Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence

Robust Cuts Over Time: Combatting the Spread of Invasive Species with Unreliable Biological Control

Gwen Spencer*
Cornell University
Ithaca, NY 14853

Abstract

Widespread accounts of the harmful effects of invasive species have stimulated both practical and theoretical studies on how the spread of these destructive agents can be contained. In practice, a widely used method is the deployment of *biological control agents*, that is, the release of an additional species (which may also spread) that creates a hostile environment for the invader. Seeding colonies of these protective biological control agents can be used to build a kind of living barrier against the spread of the harmful invader, but the ecological literature documents that attempts to establish colonies of biological control agents often fail (opening gaps in the barrier). Further, the supply of the protective species is limited, and the full supply may not be available immediately. This problem has a natural temporal component: biological control is deployed as the extent of the harmful invasion grows. How can a limited supply of unreliable biological control agents best be deployed over time to protect the landscape against the spread of a harmful invasive species?

To explore this question we introduce a new family of stochastic graph vaccination problems that generalizes ideas from social networks and multistage graph vaccination. We point out a deterministic $(1-1/e)$-approximation algorithm for a deterministic base case studied in the social networks literature (matching the previous best randomized $(1-1/e)$ guarantee for that problem). Next, we show that the randomized $(1-1/e)$ guarantee (and a deterministic 1/2 guarantee) can be extended to our much more general family of stochastic graph vaccination problems in which vaccinations (a.k.a. biological control colonies) spread but may be unreliable. For the non-spreading vaccination case with unreliable vaccines, we give matching results in trees. Qualitatively, our extension is from computing "cuts over time" to computing "robust cuts over time."

Our new family of problems captures the key tensions we identify for containing invasive species spread with unreliable biological control agents: a robust barrier is built over time with unreliable resources to contain an expanding invasion.

*gwenspencer@gmail.com. Work supported in part by an NSF Graduate Research Fellowship, and under grants no. CCR-0635121, DMS-0732196, CCF-0832782, CCF-1017688.
Copyright © 2012, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Introduction

The devastating impacts of *invasive* (a.k.a. non-native) species are being increasingly documented around the globe. Both plant and animal invaders can spread across landscapes, quickly overwhelming local ecosystems through direct competition for resources and predation. In addition to ecological concerns, economic forecasts predict massive ramifications on renewable natural resource industries (e.g. timber) and residential land values (Kovacs et al. 2010). These accounts have stimulated both practical and theoretical studies on how the spatial spread of these destructive agents can be contained (Kovacs, Haight, and Mercader 2012) (Shmoys and Spencer 2011).

A commonly-practiced technique for invasive species containment is the introduction of a *biological control agent*, that is, another species (which may also spread) whose presence renders the habitat inhospitable to the harmful invasive species. A problem explored in the ecological literature is that introductions of these control species often fail to establish viable colonies (Shea and Possingham 2000). The probability of establishment is often related to the size of the seed colony (Shea and Possingham 2000), but since supplies of the biological control agent are usually limited, and breeding additional agents takes time, land managers must proceed with unreliable deployments. In addition to the many unresolved scientific and ecological questions about biological control strategies (particulary as regards unintended negative consequences of control species introductions), there are basic mathematical issues here that are not well understood. Where should a limited supply of unreliable biological control agents be deployed over time to contain the spread of a harmful invader?

To explore this question we introduce a new family of stochastic graph vaccination problems that generalizes ideas from social networks and multistage graph vaccination.

Literature Review. Hartnell's Firefighter Problem (Hartnell 1995) concerns the containment of an infection outbreak in a graph when the amount of action per time step is limited. In each time step all neighbors of an infected node become infected. The planner has the option to vaccinate k nodes per time step (so that they will never become infected). Where should these vaccinations be placed at each time-step to minimize the total number of nodes infected? Qualitatively, the

goal is to compute the best "cut over time."

