



AFRL-RI-RS-TR-2018-275

ANTAGONISTIC GRAPH COLORING UNDER UNCERTAINTY

UNIVERSITY OF PITTSBURGH

OCTOBER 2018

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RI-RS-TR-2018-275 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /

ANDREW GORCZYCA
Work Unit Manager

/ S /

JAMES S. PERRETTA
Acting Deputy Chief, Information
Exploitation & Operations Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) OCTOBER 2018		2. REPORT TYPE FINAL TECHNICAL REPORT		3. DATES COVERED (From - To) FEB 2017 – MAY 2018	
4. TITLE AND SUBTITLE ANTAGONISTIC GRAPH COLORING UNDER UNCERTAINTY				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER FA8750-17-1-0090	
				5c. PROGRAM ELEMENT NUMBER 61101E	
6. AUTHOR(S) Jeffrey P. Kharoufeh				5d. PROJECT NUMBER AGCU	
				5e. TASK NUMBER UP	
				5f. WORK UNIT NUMBER IT	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Pittsburgh 1025 Benedum Hall 3700 O'Hara Street Pittsburgh, PA 15261				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/RIGB 525 Brooks Road Rome NY 13441-4505				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RI	
				11. SPONSOR/MONITOR'S REPORT NUMBER AFRL-RI-RS-TR-2018-275	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The primary objective of this research is to determine optimal, or near-optimal, ways of identifying a subset of nodes of interest in a graph and then "coloring" those nodes, subject to a time constraint. The term coloring is general and can be interpreted to mean that we alter the binary status of a node. For example, biological systems may be moved between healthy and infected states, physical systems between working and failed, social systems between organic and influenced, and digital systems between functioning and corrupt. Consequently, this work is applicable to many problems in epidemiology, warfare, influencing social networks and cybersecurity.					
15. SUBJECT TERMS Graph Coloring, Stochastic Optimization, Applied Probability, Networking, Markov Decision Processes					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 21	19a. NAME OF RESPONSIBLE PERSON ANDREW GORCZYCA
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) N/A

TABLE OF CONTENTS

List of Figures	ii
1.0 SUMMARY	1
2.0 INTRODUCTION	2
3.0 METHODS, ASSUMPTIONS AND PROCEDURES	2
4.0 RESULTS AND DISCUSSION	3
4.1 A Baseline Policy for Connected Graphs	3
4.2 Fully-Connected Graphs Without Uncoloring	6
4.3 Fully-Connected Graphs With Uncoloring	8
4.4 Numerical Illustrations and Results	11
5.0 CONCLUSIONS	15
LIST OF ACRONYMS & GLOSSARY	16

List of Figures

Figure 1	A randomly-generated graph with $N = 25$ and $M = 5$	4
Figure 2	Effect of varying the coloring success probability p_c	5
Figure 3	Effect of varying the coloring budget B	5
Figure 4	Comparison of failure modes.	6
Figure 5	Greedy & fully-connected policies when greedy policy is budget- and time-constrained.	8
Figure 6	Behavior of the success probability as a function of B and p_u	11
Figure 7	Best-fit curves for best attainable success probability..	12
Figure 8	Value function comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). For both cases, the time is fixed at $t = 5$ and no nodes in S have been colored.	12
Figure 9	Comparing optimal actions with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). Brighter color indicates attempting to color a larger number of nodes. In both cases, the time is fixed at $t = 5$ and no nodes in S have been colored.	13
Figure 10	Value function comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). In both cases, the time is fixed at $t = 16$ and no nodes in S have been colored.	14
Figure 11	Optimal action comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). Brighter color indicates attempting to color a larger number of nodes. In both cases, the time is fixed at $t = 16$ and no nodes in S have been colored.	14

1.0 SUMMARY

This document represents the final technical report for the DARPA project, “Antagonistic Graph Coloring Under Uncertainty.” The primary objective of this research is to determine optimal, or near-optimal, ways of identifying a subset of nodes of interest in a graph and then “coloring” those nodes, subject to a time constraint. The term coloring is general and can be interpreted to mean that we alter the binary status of a node. For example, biological systems may be moved between healthy and infected states, physical systems between working and failed, social systems between organic and influenced, and digital systems between functioning and corrupt. Consequently, this work is applicable to many problems in epidemiology, warfare, influencing social networks and cybersecurity.

