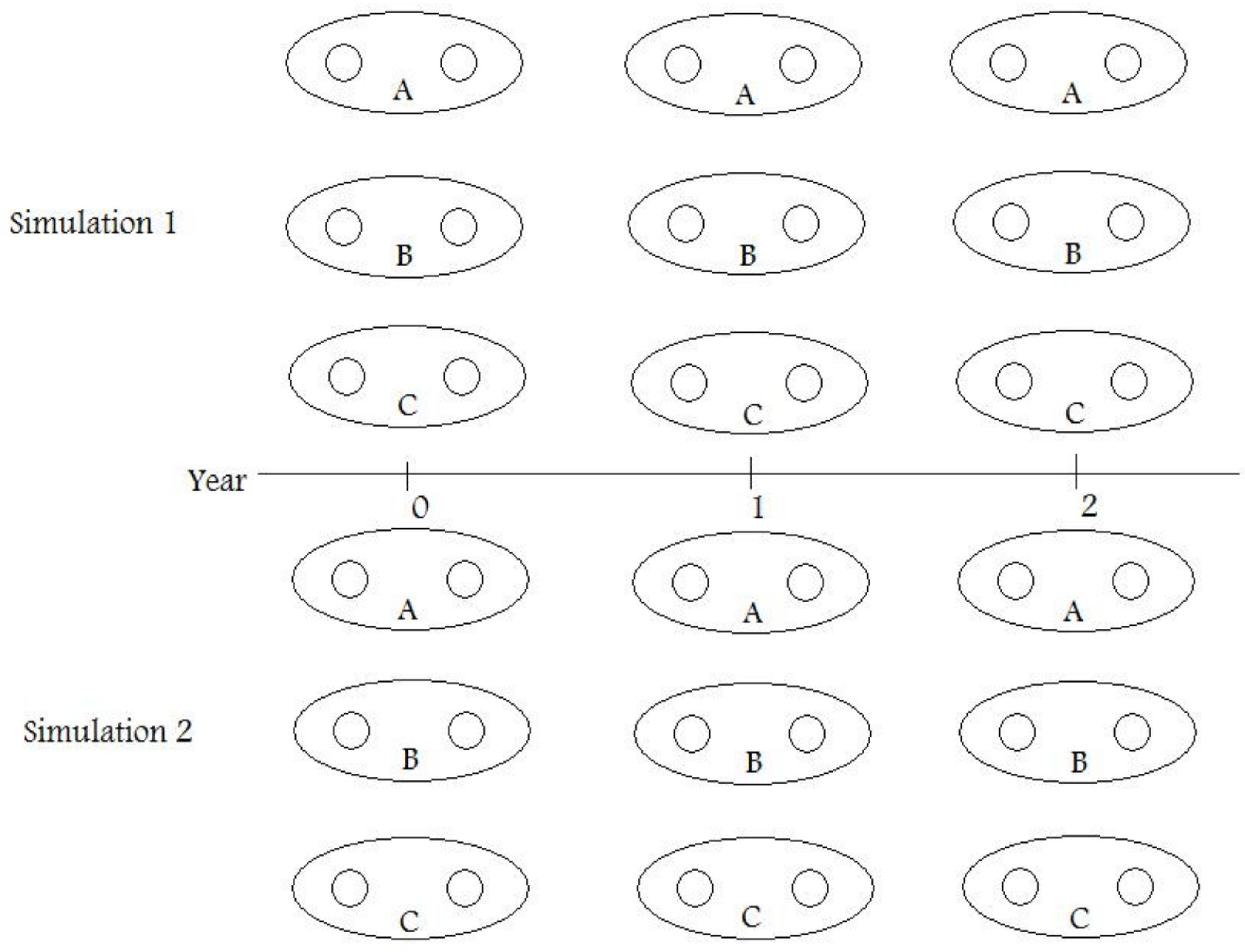
- Maximize $E_P(F(x,y))$
- subject to $y \in Y$
- where P is a probability distribution over possible inputs x

- Sample Average Approximation

- Draw m samples $x_1, x_2, \ldots, x_m$ independently from P
- and instead

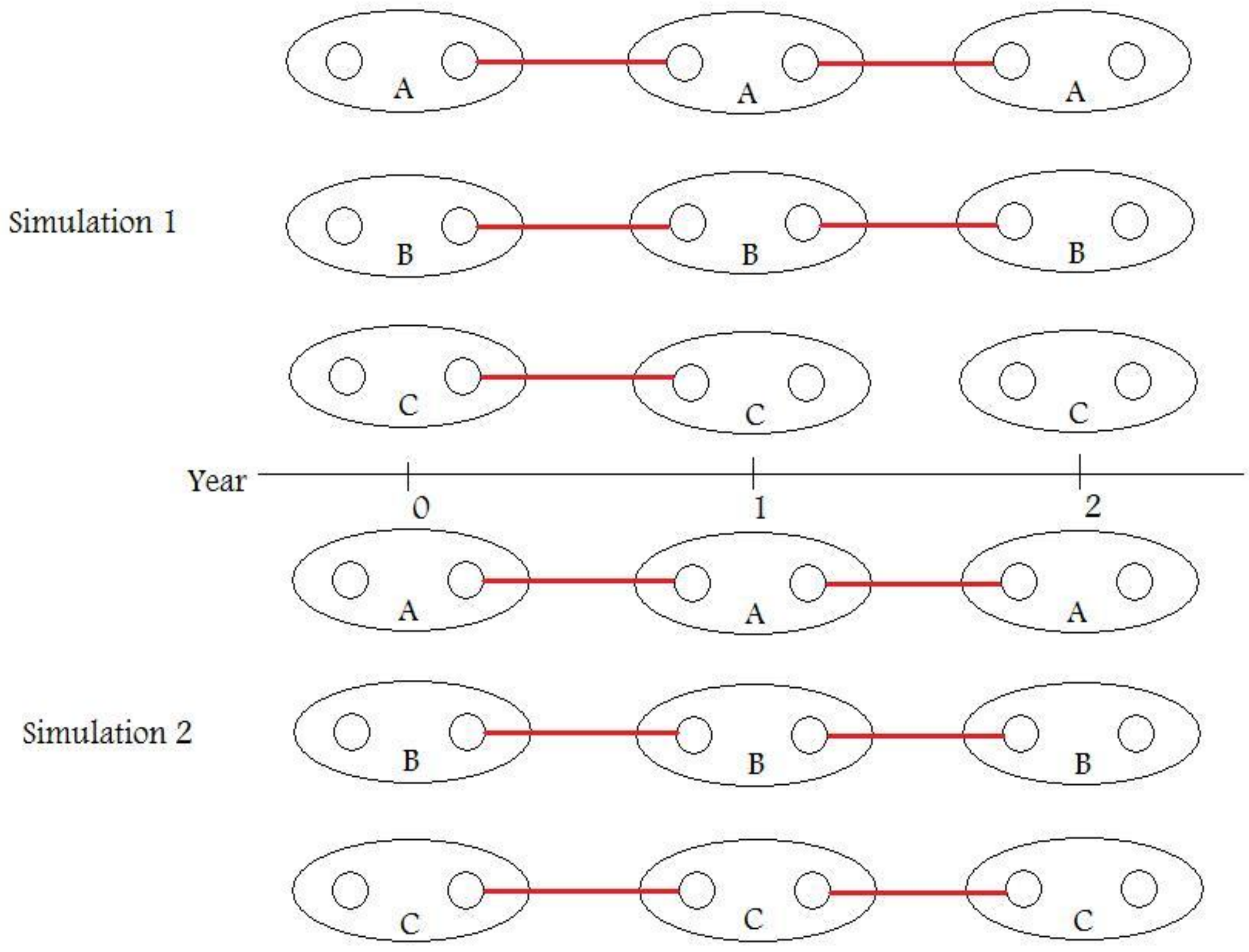
- Maximize $(1/m) \sum_i F(x_i,y)$
- subject to $y \in Y$