In addition to literal diseases that can be spread between contacts in social networks, recent work on disease spread and vaccination in networks has proposed the notion of harmful ideas or ideology spreading through social networks. The antidote to harmful ideology is positive ideology or ideas which also spread through the network vaccinating people who are exposed to them against the harmful ideology (Anshelevich et al. 2009). Once an individual is convinced by either harmful or positive ideology we suppose that they can no longer be converted to the opposing viewpoint. Considering this setting with an additional temporal component has motivated interest in a spreading-vaccination version of the Firefighter problem.

For the spreading-vaccination version of the Firefighter problem (in which vaccination also spreads to neighbors at the same rate as infection) Anshelevich, Chakrabarty, Hate, and Swamy show that the problem reduces to maximizing a submodular function subject to a partition-matroid constraint (Anshelevich et al. 2009). They use this reduction in conjunction with the recent result of Calinescu, Chekuri, Pál and Vondrák (Calinescu et al. 2007) for maximizing a submodular function subject to a matroid-constraint to give a randomized $(1-1/e)$-approximation (or to get a deterministic $1/2$ approximation by applying a greedy result of Fisher, Nemhauser and Wolsey (Fisher, Nemhauser, and Wolsey 1978)).

In fact, we note that since (Anshelevich et al. 2009) show that the spreading-vaccination Firefighter objective is a coverage function, a deterministic $(1-1/e)$-approximation can be obtained immediately by reducing the problem to maximum coverage subject to group-constraints and applying a decade-old result of Ageev and Sviridenko for that problem (Ageev and Sviridenko 2004). This deterministic $(1-1/e)$ guarantee is tight (unless $NP \subseteq DTIME(n^{\mathtt{polylog}(n)})$ due to another result from (Anshelevich et al. 2009)), and it matches the current best guarantee in trees for the non-spreading Firefighter Problem due to Cai, Verbin, and Yang (Cai, Verbin, and Yang 2008).

In the ecology literature, Shea and Possingham give curves that describe probability of control colony survival as a function of initial population (Shea and Possingham 2000), and point out that biological control resources may not all be available at the start of a suppression effort. Through field experiments, Norris, Memmott and Lovell tune parameters impacting colony survival of an insect control, thrips (*Sericothrips staphylinus*), for the models in (Shea and Possingham 2000); their study shows dependence on stochastic weather events (rainfall) (Norris, Memmot, and Lovell 2002).

As mentioned in (Shmoys and Spencer 2011), tree landscapes are used to study invasive-species spread through stream and river systems (for example, (Cumming 2002) explores how different stream branching affects invasions). Invasions of tree-landscape topologies also occur in the riparian-vegetation zones bordering river systems. For example, Tamarisk is an exotic noxious weed that invades such zones, driving out local wetland plants; the Colorado Department of Agriculture has suppressed Tamarisk invasion using the tamarisk leaf beetle, *Diorhabda elongata* along the Dolores, Colorado, Yampa and Green Rivers (Colorado.Dept.Agriculture 2009).

Our Results. The conditions required to apply the result of Calinescu, Chekuri, Pal and Vondrák (CCPV) are more general than are required for the spreading-vaccination Firefighter Problem: in exploring how much can we generalize the spreading-vaccination Firefighter Problem and still benefit from reducing to maximizing a submodular function subject to a partition-matroid constraint, we introduce a natural new family of general stochastic cut problems that capture the key tensions in containing invasive species with unreliable biological control agents. Applying CCPV and (Fisher, Nemhauser, and Wolsey 1978)), we immediately obtain randomized $(1-1/e)$ and deterministic 1/2-approximations for the spreading cases in graphs and for the non-spreading cases in trees).

Our Generalization: Unreliability. At the conclusion of their paper, Anshelevich, Chakrabarty, Hate, and Swamy (Anshelevich et al. 2009) raise the very natural question of what happens when vaccination is less virulent than infection. In particular, they mention why a positive spread rate for vaccination of less than 1 voids a reduction similar to the rate 1 case. We introduce an alternative way of blunting the strength of vaccination: spread of vaccination remains deterministic at rate 1 but we allow the possibility that a vaccine will be faulty with some probability (which is given in the input). First we give a direct generalization of the Firefighter Problem (in which k interchangeable vaccines are distributed per time step), then we generalize further.