The first set of results presented herein are based on randomly-generated graphs and reveal explicit performance limits for graphs that are *not fully-connected* (i.e., those in which there does not exist an edge connecting each node to all other nodes in the network). By randomly-generating thousands of such graphs, we first examine the performance of a simple baseline algorithm to reveal insights into the problem and explore the effects of coloring probability and coloring budget. It was found that there exist problem regimes in which small increases in coloring budget can result in exponentially higher success probabilities; however, there also exist regimes in which large budget increases provide little to no substantive improvements. This phenomenon is shown to be the result of the existence of two distinct failure modes that behave predictably as a function of the coloring budget and coloring probability. These results motivated an exploration of the case when the underlying graph is fully-connected. For that class of problems, we were able to determine optimal dynamic coloring strategies using discrete stochastic optimal control techniques (namely Markov decision process models). This framework has dual benefits of exactness for fully-connected graphs, and the ability to provide theoretical performance bounds for any possible coloring policy over any possible graph. Combining these results, we demonstrate that the baseline algorithm achieves near-optimal performance in certain problem instances, and poor performance in others. We further extend the fully-connected analysis to include problems in which the nodes can spontaneously uncolor with a certain probability. Using the MDP framework, we were able to characterize optimal policies that, again, demonstrate performance bounds for more general graphs.

2.0 INTRODUCTION

To provide some context for the problem, we first present a mathematical model description and then review the objectives of this research. Let $G(V, E)$ be an undirected graph, where V denotes the set of vertices (or nodes) and E is the set of edges. There are $N \equiv |V|$ nodes in the graph, and there exists a subset of nodes of interest, $S \subset V$, with cardinality M ($M \leq N$). Both N and M are assumed to be finite. The fundamental problem can be described as follows. We seek to “color” each of the nodes in S within a particular time interval $T = [\ell, u]$, subject to a coloring “budget.” That is, we are only allotted a fixed, finite number of coloring attempts, after which no coloring can be attempted. We consider discrete time periods in the time horizon $\mathcal{T} = \{1, 2, \dots, \ell, \ell + 1, \dots, u\}$. If all of the nodes in S are colored within $[\ell, u]$, then the event is said to be a *success*; otherwise, it is a *failure*. ***Our aim is to maximize the probability of success (or the proportion of experiments that lead to success).***

Initially, the decision maker knows only the number of nodes N , the number of special nodes M and the probability of successfully coloring a node, which is denoted by p_c ($0 < p_c \leq 1$). In the most generic form of the model, nodes may also independently uncolor with probability p_u ($0 \leq p_u \leq 1$); we present results for this more complicated version of the problem. First, one node is selected randomly from V and is colored, thereby revealing its adjacent nodes and whether or not this node belongs to S . During each subsequent time period, the decision maker must decide which nodes, from among those that are visible, to attempt to color. If the coloring is successful, the decision maker learns whether the node is in S , and the nodes adjacent to the newly colored node become available for coloring in future time periods. If the coloring is unsuccessful, the node remains uncolored, and no new knowledge of the topology is gained by the decision maker. Denote by B the total coloring budget over the time horizon. That is, the decision maker depletes one unit from the budget for any coloring attempt made (irrespective of the outcome). Once the budget is depleted, no additional coloring attempts can be made, thereby inducing a tradeoff between aggressively coloring and depleting the budget. Intuitively, with an unlimited budget, one would attempt to color at will until all the nodes in S are identified, ensuring success. However, with a constrained number of coloring attempts, and a finite time limit, one must carefully select how these attempts will be made in each period. Initially, we consider models in which the nodes cannot uncolor (i.e., $p_u = 0$), but later extend the preliminary models to consider this complication.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

The primary methods employed in this research include graph theory, stochastic modeling, Markov decision process models (discrete stochastic optimal control) and optimization theory. When analyzing graphs that are *not* full-connected, exact solutions cannot be obtained via analytical methods. For these cases, we created, executed and analyzed several computer simulation models (in the MATLAB computing environment) to gain insights into the behavior of coloring strategies. Specifically, the simulation model was used to: (i) randomly generate graphs having random topologies; (ii) simulate coloring and uncoloring probabilistic behavior; (iii) simulate a greedy, baseline coloring heuristic for non-fully-connected graphs in order to estimate the success probability.