# This can be modeled as a network connectivity problem



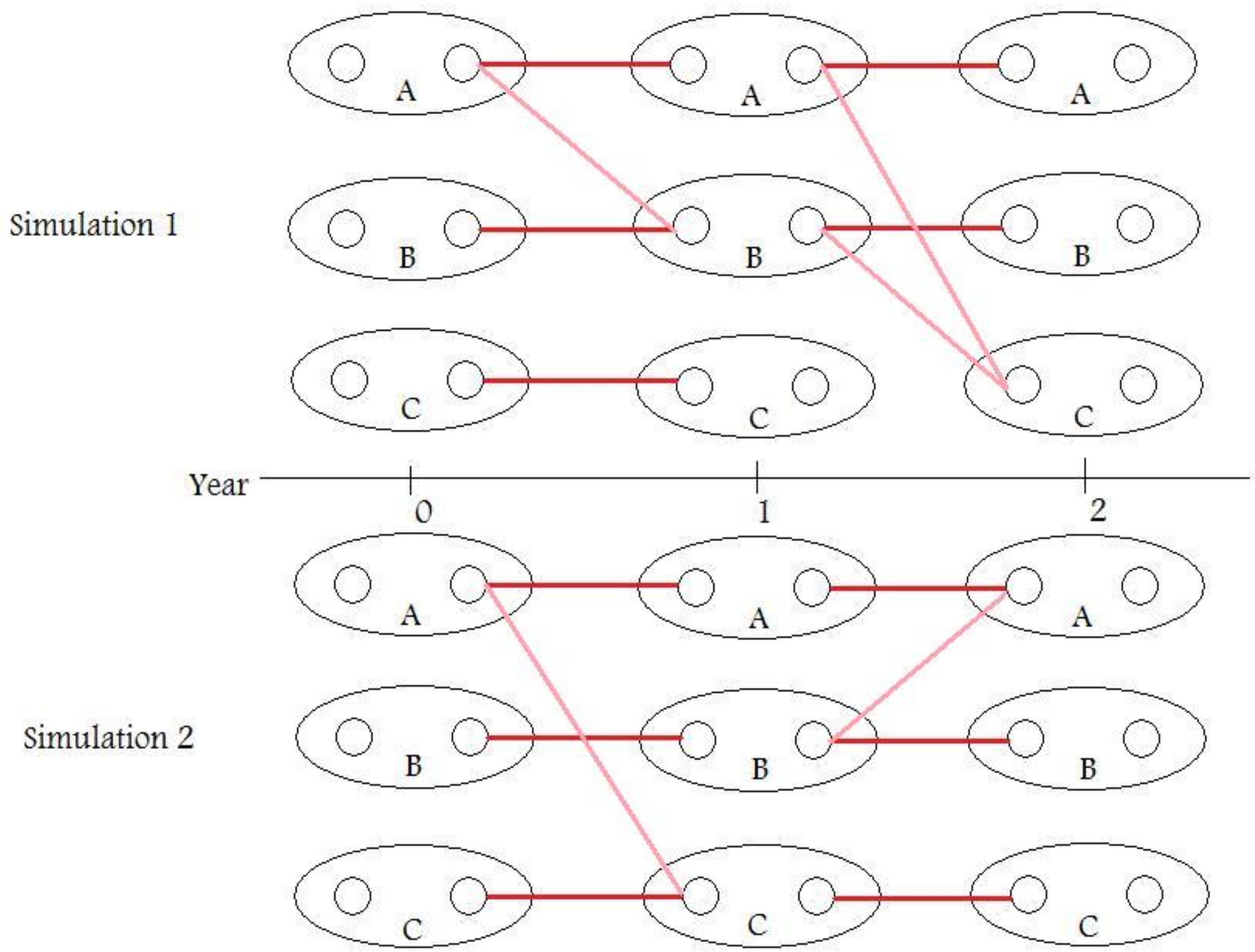

- Territories A,B,C
- 2 Years
- 2 "Trials"