One interesting model among our new class considers vaccines that become better with time (their failure probability decreases) but which may be deployed at most once over the time horizon; what is the correct balance between acting early with unreliable resources when the extent of the infection is small and acting later with more reliable resources once the infection is more extensive? This corresponds to the tradeoff between immediately deploying small biological control colonies (which have high failure probability) vs. postponing deployment while the invasive species spreads in order to breed larger and more reliable biological control colonies under protected conditions.

We describe our models as offline problems: all decisions are made without knowing which vaccines will be realized as effective. In the online setting where reliability/unreliability is realized and observed at each time step, simply extracting nodes which will be covered by a vaccination that has already been realized as effective (which is polynomial-time computable) will give a new problem of an identical type (so that our approximation guarantees hold from that time step forward).

Qualitatively, our new notion could be described as computing "robust cuts over time" in which unreliable vaccine resources are deployed over a series of time steps in order to cut (or cover, in the spreading case) a graph against the spread of an infection. In terms of idea-spread through social networks we could describe this as a problem of wag-

ing an ideological war when we are unsure which positive ideologies we plant with "go viral." An interesting feature of this expanded model is that the optimal solution may repeatedly target nodes which are in some sense "highly connective." In these features our results add to the exploration begun in the seminal paper of Kempe, Kleinberg, and Tardos on maximizing influence spread in social networks (Kempe, Kleinberg, and Tardos 2003) which was motivated by emerging applications in viral marketing. Our results for the non-spreading cases in trees are closely related to the recent work of Shmoys and Spencer (Shmoys and Spencer 2011) on containing stochastic outbreaks of invasive species in tree landscapes via imperfect (a.k.a. probabilistic) edge removal; that work addresses only connectivity vs. disconnectivity and lacks the natural temporal nature of an expanding invasion that we address here.

We show that for the cases of spreading-vaccination and non-spreading vaccination in trees the general stochastic models we describe reduce to maximizing a submodular objective subject to a partition-matroid constraint. Thus, the result of Calinescu, Chekuri, Pal and Vondrák (CCPV) (Calinescu et al. 2007) gives a randomized $(1-1/e)$-approximation (or we can get a deterministic $1/2$ approximation by applying an older result of Fisher, Nemhauser and Wolsey (Fisher, Nemhauser, and Wolsey 1978)). We will need that the failure of vaccines are independent events so that we can compute the values required by the CCPV method in polynomial time.

From Classical Firefighter Models to Stochastic Graph Vaccination

We are given a directed graph $G = (V,E)$ and a source node s. Denote $|V|$ by n, and $|E|$ by m. All nodes in the graph can have one of three states: *infected*, *vaccinated*, or *vulnerable*. At time $\tau = 0$, node s is infected and all other nodes are *vulnerable*. Once a node has become *infected* or *vaccinated* its state will never change. The goal of the planner is to maximize the expected number of members of the network that are protected from infection. (Hartnell 1995)

- **Classical (non-spreading):** At time $\tau = i > 0$ at most $k \leq n$ vaccines can be deployed at vulnerable nodes, where each vaccine has probability 1 of effectiveness (and probability 0 of not being effective). When a vaccine is realized as effective, then the node it was deployed at becomes *vaccinated*. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.
- **Classical (spreading-vaccination):** At time $\tau = i > 0$ at most $k \leq n$ vaccines can be deployed at vulnerable nodes, where each vaccine has probability 1 of effectiveness (and probability 0 of not being effective). When a vaccine is realized as effective, then the node it was planted at becomes *vaccinated*. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *vaccinated* node become *vaccinated*. Then, in the same time step, all remaining *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.