When the graphs are assumed to be fully-connected, we are able to provide exact problem

formulations by creating Markov decision process (MDP) models. This modeling approach assumes a discrete time horizon and requires a state-space model, along with one-step transition probabilities. We formulated two distinct, yet related, problems: (i) the case when nodes cannot uncolor after having been colored; and (ii) the case when nodes uncolor independently with probability p_u after having been colored. In the latter scenario, the state space is (necessarily) larger and, therefore, the computation time increases substantially.

In summary, we examined three different models and provide interesting insights into the behavior of the success probability for each of the models. The three models are summarized as follows:

1. A simulation model to *estimate* the success probability for non-fully-connected graphs. The model implements a greedy coloring heuristic and assumes that nodes cannot independently uncolor;
2. A Markov decision process (MDP) model to compute the *exact* success probability for fully-connected graphs. The model implements the value iteration algorithm and assumes that nodes cannot independently uncolor;
3. A second MDP model to compute the *exact* success probability for fully-connected graphs. The model also implements the value iteration algorithm but allows the nodes to independently uncolor after they have been colored.

4.0 RESULTS AND DISCUSSION

In this section, we summarize the main results obtained in this research and provide numerical illustrations. Some further closing remarks are provided in Section 5.0.

4.1 A Baseline Policy for Connected Graphs

Consider first a simple, greedy heuristic policy that serves as a baseline upon which we will build more sophisticated policies. Roughly speaking, this policy seeks to color all but one of the nodes in set S before time ℓ ; subsequently, it attempts to color the remaining nodes in S during $[\ell, u]$. More precisely, if at some time $t < \ell$ the remaining budget is b_t , the number of available nodes to color is a_t and the number of nodes in S that remain to be colored is r_t , this baseline policy prescribes attempting to color a total of

$$k_t = \min\{r_t - 1, a_t, b_t\}$$

randomly selected nodes (from the set of available nodes). For any time $t \geq \ell$, the policy calls for coloring $k_t = \min\{a_t, b_t\}$ randomly selected nodes. This policy is intuitive in that it will not allow for the complete coloring of S before the time window $[\ell, u]$; however, it is greedy in that within the time interval $[\ell, u]$ it will attempt to color as many nodes as possible.

To examine the performance of this baseline policy, we performed a series of computer simulation experiments. Specifically, we randomly generated 5000 graphs inside the unit square. For each graph, the locations of the vertices V are uniformly distributed, and the edges in E are generated by constructing the Delaunay triangulation of V . We used this method because it induces favorable properties, namely that the graphs are connected (though not fully-connected), nearest neighbors in V are adjacent in G and distant nodes

are rarely adjacent (sliver triangles are avoided). Figure 1 depicts one of the randomly-generated graphs with 25 nodes and five nodes of interest (which are identified with bold font). Note that, with the exception of the nodes on the outer edges, nodes tend to be

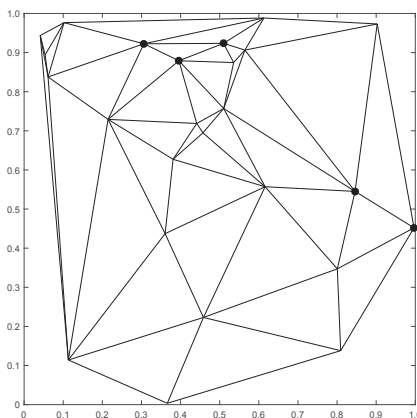


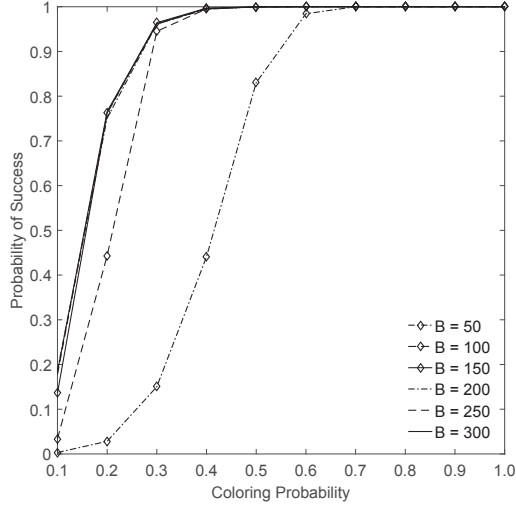
Figure 1: A randomly-generated graph with $N = 25$ and $M = 5$.

adjacent to their nearest neighbors. Lastly, the set of special nodes S was selected randomly as a subset of nodes within V . In order to promote spatially-clustered nodes of interest, the nearest neighbors of S are added to S with some probability p ($0 \leq p \leq 1$). After a graph is generated, the coloring of G is simulated by employing the baseline policy described above to determine which actions are taken.