## Red lines indicate the chance of a territory remaining occupied in 1 year



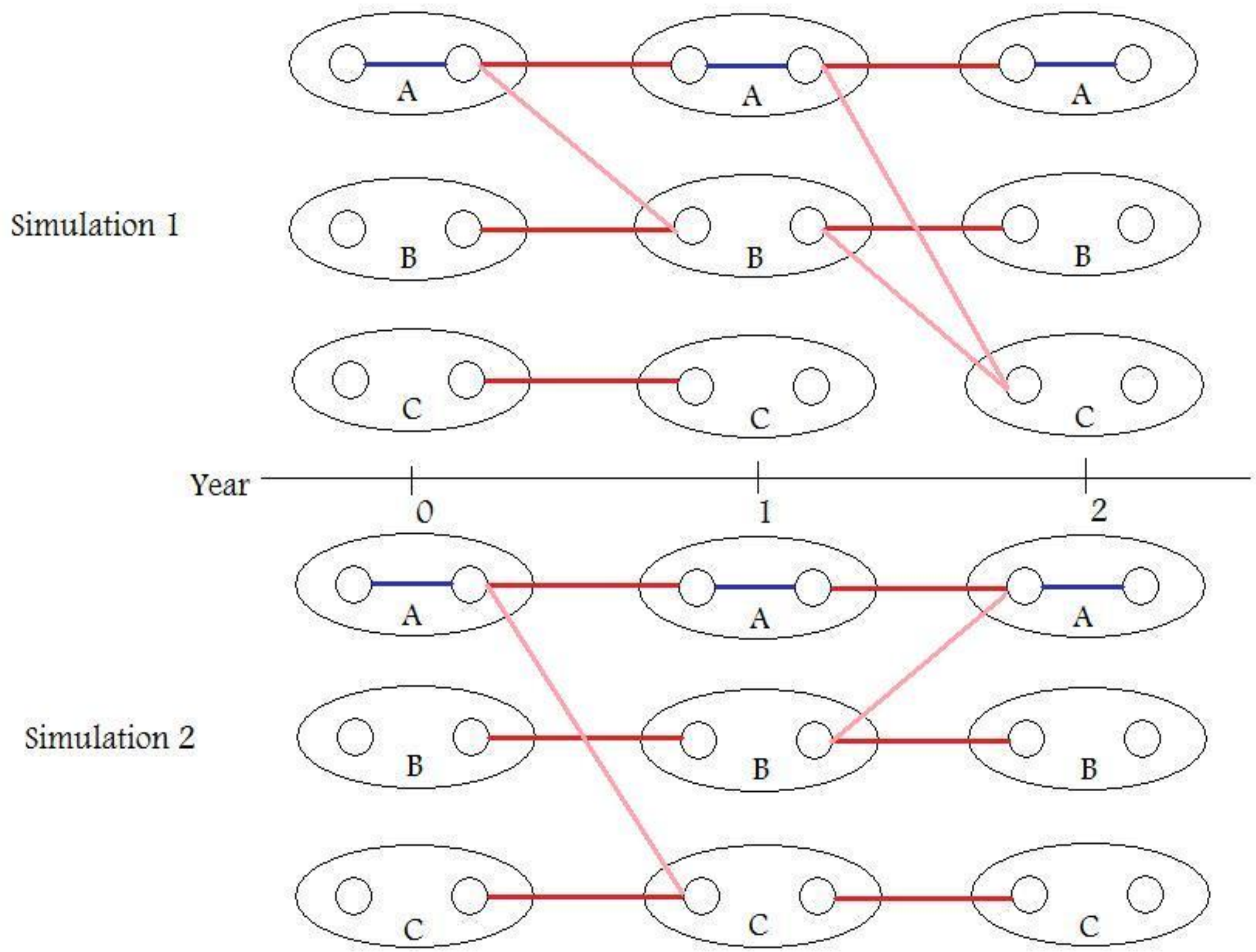

- A line from one oval to another represents the ability for a bird from the first territory to colonize the second
- Red lines indicate that if birds occupy a territory, then they will continue occupying it in the next time step
- Birds at C in year 1 in simulation 1 won't make it…

## Pink lines indicate the chance that one territory will occupy another



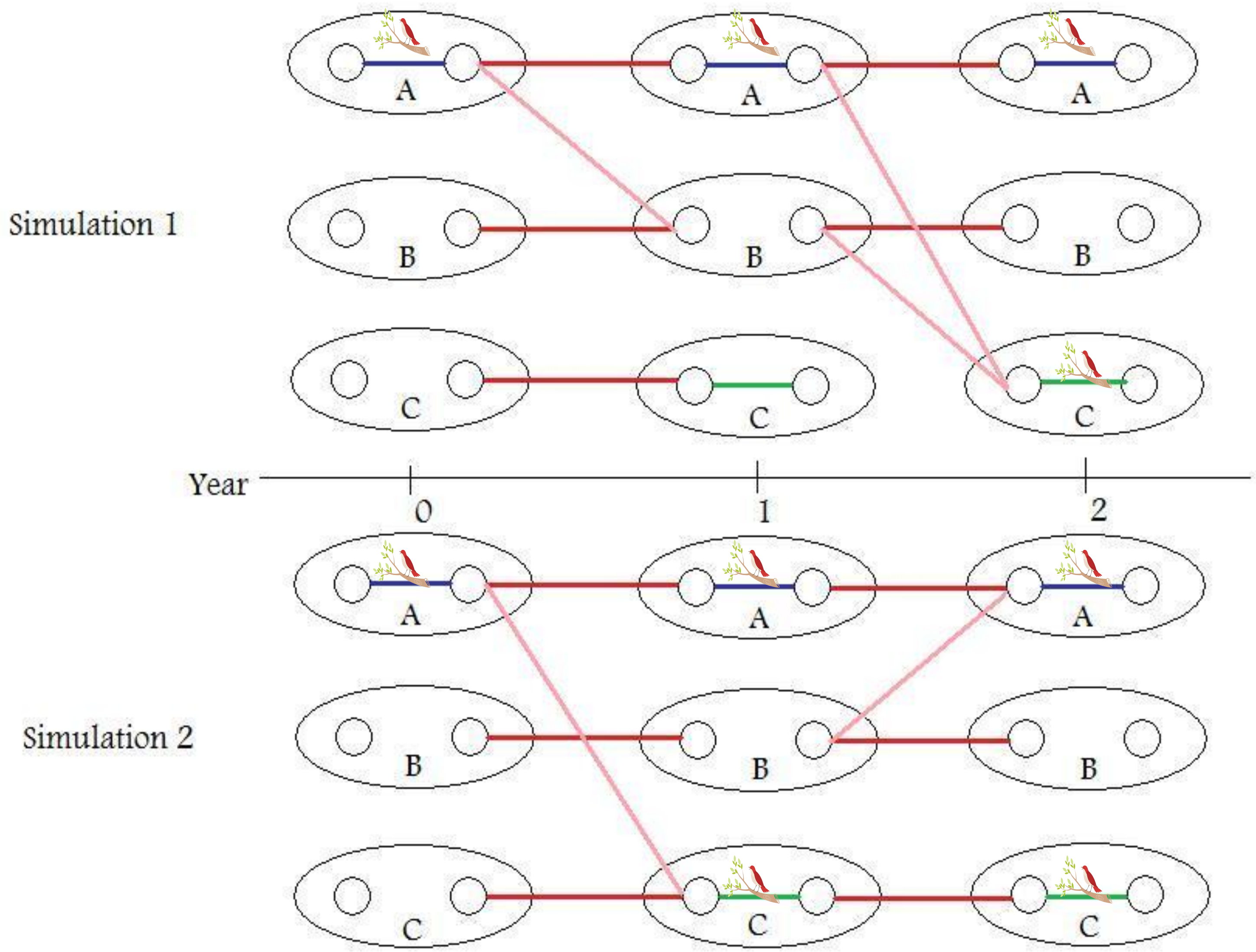

- Using data we can estimate the probability of a bird in one territory occupying another territory in one time step
- The pink lines represent the outcomes of the simulation using these probabilities
- If there are birds at B in year 1 in sim. 1, they will colonize C

Blue line represents territories already suitable (& occupied in example)




- The reason for having two nodes represent each territory is to indicate whether or not it is suitable
- If suitable, then there will be a line, allowing the birds to inhabit the territory from one time step to the next

Green lines indicate which territory which we should purchase




•Start with only territory A occupied

•Now we want to decide which territory to purchase, B or C?