First we give the direct stochastic generalizations of these models (retaining the properties that actions within each time step are identical, and that there is no dependence on location of deployment):

- **Stochastic (non-spreading):** At time $\tau = i > 0$ at most $k \leq n$ vaccines can be deployed at vulnerable nodes, where each vaccine has independent probability p_i of effectiveness (and probability $1-p_i$ of not being effective). When a vaccine is realized as effective, then the node it was deployed at becomes *vaccinated*. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.
- **Stochastic (spreading-vaccination):** At time $\tau = i > 0$ at most $k \leq n$ vaccines can be deployed at vulnerable nodes, where each vaccine has independent probability p_i of effectiveness (and probability $1-p_i$ of not being effective). When a vaccine is realized as effective, then the node it was planted at becomes *vaccinated*. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *vaccinated* node become *vaccinated*. Then, in the same time step, all remaining *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.

While we've specified that $k \leq n$ above, in fact any k that is bounded above by a polynomial in n will work. Our choice that vaccination prevails when exposure to vaccination and infection is simultaneous is not required for any of the described guarantees.

Figure 1 below motivates our next generalization: we relax the idea that there is only a single timestep in which a particular vaccine is effective.

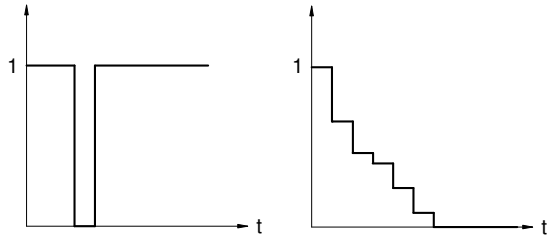

Figure 1: **Examples of $f_j(x,t)$:** On the left a vaccine fails with probability 1 except during timestep 3 in which it is effective with probability 1. On the right, a vaccine has a failure probability that decreases with time. Not shown here: the dependence of $f_j(x,t)$ on the node x at which j is deployed can be used to capture that some sites are more suitable for the establishment of biological control colonies.

- **General Stochastic Graph Vaccination (non-spreading):** The input gives a set of vaccines J. For each $j \in J$ the input gives a function $f_j(x,t) : \{1,2,...,n\} \times \{1,2,...,n\} \rightarrow [0,1]$ that describes the effectiveness probability of j if it is deployed at node x at time t. Each vaccine is deployed at some node at some time step, and effectiveness of each vaccine is realized independently. When a vaccine is realized as effective, then the node it was deployed at becomes *vac-*

cinated. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.

- **General Stochastic Graph Vaccination (spreading-vaccination):** The input gives a set of vaccines J. For each $j \in J$ the input gives a function $f_j(x,t) : \{1,2,...,n\} \times \{1,2,...,n\} \rightarrow [0,1]$ that describes the effectiveness probability of j if it is deployed at node x at time t. Each vaccine is deployed at some node at some time step, and effectiveness of each vaccine is realized independently. When a vaccine is realized as effective, then the node it was planted at becomes *vaccinated*. In time step $i+1$ all *vulnerable* nodes reachable by a single edge from any *vaccinated* node become *vaccinated*. Then, in the same time step, all remaining *vulnerable* nodes reachable by a single edge from any *infected* node become *infected*.

It is immediately obvious that the classical problems are special cases of the stochastic problems (setting $p_i = 1$ for all i), but seeing that the stochastic problems are special cases of the general stochastic problems requires a slightly subtle observation.

Since the diameter of the graph is at most n, at time $\tau = n$ every node of the graph will either be infected or we will know that it will never be infected (Anshelevich et al. 2009). Thus, in the stochastic problem there is no need to consider any vaccine deployments at time steps beyond n. This is true for both the spreading-vaccination and non-spreading problems. Consider an instance of the stochastic problem: for each time-step i, create k copies of vaccine j which has $f_j(x,i) = p_i$ for all $x \in V$, and $f_j(x,i') = 0$ for all $i' \neq i$. This gives an input for the most general problem with kn vaccines.

Notice that the stochastic elements we introduce can cause the form of the solutions to be very different than in the deterministic vaccination case: in the typical spreading-vaccination Firefighter model once a node is vaccinated its status is known conclusively, and as each time step elapses expanding neighborhoods around it are removed from the set that could possibly be vaccinated in an optimal solution, while in our stochastic model the optimal solution may repeatedly target some "important nodes" in a single time step or repeatedly over a series of time steps. See Figure 2 for an example.