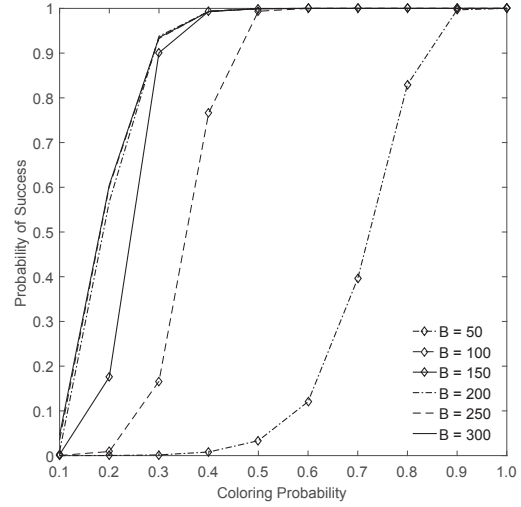
The output of the simulation run is saved as the entire history of visible portions of G at each time step, coloring attempts, coloring successes and the final result, which must be one of three mutually exclusive outcomes:

1. Successful coloring of all nodes in S ;
2. Failure due to expiration of the time window; or
3. Failure due to depletion of the coloring budget.

We examined the performance of the baseline policy over a large set of parameter values. Specifically, we considered $T = [12, 24]$, $N \in \{25, 30, \dots, 100\}$, $M \in \{5, 7, \dots, \lfloor N/4 \rfloor\}$, $B \in \{30, 40, \dots, 300\}$, and $p_c \in \{0.1, 0.2, \dots, 1.0\}$. For each set of parameters, 5000 graphs were randomly generated, the baseline policy is implemented, and the results recorded. In Figure 2(a), we fix $N = 25$ and $M = 5$ and show the impact of varying p_c for different values of B ; similarly, we fix $N = 40$ and $M = 9$ in Figure 2(b). We observe (intuitively) that increasing the coloring probability (p_c) improves outcomes, and that for a sufficiently high coloring budget, only a modest coloring probability is required to achieve a high probability of success. Additionally, we observe that there are diminishing marginal returns as the coloring budget increases. To further investigate this latter point, we varied the budget for different, fixed coloring probabilities and plotted the results in Figure 3. In each curve, we observe that there exists a point at which increasing the coloring budget has little-to-no impact on the outcome of the experiment. The reason for this is intuitive: with a sufficiently large budget, a failure event is more likely to be the result of time expiration.

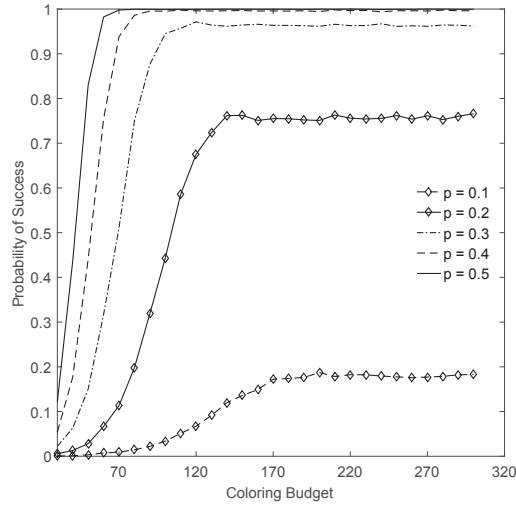


(a) 25 nodes and 5 nodes of interest.

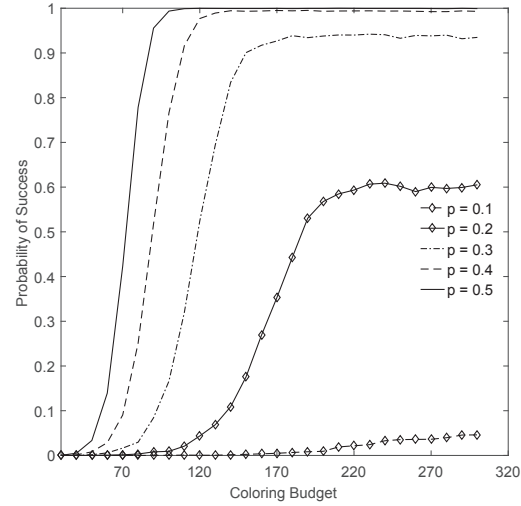


(b) 40 nodes and 9 nodes of interest.