•What maximizes average number of occupied nodes at time 2?

•In simulation 2, the birds from A can never get to B. In both simulations we can get to C, so C is better

# The Flow-Type IP Formulation

$$
\begin{aligned}
\text{maximize} \quad & (1/K)\sum_{k=1}^{K}\sum_{i\in\mathcal{R}} x^{ik}(i,T) \\
\text{subject to} \quad & \sum_{i\in\mathcal{R}}\sum_{t=1}^{T} b(i,t)y(i,t) \leq B, \\
& \sum_{t=1}^{T} y(i,t) \leq 1, \quad \forall i \\
& x^{rk}(i,t) \leq \sum_{s=1}^{t} y(i,s), \quad \forall\, r,k,i,t \\
& z^{rk}(i,j,t) \leq a^{k}(i,j,t), \quad \forall\, r,k,i,j,t \\
& \sum_{i\in\mathcal{R}} z^{rk}(i,j,t) = x^{rk}(j,t), \quad \forall\, r,k,j,t \\
& x^{rk}(i,t) = \sum_{j\in\mathcal{R}} z^{rk}(i,j,t+1), \quad \forall\, r,k,i,t \\
& x^{rk}(i,t),\ y(i,t),\ z^{rk}(i,j,t) \in \{0,1\} \quad \forall\, r,k,i,j,t
\end{aligned}
$$

Budget constraint

Purchase constraints

Suitability constraints

Colonization constraints

Flow constraints

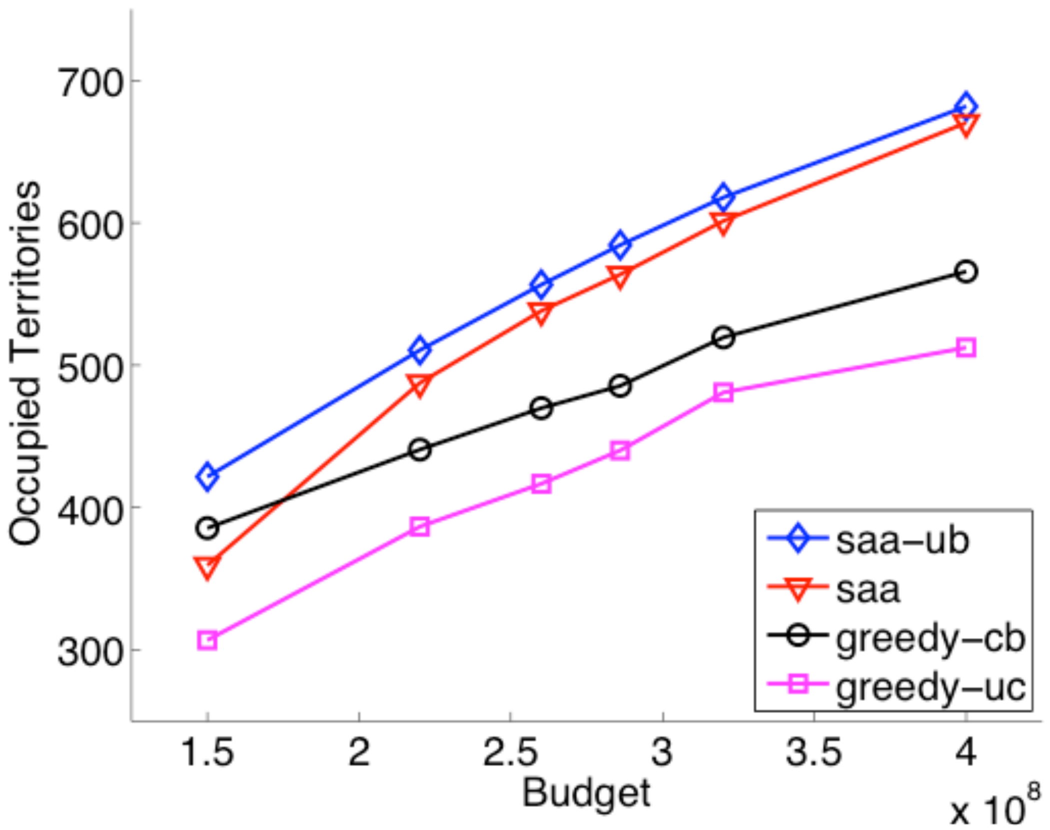
# Sample Average Approximation MIP vs Greedy

## Experimental Results

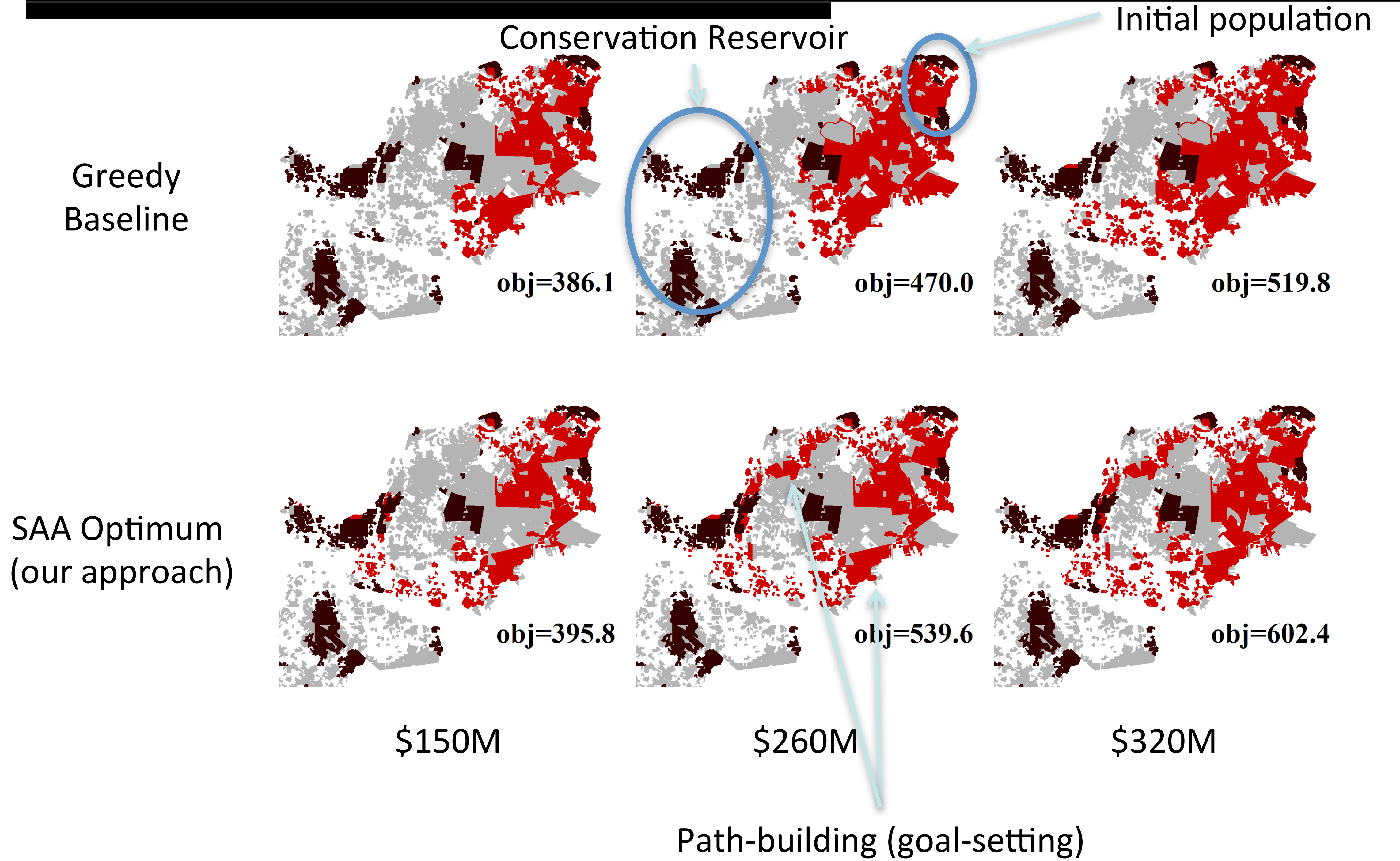
- In our experiments, we use two greedy algorithms as baselines.