Further, if a larger node cut could protect a large portion of the network but cannot be reliably imposed with the vaccines available, a moderate portion of the network may be sacrificed to more reliably impose a smaller cut further from the source. The most general model softens the hard limits on how much vaccination can be performed per time step: the freedom to design $f_j(x,t)$ allows exploration of a number of interesting tensions: vaccines that each become more powerful with time but can be deployed only once, vaccines that will be most effective within a particular time step (as in the stochastic case), but can also be shifted to neighboring time steps in exchange for a penalty on their reliability, etc.

Hardness

The classical case of non-spreading Firefighter is NP-complete in trees. The current best guarantee in general trees is a deterministic $(1-1/e)$-approximation due to Cai, Verbin, and Yang (Cai, Verbin, and Yang 2008). Some polynomial-time algorithms exist for restricted classes of trees. We give a randomized $(1-1/e)$-approximation for the General Stochastic Graph Vaccination Problem (non-spreading) described above in trees.

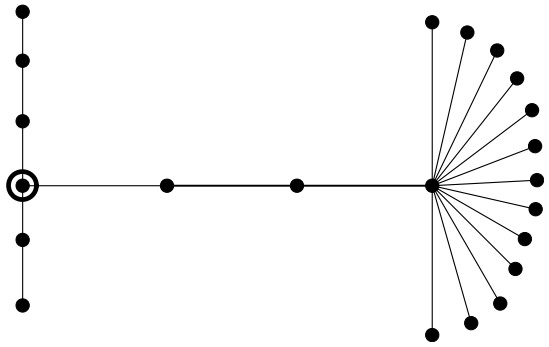

Figure 2: **Repeated-Targeting Example:** The initially-infected node is circled. The optimal solution to the robust-cuts-over-time problem will repeatedly target the long path leading to the star.

The classical case of the spreading-vaccination Firefighter Problem is hard to approximate within $c > (1-1/e)$ unless $NP \subseteq DTIME(n^{\texttt{polylog}(n)})$ (Anshelevich et al. 2009). We'll give a randomized $(1-1/e)$-approximation for the General Stochastic Graph Vaccination Problem with spreading vaccination.

Reductions to Maximum Coverage Subject to Probabilistic Element Failure

Theorem 1. *The non-spreading case of the General Stochastic Graph Vaccination Problem in trees reduces in polynomial time to the problem of maximizing a (polynomial-time computable) submodular function subject to a partition-matroid constraint. Thus, applying CCPV (Calinescu et al. 2007) gives a randomized* $(1-1/e)$*-approximation algorithm, and applying (Fisher, Nemhauser, and Wolsey 1978) gives a deterministic* $1/2$*-approximation algorithm.*

Theorem 2. *The spreading-vaccination case of the General Stochastic Graph Vaccination Problem reduces in polynomial time to the problem of maximizing a (polynomial-time computable) submodular function subject to a partition-matroid constraint. Thus, applying CCPV (Calinescu et al. 2007) gives a randomized* $(1-1/e)$*-approximation algorithm, and applying (Fisher, Nemhauser, and Wolsey 1978) gives a deterministic* $1/2$*-approximation algorithm.*

In fact we will show that both problems reduce in polynomial-time to Maximum Coverage subject to probabilistic set failure (which we'll show is a submodular function, similar to the argument in (Shmoys and Spencer 2011)) subject to a partition-matroid constraint.

Maximum Coverage subject to probabilistic set failure:

Let S be a collection of sets $S_1, S_2, ..., S_n$ over a ground set of elements Y. Each set S_i has an associated independent failure probability p_i. The objective is to choose a set $O \subseteq S$ such that the expected cardinality of the union of sets realized from O according to the p_i is maximized. *With a partition-matroid constraint:* For some partition of S into $P_1, ..., P_q$, at most k_j sets from P_j can be included in O.

Reduction 1: Non-spreading General Stochastic Graph Vaccination in trees (nSGVt) to Maximum Coverage subject to probabilistic set failure with partition-matroid constraint (pMCPM).