Figure 2: Effect of varying the coloring success probability p_c .



(a) 25 nodes and 5 nodes of interest.



(b) 40 nodes and 9 nodes of interest.

Figure 3: Effect of varying the coloring budget B .

Figure 4 depicts the case with $N = 25$, $M = 5$, $p_c = 0.2$, illustrating the contribution of each failure mode. This suggests two regimes should be considered, based on parameter values, namely budget-constrained (BC) and time constrained (TC) regimes. In the BC regime, little can be done to improve the success probability without increasing the coloring budget, increasing p_c , or leveraging known structure about G . Within TC, modifying the exploration strategy to more aggressively explore before time ℓ could prove useful (trading off the risk of time expiration with that of prematurely coloring all of S).

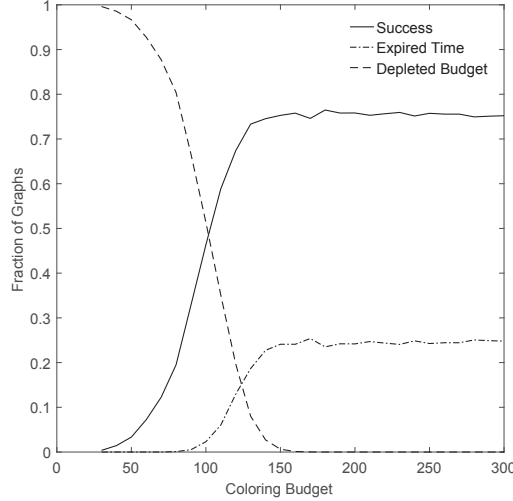


Figure 4: Comparison of failure modes.

4.2 Fully-Connected Graphs Without Uncoloring

If we consider the case of a fully-connected (or complete) graph, i.e., there exists an edge E between any two nodes in V , then it is possible to exploit the problem structure and simplify the analysis. We have created a stochastic optimization model for this problem, and the framework provides an upper bound on the maximum probability of successfully coloring all of the nodes in S . We next describe this optimization model.

First, it should be noted that in this case, all nodes are visible and available for coloring at the start of the problem. This is because the initial node that is colored must be connected to every other node in V . In this case, nodes can no longer be discriminated by their edge sets and as such, node identities can be pooled together into the following categories: nodes in S that are currently colored (w_1), nodes in S that are currently uncolored (x_1), nodes in $V \setminus S$ that are currently colored (w_2), nodes in $V \setminus S$ that are currently uncolored (x_2), and nodes that have never been colored. The decision maker's actions, at any decision epoch, can be similarly reduced to the number of currently uncolored nodes to attempt to recolor (y) and the number of nodes of unknown type to attempt to color (z). With these definitions we can formally define a Markov decision process (MDP) that has a state space of polynomial size in the model parameters. At time $t \in \mathcal{T}$ we define the system state to be $s^t = (w_1^t, w_2^t, x_1^t, x_2^t, b^t)$, with state space $\Gamma = \{0, 1, \dots, M\}^2 \times \{0, 1, \dots, N - M\}^2 \times \{0, 1, \dots, B\}$, action $a_t = (y^t, z^t)$, and action space $A(s^t) = \{0, 1, \dots, \min\{x_1^t, b^t\}\} \times \{0, 1, \dots, \min\{N - x_1^t - x_2^t - w_1^t - w_2^t, b^t\}\}$. It should also be noted that the permissible actions are further constrained by $y^t + z^t \leq b^t$. Additionally, in this model, there are no immediate rewards except for the terminal rewards. In particular, at time u and state s^u the reward $R_{a^u}(s^u)$ is simply the probability that the remaining nodes in S are successfully colored after taking action a^u . It should be noted that in case $p_u = 0$, the state reduces to $s^t = (w_1^t, w_2^t, b^t)$ and the action is simply $a^t = z^t$.

