

We apologize for the delay in returning this revision.

Below we have included the entire decision letter including the reviews in black text. Our responses to the critical comments are in red text.

We think that this has been a useful exercise and that the paper has been improved.

Thanks for the effort.

Dear Drs. Xue, Fern, and Sheldon,

We have received all the reviews for the paper you submitted to JAIR, entitled "Scheduling Conservation Designs for Maximum Flexibility via Network Cascade Optimization". As you can see below, the reviewers had a maximal variety of recommendations: one accept, one accept with minor revisions, and one reject. I read the paper in detail myself and have provided comments below based on both the reviewers' comments and my reading of the paper. While the paper is an improvement on the earlier version, there are several difficulties which preclude publication at this stage. Since JAIR's policy is to accept papers that only require minor revisions, we cannot currently accept the paper. However, we would be happy to see a revised paper on this subject--addressing the reviewers' concerns--published in JAIR. Thus, I strongly encourage you to resubmit the paper.

Note that this decision is an exception from JAIR's usual policy of allowing only one resubmission. In my view the paper is interesting and relevant both in terms of the problem being solved and the approach to solving it. Therefore, I would like to see it eventually appear in JAIR. However, I cannot justify the paper's acceptance in its current state. Please see my detailed comments below plus the comments of the reviewers. I am confident that these modifications can be made but I also believe that because they are at a fundamental level for the paper that some time should be spent on them.

If you decide to re-submit it, I plan to read the paper carefully myself and seek the advice of a subset of the current reviewers. Depending on that initial evaluation, the paper may not require another full round of reviewing.

As an administrative comment: please note that one of the reviewers for the original submission of your paper was not available for this draft and so a new (fourth) reviewer was recruited. However, in turning down the request to review, the original Reviewer #1 reiterated the opinion that the paper was not relevant to AI. I do not

share this opinion.

Attached you will find the reviewers' comments and suggestions.

If you choose to revise the paper, please submit it through our standard web process, and include a cover letter describing the changes and responding to the reviewers' criticisms.

Thank you for submitting your paper to JAIR.

Regards,

Chris Beck
Associate Editor

Comments by Associate Editor

Major Comments

- Complexity of the original deterministic optimization problem: On p3, the authors state "In particular, our deterministic optimization problem is one of network cascade optimization, which we show is equivalent to a novel variant of the directed Steiner tree problem." As pointed out by Reviewer #4, this statement is not actually supported in the paper as a general variant of the Steiner tree problem is proposed. While it is clear that the deterministic optimization problem can be posed as the specified Steiner tree variant, it is not clear that these problems are equivalent as the former contains specific structure that the latter lacks (see Review #4). So while the complexity proof is fine, it is about the Steiner tree variant, not about the original deterministic problem. I think that it is crucial to make this distinction because the reader is left with the impression that the paper contains a complexity proof about the original deterministic optimization problem and this is not the case. It is even possible (though I doubt it) that the original deterministic problem is polynomial. Ideally, the paper should include a proof of the equivalence of the original problem and the Steiner tree variant.

The detailed changes of paper are in the response to under reviewer #4's comments. In summary, the problems are both NP-complete (hence equivalent), which is shown now in the paper. We also highlight that the primal-dual algorithm obtained for the Steiner problem is

equivalent to the algorithm obtained by following a standard primal-dual schema for the original problem. The derivation from the Steiner problem, however, is much simpler and hence is our preferred presentation vehicle.

- Where does the flexibility come from? As pointed out by Reviewer #2 in the review of the first submission of this paper to JAIR and Reviewer #4 here, it is unsatisfying that the flexibility of the schedules (for $\epsilon = 0$) is somewhat accidental in that it arises from the limited number of samples. As Reviewer #4 points out, N has a double role as it determines the error in approximating the true rewards (higher $N \rightarrow$ lower error), and this is true of the up-front schedule too, but it also provides the only scope for flexibility (higher $N \rightarrow$ less flexibility, as eventually the only optimal schedule is the up-front). This is a rather complex and confusing situation since these two roles are essentially distinct. While SAA provides some convergence guarantees to the former, the latter has no treatment whatsoever. A formal treatment is not necessary for this paper (though it would be welcome and valuable) however a clear presentation of where the flexibility comes from and this role of N is needed.