Let r denote the node where the infection starts. For each node x in T: let d denote the distance from x to r. The ground set of elements in the instance we construct is the set of nodes of tree T which we'll call Y. For every pair of $j \in J$ and $t \in \{1, 2, ...d\}$, create a set S_{jxt} which contains all nodes y whose shortest path to r contains x. Let $p_{jxt} = 1 - f_j(x,t)$. This gives at most $|J| \cdot |V| \cdot |V|$ sets which can each be computed in polynomial time. Define a partition of the collection of sets by letting all sets which have a common j be in a single partition piece, and specify that at most one set can be selected from each partition piece.

Let O denote the optimal solution of the resulting pMCPM instance, and use it to construct a solution for nSGVt as follows. If $S_{jxt} \in O$ then deploy vaccine j at node x at time t. Since O obeys the partition-matroid constraint this constructed solution deploys vaccine j at most once (at some particular node, at some particular time). What is the expected number of nodes saved from infection by the constructed strategy? Using linearity of expectation, it is the sum over all $y \in V$ of the expectation of an indicator variable that is 1 when y is saved, and 0 otherwise.

A node y is saved from infection exactly when some node on the path from r to y is *vaccinated effectively* before the infection arrives at that node; that is, when not every vaccination attempt on nodes on the r to y path between time 0 and $d(r,y)$ fails. By our construction, this event (and its probability) corresponds exactly to the event that y is in some non-failing set that contains it from O. Thus, the expected number of nodes saved from infection is exactly the expected cardinality of union of realized sets from O.

In the reverse direction, every valid solution for nSGVt defines a valid solution for the pMCPM instance: if vaccine j is to be deployed at node x at time t, include S_{jxt} in the solution. Since each vaccine is deployed at most once, the partition-matroid constraint is obeyed. Because the objective value exactly coincides, if the solution we constructed from O were not optimal for the nSGVt instance, it would imply that O was not optimal for the pMCPM instance.□

In the following reduction we add a probabilistic component to an observation about coverage of nodes by expanding vaccinated sets from (Anshelevich et al. 2009), then exploit the fact that only $|V|$ time steps must be considered so that the partition can force a single deployment of a vaccine (since the requirement that only k vaccines are deployed per time step need no longer be explicitly articulated via a partition-matroid constraint as in (Anshelevich et al. 2009)).

Reduction 2: Spreading-vaccination General Stochastic Graph Vaccination (sSGV) to Maximum Coverage subject to probabilistic set failure with partition-matroid constraint (pMCPM).

Let r denote the node where the infection starts. For any two nodes x and y, let $d(x,y)$ denote the distance from x to y. The ground set of elements in the instance we construct is the set of nodes of G which we'll call Y.

For every triple (j,x,t) of $j \in J$, $x \in V$, and $t \in \{1, 2, ..., n\}$, create a set S_{jxt} which contains all nodes y for which $d(x,y) \leq d(y,r) - t$. Let $p_{jxt} = 1 - f_j(x,t)$. This gives at most $|J| \cdot |V| \cdot |V|$ sets which can each be computed in polynomial time. Define a partition of the collection of sets by letting all sets which have a common j be in a single partition piece, and specify that at most one set can be selected from each partition piece.

Let O denote the optimal solution of the resulting pMCPM instance, and use it to construct a solution for sSGV as follows. If $S_{jxt} \in O$ then deploy vaccine j at node x at time t. Since O obeys the partition-matroid constraint this constructed solution deploys vaccine j at most once (at some particular node, at some particular time). What is the expected number of nodes saved from infection by the constructed strategy? Using linearity of expectation, it is the sum over all $y \in V$ of the expectation of an indicator variable that is 1 when y is saved, and 0 otherwise.

A node y is saved from infection exactly when there exists some vaccine deployment that is realized as effective at node x and time t such that vaccination spreading from x starting at t reaches y before infection reaches y. That is, when $d(x,y) \leq d(r,y) - t$ and a vaccine deployed at x at time t is realized as effective. By our construction, this event (and its probability) corresponds exactly to the event that y is in some non-failing set that contains it from O. Thus, the expected number of nodes saved from infection is exactly the expected cardinality of union of realized sets from O.