Next, we will derive the transition probabilities in the fully-connected case. Let us define the following random variables:

Z^t : the number of nodes belonging to S among the z_t that we attempt to color at t

X^t : the number of nodes in S that we successfully color at time t

W^t : the number of nodes in $V \setminus S$ that we successfully color at time t

In what follows, $HG(a, b, c)$ denotes a hypergeometric random variable with parameters a , b and c , and $Bin(n, p)$ denotes a binomial random variable with n trials and success probability p . At time t , we attempt to color z^t nodes from a finite population of size $N - w_1^t - w_2^t$ containing exactly $M - w_1^t$ nodes in S ; therefore, $Z^t | s^t, z^t \sim HG(N - w_1^t - w_2^t, M - w_1^t, z^t)$. For notational convenience, define the (conditional) random variable $\xi^t = Z^t | s^t, z^t$. Now, of the ξ^t nodes in S that we attempt to color, each node has a probability p_c of being successfully colored; hence, $X^t | s^t, z^t \sim Bin(\xi^t, p_c)$. Similarly, $W^t | s^t, a^t \sim Bin(z^t - \xi^t, p_c)$. With this information, we can explicitly characterize the probability of successfully coloring all the nodes in S by time u . In particular, this probability of success is $\mathbb{P}(X^u = M - w_1^u | s^u, z^u)$. Since this probability is monotone nondecreasing in z^u , we have that the optimal action at the terminal decision epoch is given by $z^*(s^u) = \min\{b^u, N - w_1^u - w_2^u\}$, i.e., the largest permissible action. Next, define value function $v(s^t)$, as the probability of success starting in state s^t . Then for the terminal state s^u , we have

$$\begin{aligned} v(s^u) &= \mathbb{P}(X^u = M - w_1^u | s^u, z^*(s^u)) \\ &= \begin{cases} p_c^{M-w_1^u}, & b^u \geq N - w_1^u - w_2^u, \\ p_c^{M-w_1^u} \frac{\binom{N-M-w_2^u}{b^u-M+w_1^u}}{\binom{N-w_1^u-w_2^u}{b^u}}, & M - w_1^u \leq b^u < N - w_1^u - w_2^u, \\ 0, & b^u < M - w_1^u. \end{cases} \end{aligned} \quad (1)$$

For decision epoch $t < u$, we can determine $v(s^t)$ by backwards recursion. In particular,

$$v(s^t) = \max_{a \in A(s^t)} \{ \mathbb{E}(v(s^{t+1}) | s^t, a) \} \quad (2)$$

$$= \max_{a \in A(s^t)} \left\{ \sum_i \sum_j \mathbb{P}(X^t = i, W^t = j | s^t, a) v(w_1^t + i, w_2^t + j, b^t - a) \right\}. \quad (3)$$

Now, it is clear that X^t and W^t are not independent random variables; however, they are *conditionally independent*, given Z^t . Thus, by conditioning and removing all zero terms, we obtain the following form of $v(s^t)$, which is useful for numerical implementation:

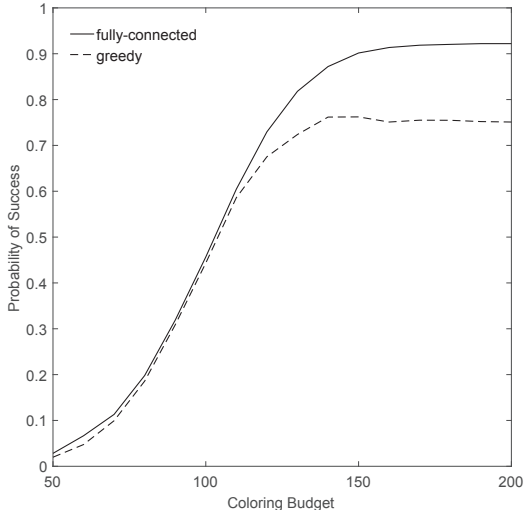
$$\begin{aligned} v(s^t) &= \max_{a \in A(s^t)} \left\{ \sum_{k=m}^{\bar{m}} \sum_{i=0}^k \sum_{j=0}^{a-k} \mathbb{P}(X^t = i | Z^t = k) \mathbb{P}(W^t = j | Z^t = k) \right. \\ &\quad \left. \times \mathbb{P}(Z^t = k | s^t, a) v(w_1^t + i, w_2^t + j, b^t - a) \right\}, \end{aligned} \quad (4)$$