We agree that this was not described well and it was an unnecessary distraction. Please see our response to reviewer #4 for details. We have clarified the high-level algorithm description (Section 3.4) to make it clear that “early stopping” is the primary way that our algorithm allows for flexibility via setting $\epsilon > 0$. This provides flexibility even for infinite N .

In practice, even for relatively large finite values of N we will get non-trivial flexibility for $\epsilon = 0$ (no early stopping). But this does not mean that N is our intended knob for controlling flexibility. $\epsilon > 0$ is the main knob.

- The continued use of the $<_c$ partial ordering. There are a number of points here. First, the practice of discounting future costs is widely accepted in a variety of areas including AI. It seems more realistic to have a decreasing cost function as part of the standard definition of your problem due to economic realities such as inflation. Especially with the time-frames being considered, it is quite strange to equally weigh the cost of a purchase now with the cost of a purchase in 80 years (even if they have the “same” dollar value) which is what the original cost curve formulations do, in the name of flexibility. In other words, it seems more realistic that the surrogate cost function should be the cost function in the problem definition. This is particularly appropriate since the actual problem solved is based on such a cost function. Reviewer #4 points out that the discussion of $<_c$ and the proof of the valid use

of a surrogate function could be avoided and the paper therefore simplified. I see this as a second benefit: removing the <c discussion would make a more realistic problem *and* a simpler paper. Third, as pointed out by Reviewer #2, in the original submission, the proof of proposition 1 is straightforward and could be omitted. You have decided not to do this for completeness, however, at most, this proof should appear in an Appendix as it currently breaks up the flow. To summarize: I believe the entire presentation of partial order should be removed and the original problem should have a descending cost function. However, there may be some subtleties I am missing that the authors can convincingly argue for its inclusion. If this is the case, the proof of Proposition 1 should be removed or at least moved to an Appendix.

We have thought carefully about this and agree with you. We have modified the text to describe the motivation for this in terms of both flexibility and economic factors, which are described on page 7. Then we directly define the original problem in terms of the time-weighted cost function, where the weights are non-increasing with time (formula (1)).

Minor Comments

p13: "optimal Steiner graph needs not be a tree" -> need
Corrected.

p23: Are cost curves discounted costs? Please specify what is being plotted. Also why isn't the upfront cost shown in Fig 6?

The cost curves are the accumulated costs over time as defined in paragraph 3 in section 3.2. They are not discounted. This is why in the graph the cost curve with $H = 100$ goes to 3.2×10^8 in the end, meaning the entire budget is used. We have made this clear by re-visiting the definition of cost curve when introducing Figure 6 in section 6.1.

It is a good idea to add the cost curve of the upfront schedule. We now have it in the graph, which is a flat line with cost = 3.2×10^8 since all the purchases are made at time 0. In the same paragraph about Figure 6, we now point the upfront schedule to this upfront cost curve in the figure to better explain the flexibility of our schedules over the upfront one.

p23: The discussion at the end of Sect 6.1 (last paragraph on p23) is interesting from the perspective of asking if the problem being modeled is the right one. It appears to me that solutions that delay buying parcels until the last (reliable) minute are high risk from an ecological perspective and not necessarily high quality for the population. Given the uncertainty of natural occurrences (e.g., forest fires) over a time-span of decades, does it really make sense that your reward function should be the spread of population at some fixed time (H) in the future? Another perspective is to ensure that the

population is as widespread as it can be across the planning horizon so that it is best able to respond to unexpected negative events. This suggests that you really want a reward function based on a population curve over time (analogous to the cost curves) or to optimize the "net present spread" with some monotonically **increasing** weight (i.e., spread in the future is more valuable than spread now). Your current reward function can be seen as an extreme of this as the weight at H is 1 and the weight at all other times is 0. I am certainly not suggesting that this is investigated in depth in the current paper. However, a discussion of the appropriateness of the reward with respect to true ecological goals in the context of possible future work would be valuable, in my opinion.