In the reverse direction, every valid solution for sSGV defines a valid solution for the pMCPM instance: if vaccine j is to be deployed at node x at time t, include S_{jxt} in the solution. Since each vaccine is deployed at most once, the partition-matroid constraint is obeyed. Because the objective value exactly coincides, if the solution we constructed from O were not optimal for the sSGV instance, it would imply that O was not optimal for the pMCPM instance.□

The following result is similar to that in described by analogy in (Shmoys and Spencer 2011) for a knapsack version; for completeness, we include it here.

Theorem 3. *Maximum Coverage subject to independent Probabilistic Set Failure is a submodular function.*

Proof. Each S_i is an element our solution can purchase. Let $E(\cdot)$ denote the objective. To prove submodularity we will establish that the law of diminishing returns holds: for an arbitrary S_i, if $A \subseteq B \subseteq S$, then

$$E(A \cup S_i) - E(A) \geq E(B \cup S_i) - E(B).$$

For a set $R \subseteq S$ which does not contain S_i we can say that the probability that S_i fails is 1. After S_i is added to either A or B the probability that S_i fails is p_i. Before S_i is "added" either: neither A or B contain S_i, or $S_i \in A$ (and $S_i \in B$), or $S_i \notin A$ and $S_i \in B$, or $S_i \notin A$ and $S_i \notin B$.

In every case, the probability that S_i fails when A is the solution is at least the probability that S_i fails when B is the solution. Letting $\wp_{S_i}(R)$ denote the probability of failure of S_i as a function of the set R:

$$\wp_{S_i}(A) \geq \wp_{S_i}(B) \Rightarrow$$
$$\wp_{S_i}(A) - \wp_{S_i}(A \cup S_i) \geq \wp_{S_i}(B) - \wp_{S_i}(B \cup S_i)$$

Next, focus on a particular element y from the ground set. Either S_i does not contain y (in which case the addition of S_i leaves the expected coverage of y unchanged) or it does. For the later case, the probability of every one of the other sets containing y failing (excepting S_i) is at least for A what it is for B. Let $P(R)$ denote the probability that every one of the sets containing y fails when the solution is R. Then $P(A) \geq P(B)$ gives that

$$\begin{aligned}
&P(A)(\wp_{S_i}(A) - \wp_{S_i}(A \cup S_i)) \geq \\
&\qquad P(B)(\wp_{S_i}(B) - \wp_{S_i}(B \cup S_i)) \\
&\Rightarrow P(A)\wp_{S_i}(A) - P(A)\wp_{S_i}(A \cup S_i) \geq \\
&\qquad P(B)\wp_{S_i}(B) - P(B)\wp_{S_i}(B \cup S_i). \\
&\Rightarrow (1 - P(A)\wp_{S_i}(A \cup S_i)) - (1 - P(A)\wp_{S_i}(A)) \geq \\
&\qquad (1 - P(B)\wp_{S_i}(B \cup S_i)) - (1 - P(B)\wp_{S_i}(B)). \\
&\Rightarrow E(A \cup S_i) - E(A) \geq E(B \cup S_i) - E(B).
\end{aligned}$$

The second to last line compares the changes in the probability that y is covered which results when S_i is added to A and when S_i is added to B. The final inequality follows from simply summing the change in expected coverage over all y in the ground set (including those for which the addition of S_i caused no changes in the probability of coverage). This establishes submodularity. □

Running CCPV

The method described in Calinescu, Chekuri, Pal and Vondrák (Calinescu et al. 2007) requires the ability to sample the value of the submodular function being maximized in polynomial time. Though for a given solution $A \subseteq S$ there are exponentially-many possible realizations of set failures, the objective can be computed in polynomial time by taking advantage of the independence of the set failures and the form of the coverage function. In particular, the expected size of the union of the sets realized as effective is just the sum of the expectations of indicators over all items of the ground set. For each such item, y, the probability of coverage is simply 1 minus the probability of non-coverage (the product of the failure probabilities of all sets in A containing y).