or, more succinctly,

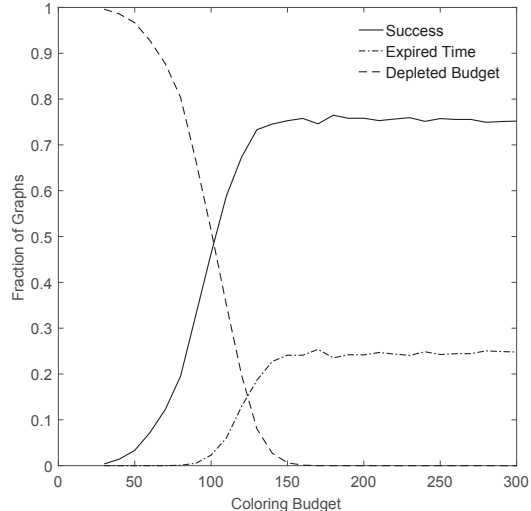
$$v(s^t) = \max_{a \in A(s^t)} \left\{ \sum_{k=m}^{\bar{m}} \sum_{i=0}^k \sum_{j=0}^{a-k} \binom{k}{i} \binom{a-k}{j} p_c^{i+j} (1-p_c)^{a-i-j} \frac{\binom{M-w_1^t}{k} \binom{N-M-w_2^t}{a-k}}{\binom{N-w_1^t-w_2^t}{a}} \right\}, \quad (5)$$

where $m := \max\{0, a - N + M + w_2^t\}$ and $\bar{m} := \min\{a, M - w_1^t\}$.

Next, we compare the performance of the MDP model with the baseline policy described in Section 4.1. We graphed the probability of success using the greedy policy and the optimal policy obtained via the MDP model in Figure 5(a) when $T = [12, 24]$, $N = 25$, $M = 5$, $p_c = 0.2$, and the budget B ranges from 50 to 200 in increments of 10. First, because



(a) Baseline versus optimal policies.



(b) Comparison of failure modes: baseline policy.

Figure 5: Greedy & fully-connected policies when greedy policy is budget- and time-constrained.

the complete graph is a relaxation of any other graph, it provides an upper bound on the expected performance of any policy. However, notice that when the budget is relatively small ($B \leq 100$), the primary failure mode of the baseline policy is a depleted budget (see Figure 5(b)), but the performance is very close to that of the optimal policy in Figure 5(a). However, when the budget is sufficiently large ($B > 100$) and the primary failure mode of the baseline policy is time expiration, the performance of the two policies differ significantly. This is because the optimal policy attempts to activate nodes more aggressively in the early time periods, resulting in (often significant) performance gains. Though not displayed here, it is worth noting that the value function exhibits monotonicity that can be exploited to improve scalability.

4.3 Fully-Connected Graphs With Uncoloring

We modify the state at time $t \in \mathcal{T}$ to be $s^t = (w_1^t, x_1^t, w_2^t, b^t)$, where w_1^t, x_1^t , and b^t are defined as before, $w_1^t + x_1^t \leq M$, and w_2^t is the number of nodes in $V \setminus S$ that have ever been colored. Therefore, the state space Γ satisfies $\Gamma \subseteq \{0, 1, \dots, M\}^2 \times \{0, 1, \dots, N - M\} \times \{0, 1, \dots, B\}$ and $|\Gamma| = (M + 1)(M + 2)(B + 1)(N - M + 1)/2$. The action at time t is $a_t = (z_t, y_t)$, where z_t is the number of nodes of unknown type to attempt to color, and y_t is the number of nodes known to be in S that we attempt to *recolor* at time t . For each $t \in \mathcal{T}$, define these random variables:

Z_c^t : of the z_t nodes of unknown type, Z_c^t is the number that belong to S

X_c^t : the number of nodes in S successfully colored for the first time

Y_c^t : the number of nodes in S successfully *recolor*ed

W_c^t : the number of nodes in $V \setminus S$ successfully colored for the first time

X_u^t : the number of nodes in S that uncolor.

After conditioning on the current state and action, these random variables are distributed as follows:

$$Z_c^t | s^t, a^t = Z_c^t | s^t, z^t \sim HG(N - x_1^t - w_1^t - w_2^t, M - x_1^t - w_1^t, z^t)$$