- The MIP finds better solutions than the greedy baseline while proving near-optimality

- Results show dramatic improvement over greedy algorithm.



# Results




## Spatio-Temporal Aspects Make These Problems Hard

# Approximation algorithms for fragmenting a graph against a stochastically-located threat (WAOA 2011)

David Shmoys and Gwen Spencer

September 21, 2011

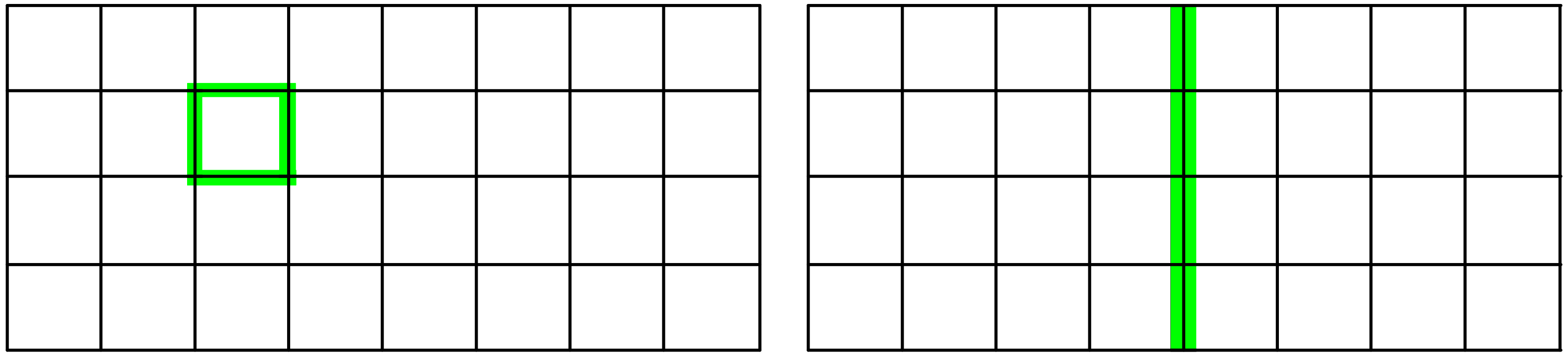
# A preventative approach to wildfire management

- Fuels build-up on landscape,
  fires threaten habitat and infrastructure:
  USFS interest in preventative fuel reductions.
- Current planning methods: don't simultaneously incorporate distributional information about ignition site and spatial information.
- Best fire simulation models seem too complex for optimization of stochastic objective.

**Goal:** Better understand spatial/stochastic interplay.
**Particularly:** Correct balance of prevention and real-time action?

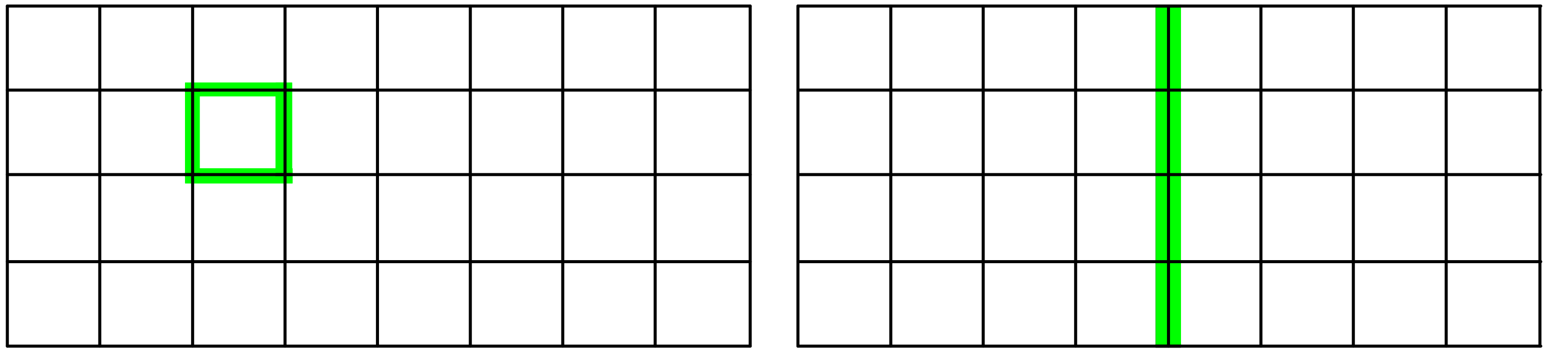
# What can we add to this prevention-planning tool box?

- Simple example (no fire ending events, assume that fuel treatment stops fire).