These are all interesting models. In this paper we focused on the previously studied model for this domain (Sheldon et al (2010) and Ahmadizadeh et al (2010)). Now in the end of the last paragraph (section 6.1), we point out that basically the observation of the population "hole" is due to our definition of the reward function. We also added these possible models to future work (section 7, the last paragraph).

Review #2

TRACKING NO.: 4322

ARTICLE: Scheduling Conservation Designs for Maximum
Flexibility via Network Cascade Optimization

AUTHORS: Shan Xue, Alan Fern and Daniel Sheldon

RECOMMENDATION: Accept with minor revisions

General comments:

All major issues in my original review have been well addressed, and I think the paper has turned out very nicely. Congratulations!

Of the less critical issues in my original review, one remains:

> 6. [...]
>
> Or maybe the discrepancy exists because the Steiner tree
> algorithm doesn't buy certain parcels **at all**? The way the
> algorithm is written up, there seems to be nothing in it that
> guarantees that every parcel is eventually bought: [...]
> [...] How is this handled,
> considering that the problem as defined requires that every

> parcel is eventually bought (at time H at the latest)?

The authors' comments clarify that indeed the algorithm doesn't buy every parcel and that there is no need for it to do so.

I would point out there that this is a mismatch with the problem definition on page 6, which explicitly requires the purchase schedule to map every parcel to a purchase time in $\{0, 1, \dots, H\}$. I agree with the authors that it is useful *not* to require purchasing every parcel. Therefore, the problem definition should be amended (e.g. by adding ∞ to the set of allowed values in a purchase schedule and amending the associated definitions).

The main reason why I think such a mismatch is problematic is that other authors might want to tackle the problem defined in the paper, and if they try to find a faithful solution that follows the definition in Section 3, they won't be able to find solutions with a budget as low as those reported in Section 6, since those solutions are impossible when following the problem definition (see the calculations in my original review). Hence, please amend the problem definition so that it matches what is computed.

Thanks for pointing out this detail. We have corrected this by including ∞ as a purchase time in the problem definition (section 3.1, paragraph 2). Parcels that are not going to be purchased are assigned purchase time ∞ . Similarly, we also define the discounted function f (section 3.2, paragraph 2) as a function from times in $\{0, \dots, H, \infty\}$ to real numbers. We further require $f(\infty) = 0$ so that the costs of unpurchased parcels do not contribute to the surrogate cost.

Apart from this, I only have the following very minor suggestions for corrections:

- p. 3: "(see Shapiro (2003))"

Please check the JAIR author instructions regarding parenthetical references within parentheses (<https://www.jair.org/authorinstrs.html>, point 1.)

Corrected to (Shapiro, 2003)

- p. 4: "(e.g. see the review ... and Levin (2005))"

ditto

Corrected to (see ... article, Williams... and Levin, 2005)

- p. 5: "successfully applied in (Bent & Hentenryck, 2004), (Chang ...)"

Please check the JAIR author instructions regarding citations used as nouns.

Corrected to applied by Bent et al. (2004), ...

- p. 7: "reward loss, ---"

=> should have no comma before dash

Corrected.

- p. 7: "using larger epsilon, increases"

=> no comma

Corrected.

- p. 8: "(see Shapiro (2003) for survey of some results)"

See above; also, "for survey" => "for a survey".

Corrected.