Immediate Extensions

- *All arguments work for node-weighted versions of the problems (where the value associated with protecting nodes in the graph, or covering the elements in the ground set, can vary and is specified in the input).*

  This follows from the proof of Theorem 3: $E(\cdot)$ can be any linear function of the probabilities that nodes are protected.

- *The multi-infection-source case of spreading-vaccination Stochastic Graph Vaccination is also maximum coverage subject to probabilistic set failure (however, the multi-infection-source case of non-spreading Stochastic Graph Vaccination in trees is not submodular). Thus, a result analogous to Theorem 2 also holds in this case.*

  This result can be obtained by redefining S_{jxt} in the reduction for the single source case. Now, if R is the set of sources, let S_{jxt} contain all nodes y for which $d(x,y) \leq \min_{r \in R} d(y,r) - t$. Pushing slightly further, *the sources may even arise at different times*: if, for each $r \in R$, r first becomes infectious at time t_r, then define S_{jxt} to be the set of all nodes y for which $d(x,y) \leq \min_{r \in R} \{d(y,r) - (t - t_r)\}$. Either described extension can also be quickly obtained by reduction to the single-source case of spreading-vaccination Stochastic Graph Vaccination: add a super source that acts as a single infection point which is connected by paths (of new degree-two nodes) of $t_r + 1$ edges to source r (for each $r \in R$). Let the failure probability of every vaccine at the all new nodes be 1 in every time step (and the value of all new nodes is 0) and shift all $f_j(x,t)$ to the right one time step (letting $f_j(x,0) = 1$ for all (x,j)).

References

Ageev, A. A., and Sviridenko, M. 2004. Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *J. Comb. Optim.* 8(3):307–328.

Anshelevich, E.; Chakrabarty, D.; Hate, A.; and Swamy, C. 2009. Approximation algorithms for the firefighter problem: Cuts over time and submodularity. In *ISAAC*, 974–983.

Cai, L.; Verbin, E.; and Yang, L. 2008. Firefighting on trees: $(1 - 1/e)$-approximation, fixed parameter tractability and a subexponential algorithm. In *ISAAC*, 258–269.

Calinescu, G.; Chekuri, C.; Pál, M.; and Vondrák, J. 2007. Maximizing a submodular set function subject to a matroid constraint (extended abstract). In *IPCO*, 182–196.

Colorado.Dept.Agriculture. 2009. Colorado biocontrol. *Published online at www.colorado.gov* Spring Issue.

Cumming, G. 2002. Habitat shape, species invasions, and reserve design: Insights from simple models. *Conservation Ecology* 6 (1).

Fisher, M. L.; Nemhauser, G. L.; and Wolsey, L. A. 1978. An analysis of approximations for maximizing submodular set functions. *Mathematical Programming Study* 8:73–87.

Hartnell, B. 1995. Firefighter! an application of domination. *Presentation, 25th Manitoba Conference of Combinatorial Mathematics and Computing*.

Kempe, D.; Kleinberg, J. M.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 137–146.

Kovacs, K. F.; Haight, R. G.; McCullough, D. G.; Mercader, R. J.; Siegert, N. W.; and Liebhold, A. M. 2010. Cost of potential emerald ash borer damage in u.s communities, 2009-2019. *Ecological Economics* 69:569–578.

Kovacs, K.; Haight, R. G.; and Mercader, R. 2012. A bioeconomic analysis of emerald ash borer control in an urban forest. *Working Paper*.

Norris, R.; Memmot, J.; and Lovell, D. 2002. The effect of rainfall on the survivorship and establishment of a biological control agent. *Journal of Applied Ecology* 39:226–234.

Shea, K., and Possingham, H. 2000. Optimal release strategies for biological control agents: an application of stochastic dynamic programming to population management. *Journal of Applied Ecology* 37:77–86.

Shmoys, D., and Spencer, G. 2011. Approximation algorithms for fragmenting a graph against a stochastically-located threat. In *Proceedings of the Workshop on Approximation and Online Algorithms (WAOA)*.