# What can we add to this prevention-planning tool box?

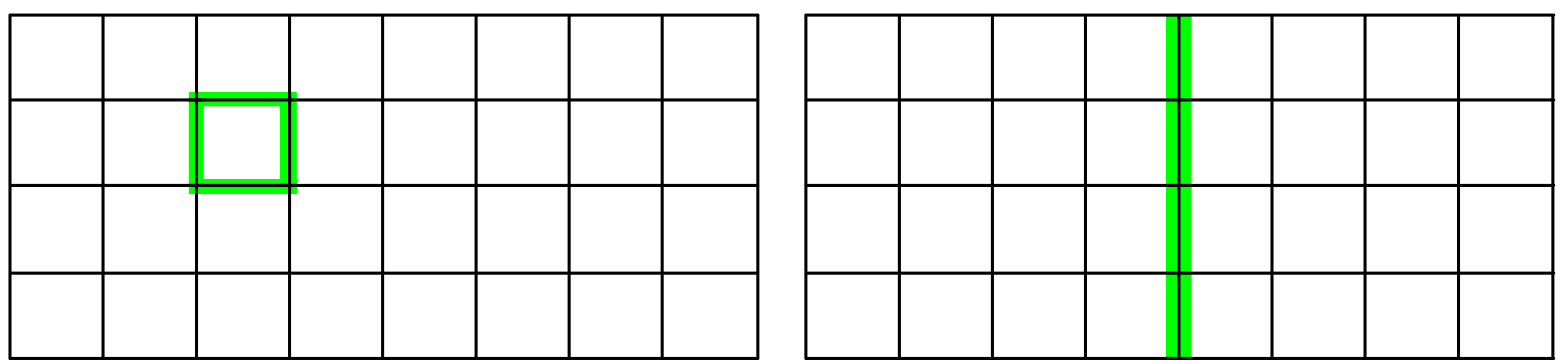
- Simple example (no fire ending events, assume that fuel treatment stops fire).



- **Uniform ignition probability:**
  single barrier beats parcel isolation by a factor of 8.

# What can we add to this prevention-planning tool box?

- Simple example (no fire ending events, assume that fuel treatment stops fire).



- **Uniform ignition probability:**
  single barrier beats parcel isolation by a factor of 8.
- **Concentrated ignition probability:**
  parcel isolation beats single barrier by a factor of 2.

## Exploring explicit dependence on distribution of fire ignition location

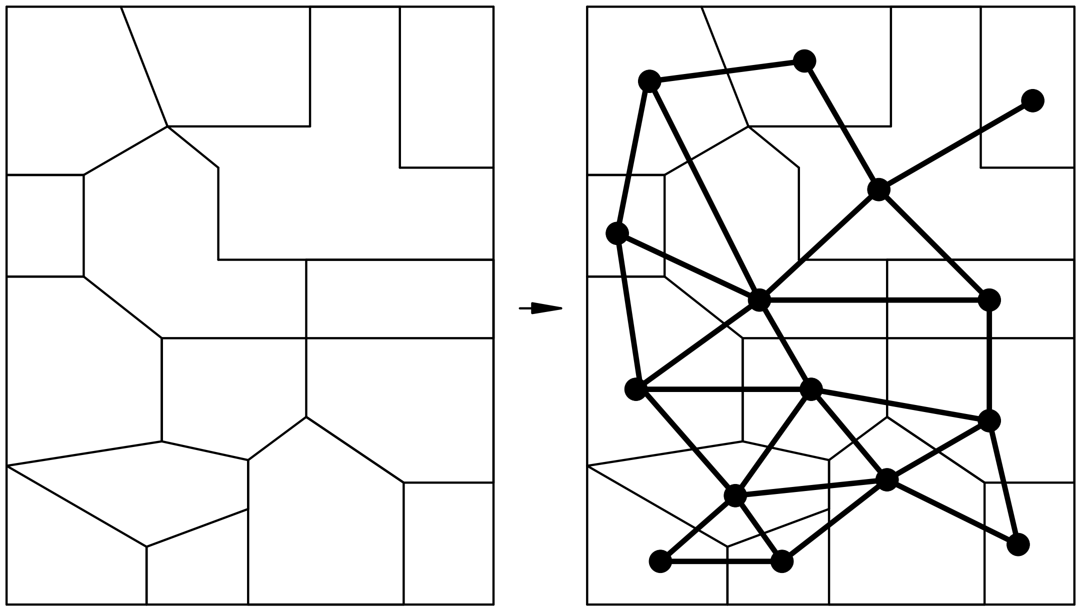
- **Distribution of ignition point matters in determining placement of preventative treatments.**
- **How can we use distributional info about ignition site explicitly?**

# This talk:

- **Propose:** Introduction of a stripped down model for budgeted placement of preventative treatments that explicitly incorporates probabilistic data on outbreak site.
- **Goal (near term):** Find scalable algorithms that perform provably close to the optimal solution for this problem.
- **Goal (longer term):** Extract design insight from methods and solution forms for simple stripped down model that can be used to inform decision making about the full complexity planning problem.

# The model: Stochastic Graph Protection

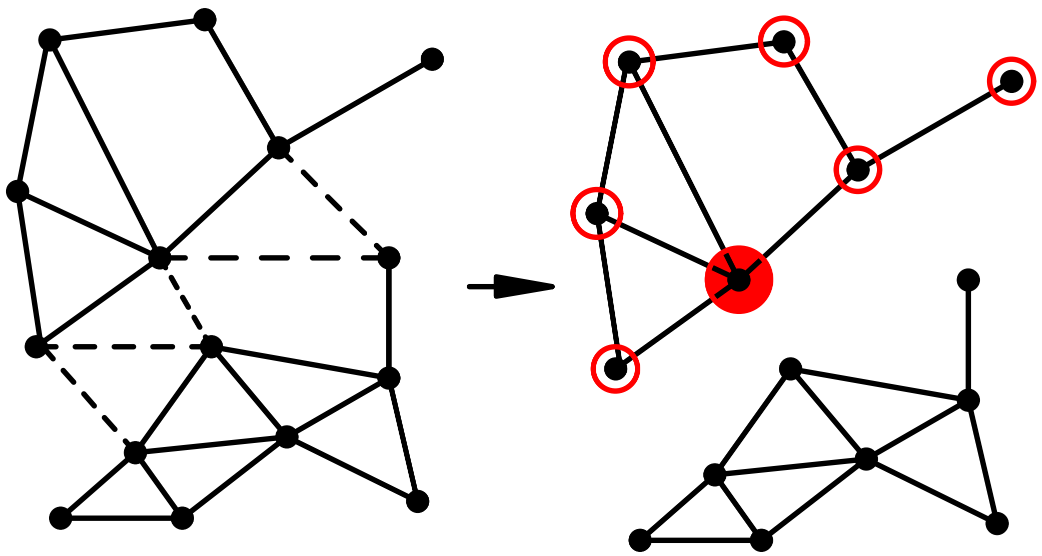
- A map becomes a graph:



- A node for each parcel, edges indicate *adjacent* parcels.
- Parcels have values, edges have costs, specified budget to spend, distribution over ignition sites.