- p. 9: Figure 1

Node $v^1_c,3$ should not be red. Edge $v^3_c,0 \rightarrow v^3_b,1$ should be bold if I understand correctly that the bold edges are the ones which go into some bought node. (If, on the other hands, they are the ones *within* the bought nodes, two of the bold edges should be non-bold.)

Yes, you are right. We've made it right.

- p. 9: "no later than time t . That is"

=> should be a comma, not a period, as the "That is" part is not a complete sentence.

Corrected.

- p. 9: "edges, that involve"

=> no comma

Corrected.

- p. 13: "for even acyclic"

=> "even for acyclic"

Corrected.

- p. 14, line 2: "there is a collection \mathcal{S} \in \mathcal{S}"

"\in" should be "\subseteqq"

Corrected.

- p. 14, line 2: "\cup_{S_j \in \mathcal{S}}"

"\cup" should be "\bigcup"

Corrected.

- p. 14, "\$i'th\$" (multiple times)

should be "\$i\$th" or "\$i\$-th"

Corrected.

- p. 18, "so that (8) remain satisfied"

=> "... *remains* ..."

Corrected.

- p. 19, "computation, then the memory"

=> comma should be "." (two full sentences)

Corrected.

- p. 19, "case, corresponds to only"

=> no comma; also, I'd drop "only" (there's another "only" later in the sentence)

Corrected.

- p. 20, "that criteria"

=> "that criterion" ("criteria" is plural)

Corrected.

- p. 25, Figure 8

The bottom of the x axis label is cut off. (See letters "g", "y", "p".)

- p. 26, "Trading-Off Flexibility" (caption of Section 6.3)

no hyphen: "Trading Off" (because it is used as a verb)

Corrected.

- p. 26, "schedules that trade-off"

ditto

Corrected.

- p. 28, "Among the complicating factor include"

There is something wrong with the grammar here.

Corrected.

- p. 28: "an existing well studied problems"

=> "an existing *well-studied* *problem*"

Corrected.

- p. 29, Bent et al. reference

"Hentenryck, P. V." => "Van Hentenryck, P."

Corrected.

- p. 29, Bent et al. reference

"online" => "Online"

Corrected.

Review #3

I have reviewed the paper (which is really a revised version of an earlier draft submitted to JAIR for which I was one of the reviewers) and have concluded the paper should be accepted as they have met or sufficiently addressed all revision requests based on the first review.

Review #4

Title: Scheduling Conservation Designs for Maximum Flexibility via Network Cascade Optimization

Author(s): Shan Xue, Alan Fern and Daniel Sheldon

Overall evaluation

- Accept as is
- Accept with minor revisions
- Reject

Comments and suggestions

Foreword: I was not a reviewer of the original version of the article and I wrote the comments and suggestions below before reading the initial reviews and the author's answer. My recommendation after reading the other comments is at the end of the review.

The article deals with a very interesting and original application for scheduling flexible conservation designs. The problem is seen as a stochastic optimization

problem where the problem is to maximize schedule flexibility (formulated as a cost minimization) while ensuring a minimum reward (amount of population spread at the schedule horizon).

The problem is solved in the following way:

- 1- A deterministic model is built (network cascade) using a scenario-based approach
- 2- This deterministic model is generalized as a new problem on graphs: set-weighted directed steiner graph (SW-DSG)
- 3- A MIP formulation of the SW-DSG problem is provided, which is shown not to scale well enough for the application
- 4- A Primal-Dual approach is proposed for the SW-DSG problem that boils down to a greedy algorithm that provides a feasible solution and a lower-bound on the cost

The article is well structured, well illustrated with figures and easy to read, except maybe the beginning of section 5.1 where some more details could be given about the dual formulation of the problem.

The adequacy of the article with AI is not evident. It for sure deals with an optimization problem but the method itself consists of a greedy algorithm (no search, no real inference mechanism) that is more based on classical OR concepts (Math Programming, Graph theory).

One thing that is not clear in the paper is whether the original problem formulated in section 3.3 is "equivalent" to the SW-DSG problem. The article proves that the original problem can be expressed as a SW-DSG problem but the original problem could be simpler. It is shown that SW-DSG problem is NP-Hard even for acyclic graphs, but in the original problem there is an additional property that is related with some particular structure of the subsets $\{ E_i \}$. For instance, there is a partition of the $\{ E_i \}$ in P_1, \dots, P_p (These subsets P corresponds to the parcels) such that (1) for $i \neq j$, any subset E_{ik} of P_i is disjoint from any subset E_{jl} of P_j (each patch is associated a unique parcel), (2) for each i , subsets E_{ik} of P_i are included in one another like russian dolls (this is the fact a parcel purchased before $t-1$ is also considered as purchased at time t) and (3) the smallest subsets E_{ik} of the P_i s form a partition of the terminal nodes (the terminal nodes are the ones at the horizon of the schedule, which corresponds to the smallest subsets E_{ik}). For me, it is still not clear whether the original problem is NP-Hard or not. For instance the reduction of weighted set cover to the SW-DSG problem results in a subset family that does not seem to satisfy the above properties. This being said, I admit that the proposed SW-DSG problem is very interesting in itself.

It is true that the original deterministic optimization problem has a specific structure in the edge sets which SW-DSG lacks. The original problem is still NP hard. The theorem and the proof in section 4.2 now are for the Shortest Path case of SW-DSG restricted to have the special edge structure in scenario graphs. The proof is quite similar to the proof we had before.

A primary reason that we choose to focus on the SW-DSG problem for the algorithmic part of the paper is the simpler exposition. In particular, we originally derived a primal-dual algorithm for the original problem following the standard schema. The resulting algorithm turns out to be identical to the one obtained by deriving a primal-dual algorithm for SW-DSG and applying it to graphs from the original problem. However, the exposition of the primal-dual algorithm for SW-DSG is significantly simpler. We now indicate this in the paper in paragraph 2 of Section 5.

It is strange that in the description of the problem, there seem to be some focus on defining a partial order between the schedules (relation \prec) and the fact there is not a unique solution. This may suggest that one could be interested in computing several non-dominated solutions. But in the sequel of the paper, relation \prec is aggregated as a unique surrogate objective function (section 4.1) and the relation between this function and the partial order is not reconsidered anymore in the paper. I'm wondering if it would not simplify the paper just to avoid mentioning this partial order and starting right from the surrogate function which can be considered as making sense as it could be related with some net present value.

We appreciate your suggestion. We have thought carefully about it and in the end agree. We have modified the text to describe the motivation for this in terms of both flexibility and economic factors (suggested by Associate Editor), which are described on page 7. Then we directly define the original problem in terms of the time-weighted cost function, describing that generally the weights will be strictly decreasing.

When $\epsilon=0$, the number of cascades N plays a double role: N should be big enough to have a good model for uncertainties (so a good estimation of the reward), but if N is too big, this will necessarily decrease the flexibility of the schedule (when N tends to infinity, the schedule tends to the upfront non-flexible schedule). So the selected value for N should I think be more motivated in the experimental section. Currently, the selected value $N=10$ is motivated only in light of the expected reward: "We also see that 10 cascades is quite close to get the best performance and the rate of improvement is slowing down. Thus, the remainder of our experiments use 10 cascades for the SAA."

We agree that this was not described well and it was an unnecessary distraction. We have clarified the high-level algorithm description (Section 3.4) to make it clear that "early stopping" is the primary way that our algorithm allows for flexibility via setting $\epsilon>0$. This provides flexibility even for infinite N .

We select N based on expected reward and it happens that for those values we happen to get significant flexibility even for $\epsilon=0$ (no early stopping). But this does not mean that N is our intended knob for controlling flexibility. $\epsilon>0$ is the main knob.