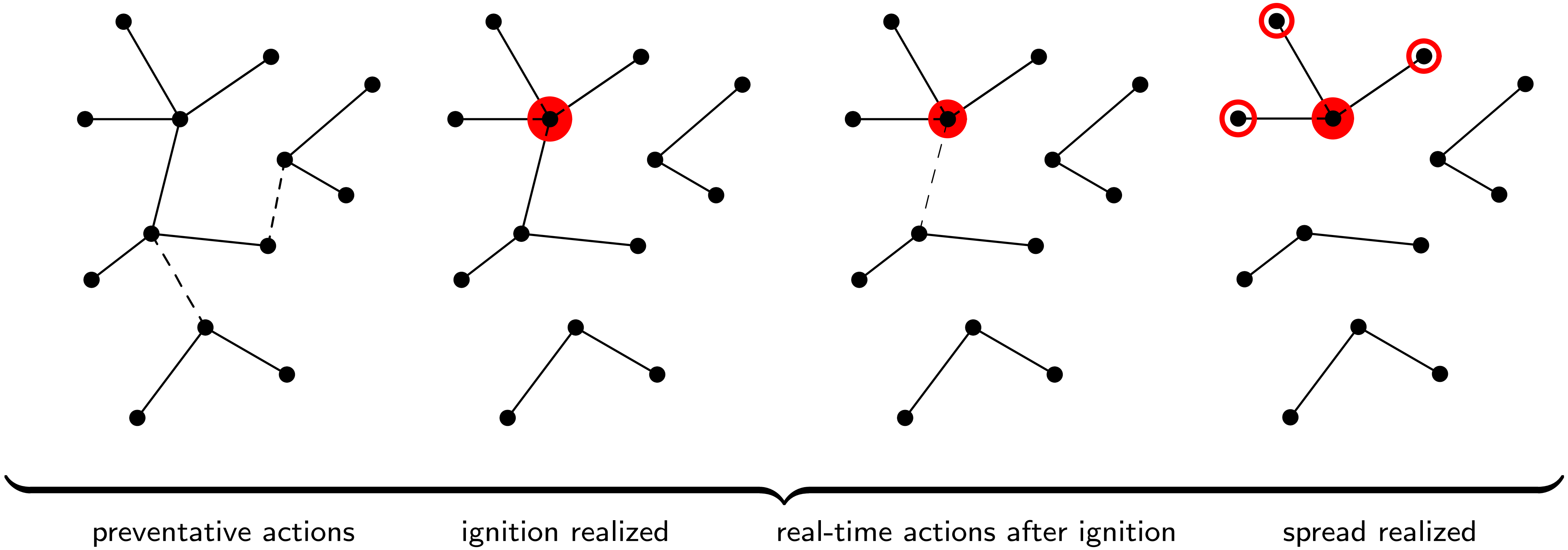
# The model: Stochastic Graph Protection Problem

- **Thematic:** Remove budget-limited set of edges, fire spreads through connected components. Protecting a node requires that every path from ignition point is blocked.



**Goal:** Spend budget removing edges from the graph to **max expected value of the landscape protected**

# 2-stage version: Preventative vs. Real-time




- Basic paradigm in multi-stage optimization: Acting in real time is more costly.

# Comments:

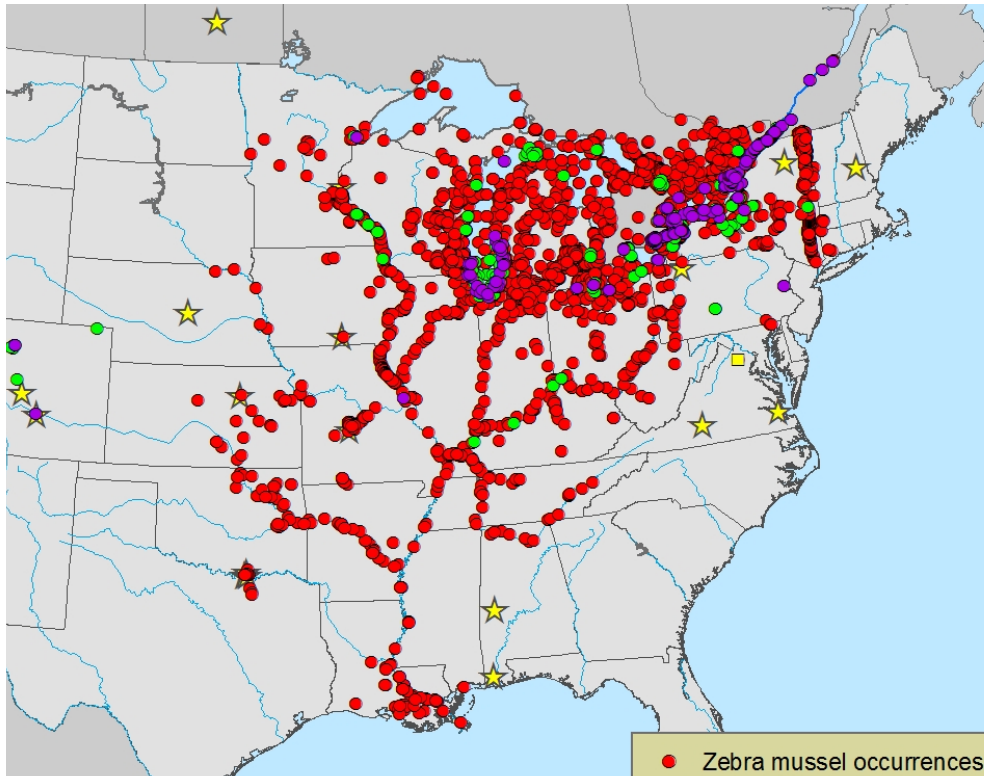
- **Stochastic Graph Protection is very general:**
  captures stochastic outbreak of harmful diffusive process in network.

  Steps toward isolating the outbreak are only effective if other paths to infection don't exist,
  variable costs (and variable inflation),
  variable loss (across nodes and scenarios).

- Other possible applications of this model:
  isolating outbreaks of pests or disease
  stemming the spread of invasive species

# Comments:

- **Restricted graph classes (e.g. trees) can be useful:** zebra mussels spreading through a river and stream system.




from CA.gov

# Results: 1-stage

| | restricted graph classes | general graphs (via [3],[4]) |
|---|---|---|
| stochastic, single source with probabilistic edges | trees: $(1-1/e, 1)$ Due to submodularity. | open |
| stochastic, single source | trees: $(1-(1-1/\delta)^{\delta}, 1)$ Reduce to MCKP, apply [1]. | $(1-(1-1/n)^n, O(\log n))$ |
| stochastic with constant support and constant source size | trees: $(1+\epsilon,\ 1)$ | $(1+\epsilon,\ O(\log n))$ |
| deterministic with arbitrary source size | bounded tree width: $(1+\epsilon,\ 1)$ | $(1+\epsilon,\ O(\log n))$ |
| deterministic with single source | bounded tree width: $(1+\epsilon,\ 1)$ [2] | $(1+\epsilon,\ O(\log n))$ [2] |

[1] Ageev, Sviridenko '04. [2] Hayrapetyan, Kempe, Pál, Svitkina '05. [3] Räcke '08.
[4] Engelberg, Könemann, Leonardi, Naor '06.