It is difficult to assess the quality of the proposed approach in terms of how close it is to the optimal solution. Well, this is unfortunately true as soon as one works

on large complex problems like here. Figure 8 suggests that the approach is about 15% above the optimal solution found by the MIP on small problems. This gap is difficult to interpret, especially as the cost objective is an aggregated function with exponential terms. It would be very informative to compare the cost curves with the optimal ones for these small problems. On a similar line, I think you should justify why you consider that a lower bound that is within a factor 2 is quite informative because in general, such a gap would be considered as quite large.

We agree the gap is difficult to interpret. However, comparing cost curves for such a small value of $N=2$ is not very meaningful. In this section we were primarily interested in assessing whether our primal dual algorithm seemed to be in the ballpark with respect to approximating the MIP solution and hence we only compared in terms of the MIP objective. A factor of 2 bound is often considered quite good in many circles, but this is ultimately in the eye of the beholder. We no longer say "quite informative" and just state we are within a factor of 2.

Some more detailed remarks and typos:

In section 3.3, when explaining the graph representation ("More concretely ..."), it would be good to explicitly mention how extinction is modeled even if this is trivial (no arc between t and $t+1$).

Page 20, "The the SAA"

Corrected.

Page 21. The fact the solution found by PD is better than the optimal solution found by CPLEX is strange. The difference is here around 0.5% which is ways larger than the default MIP gap tolerance of CPLEX (0.01%). That could be due to the fact the model is ill-conditioned with some very small and/or very large coefficients. For instance, I see that the objective function is in M\$, if the unit is 1\$ maybe you could try rescaling the objective coefficients.

Page 22, in subsection "Number of Cascades in the Scenario Graph". When you say you are running 20 simulations, these are simulations that are used to evaluate the approach. This is not related with the number N of scenarios used in the cascade mode, right? There is some possible ambiguity here.

Thanks for letting us know of this ambiguity. Yes, N is the number of scenarios in the scenario graph that is used to compute the schedules, which is 10 in our experiments. To evaluate the any schedule, we run a number of simulations (20 here) of the actual population spread process by following the purchase schedule and use the average reward to validate the quality of the schedule. We have clarified this further in the revision (page 22, in subsection "Number of Cascades in the Scenario Graph").

In the end of section 6.3, it would be good to add a figure for the cost curves of the

early-stopping schedules.

They are in Figure 11. The figure was shown ahead of the text. We have made it closer to the text so that the readers would not miss it.

Additional comments and suggestions after reading initial reviews and author's feedback:

Reading question 2 of reviewer 2, I understand that in the original paper there was even less focus on the version of the problem with a trade-off between reward an cost and that the part

dealing with epsilon parameter and "early stopping" of the greedy algorithm was added or extended

with the present version. I still find that the current version does not fully address this issue. The "flexibility" for the case $\text{epsilon}=0$ is only accidental and due to scenario sampling (see my above comment on the double meaning of "N") so it is not really interesting. The real interesting

case is the exploration of the trade-off between reward an cost, for instance with $\text{epsilon}>0$. But

this case is still treated as a secondary problem in the current version of the article, both in the justification of the selected approach and in the computational experiments.

The problem studied in the article and the different elements for its resolution are clearly very interesting but I think the approach should focus on considering the trade-off between reward an cost

as the *central* problem. The case $\text{epsilon}=0$ is not interesting per se. This would also mean to

revise the link with the SW-DSG problem (as the set of terminal nodes is not fixed). It would also be

necessary to see if the particularities of the problem (see one of my comment above) cannot be

exploited to simplify the modified SW-DSG problem. So I would recommend a rejection. There is here

enough material and experience on the problem for a very good paper but I think it will require significant reworking.

We like the perspective of viewing the main goal as considering various trade-offs between reward and cost. To emphasize this perspective, we rewrote parts of the abstract and introduction (page 3, 4). Then \epsilon provides us with different points on the trade-off but we cannot exactly quantify it precisely as it depends on the specific problem structure. Also see comments above regarding $\text{epsilon}=0$.