# Tool: Maximum Coverage subject to Knapsack Constraint

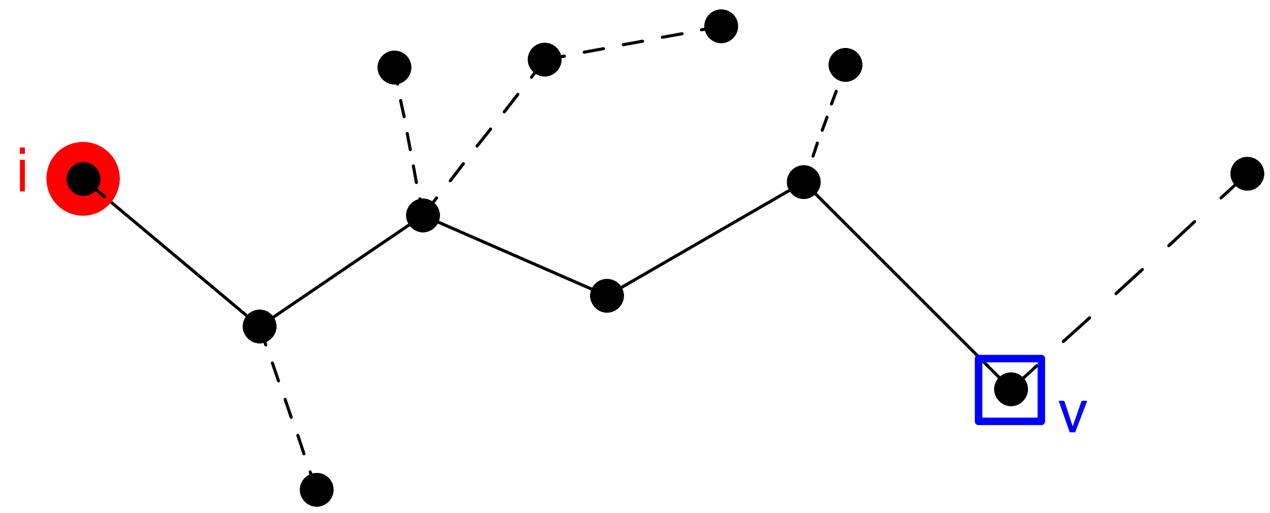
- **MCKP:**
  Family $F = \{S_j : j \in J\}$ of subsets of $I = \{1, 2, ..., n\}$
  subset weights $w_j$, item costs $c_i$, budget $B$:

  Find a budget-balanced set of items $X$ to

  $$\max \sum_{j:S_j \cap X \neq \emptyset} w_j.$$

- **Theorem.** (Ageev, Sviridenko '04)
  Exists a $(1 - (1 - \frac{1}{\delta})^{\delta})$-approximation algorithm for MCKP.

# 1-stage Stochastic GPP in trees: Connection to Maximum Coverage




- Choosing any edge on this path *covers* the pair (i,v).
- Path gives a set of items: including one of these items (or more) in solution accrues (value of v)$\times$ $p_i$.
- Apply Ageev and Sviridenko's MCKP Result: $(1 - (1 - \frac{1}{\delta})^{\delta})$-approximation. *wBSMC also.

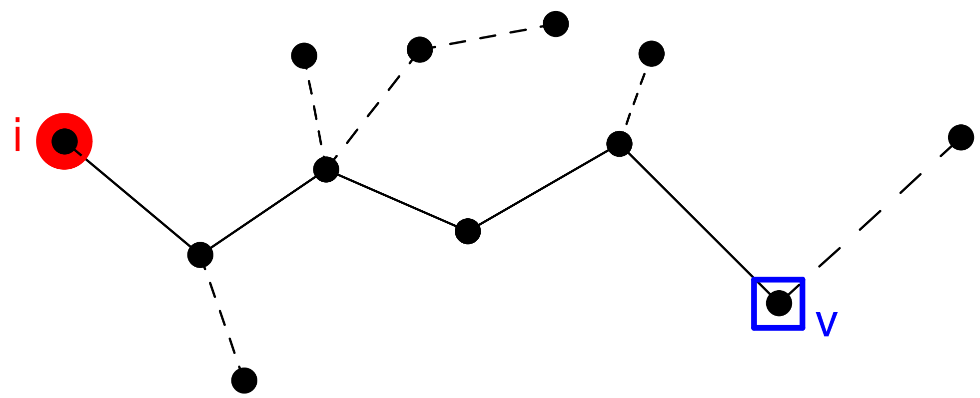
# Results: 2-stage

| | restricted graph classes | general graphs (via [3],[4]) |
|---|---|---|
| stochastic, single source | trees: $(1-(1-1/2\delta)^{2\delta}, 2)$<br>Via pipage rounding.<br><br>*Alternative:* $(0.387, 1)$ | constant # of scenarios<br>$\Rightarrow (1-(1-1/2n)^{2n}, O(\log n))$ |
| stochastic, single source with $(B_1, B_2)$ | trees: $(1-(1-1/2\delta)^{2\delta}, 1, 2)$<br>Via pipage rounding. | constant # of scenarios<br>$\Rightarrow (1-(1-1/2n)^{2n}, O(\log n), O(\log n))$ |

# The 2-stage problem

## Theorem

*There exists a bicriteria* $(1-(1-\frac{1}{2\delta})^{2\delta}, 2)$*-approximation algorithm in trees.*




- 
- Edge in $i, v$ path can *cover* in first stage or second stage. Consider these as separate items.
  $\Rightarrow$ Max Coverage objective.

# A LP for the 2-stage problem: Preventative vs. Realtime

$x_{iv} = 1$ if $v$ protected when $i$ ignites.
$y_e = 1$ if solution buys $e$ in first stage.
$z_e^i = 1$ if solution buys $e$ in second stage when $i$ ignites.

$$\max \sum_{(i,v)} (p_i v_{iv}) x_{iv} \quad \leftarrow \textit{max expected valued protected}$$

*protection constraints* $\rightarrow$ $$\sum_{e \in P(i,v)} y_e + \sum_{e \in P(i,v)} z_e^i \geq x_{iv} \quad \forall (i,v)$$

*budget constraints* $\rightarrow$ $$\sum_e y_e c_e + \sum_e z_e^i (M^{ie} c_e) \leq B \quad \forall i$$

$$x_{iv} \leq 1 \quad \forall (i,v)$$

# Comments and Directions:

- The analysis of several well-known 2-stage optimization results requires uniform inflation in the second stage.

  Are there simple algorithms for Stochastic Graph Protection that perform better when we **assume uniform inflation**?

- General problem in graphs:
  possible to **avoid resorting to capacity approximation** (and the associated $O(\log n)$ loss in budget guarantee)?

Thank you!