$$X_c^t | s^t, a^t = X_c^t | s^t, z^t \sim Bin(Z_c^t, p_c)$$

$$Y_c^t | s^t, a^t = Y_c^t | y^t \sim Bin(y^t, p_c)$$

$$W_c^t | s^t, a^t = W_c^t | s^t, z^t \sim Bin(z^t - Z_c^t, p_c)$$

$$X_u^t | s^t, a^t = X_u^t | w_1^t \sim Bin(w_1^t, p_u).$$

With these definitions, we are now prepared to derive the value function. Similar to the case when $p_u = 0$, first consider the terminal decision epoch at time $t = u$. If the current state is $w_1^u = M$, then we have already succeeded, and if $b^u < M - w_1^u$ then we have insufficient budget and our probability of succeeding is zero; otherwise, the optimal action is to attempt to recolor all of the nodes in S and expend the remaining budget on coloring nodes of unknown type. Therefore, $y^*(s^u) = x_1^u$ and $z^*(s^u) = \min\{b^u - x_1^u, N - x_1^u - w_1^u - w_2^u\}$, respectively. A successful outcome is then contingent on no nodes in S uncoloring, and the remaining uncolored nodes in S becoming colored. We can then see that the value function, at the last decision epoch u , is given by

$$\begin{aligned} v(s^u) &= \mathbb{P}(X_c^u = M - x_1^u - w_1^u, Y_c^u = x_1^u, X_u^u = 0 | s^u, z^*(s^u)) \\ &= \begin{cases} 0, & b^u < M - w_1^u \\ p_c^{M-w_1^u} (1-p_u) w_1^u \frac{\binom{N-M-w_2^u}{b^u+w_1^u-M}}{\binom{N-x_1^u-w_1^u-w_2^u}{b^u-x_1^u}}, & b^u \geq M - w_1^u, b^u - x_1^u < N - x_1^u - w_1^u - w_2^u. \\ p_c^{M-w_1^u} (1-p_u) w_1^u, & \text{otherwise} \end{cases} \end{aligned} \quad (6)$$

Next, we derive the general state-action transition probability, namely

$$\mathbb{P}(w_1^{t+1} = w_1^t + i, x_1^{t+1} = x_1^t + j, w_2^{t+1} = w_2^t + k | s^t, a^t).$$

We note that this probability is equivalent to

$$\mathbb{P}(X_c^t - X_u^t + Y_c^t = +i, X_u^t - Y_c^t = j, W_c^t = k | s^t, a^t). \quad (7)$$

Similar to the case where $p_u = 0$, W_c^t is conditionally independent of the other random variables, given Z_c^t . Thus, we begin by conditioning on Z_c^t to see that (7) is equivalent to

$$\sum_{n \in \mathcal{N}} \mathbb{P}(X_c^t - x_u^t + Y_c^t = +i, X_u^t - Y_c^t = j | Z_c^t = n, s^t, a^t) \mathbb{P}(W_c^t = k | Z_c^t = n, z^t) \mathbb{P}(Z_c^t = n | s^t, a^t), \quad (8)$$

where $\mathcal{N} = \mathbb{Z}_+ \cap [\max\{0, z^t - N + M + w_2^t\}, \min\{z^t, M - w_1^t - x_1^t\}]$. Now, we consider the second random variable in the first expression of (8), i.e., $X_u^t - Y_c^t$. By conditioning on Y_c^t , we see that

$$\begin{aligned} \mathbb{P}(X_u^t - Y_c^t = j | Z_c^t = n, s^t, a^t) &= \sum_{m=0}^{y^t} \mathbb{P}(X_u^t = j + m | Z_c^t = n, s^t, a^t) \mathbb{P}(Y_c^t = m | Z_c^t = n, s^t, a^t) \\ &= \sum_{m=0}^{y^t} \binom{w_1^t}{j+m} p_u^{j+m} (1-p_u)^{w_1^t-j-m} \binom{y^t}{m} p_c^m (1-p_c)^{y^t-m}. \end{aligned} \quad (9)$$

Additionally, we know that

$$\mathbb{P}(X_c^t = i + j | Z_c^t = n, z^t) = \binom{n}{i+j} p_c^{i+j} (1-p_c)^{n-i-j}, \quad (10)$$

$$\mathbb{P}(W_c^t = k | Z_c^t = n, z^t) = \binom{z^t - n}{k} p_c^k (1-p_c)^{z^t - n - k}, \quad (11)$$

and

$$\mathbb{P}(Z_c^t = n | s^t, a^t) = \frac{\binom{M-x_1^t-w_1^t}{n} \binom{N-M-w_2^t}{z^t-n}}{\binom{N-x_1^t-w_1^t-w_2^t}{z^t}}. \quad (12)$$

Since (8) is equivalent to

$$\begin{aligned} \sum_{n \in \mathcal{N}} \mathbb{P}(X_c^t = i + j | Z_c^t = n, s^t, a^t) \mathbb{P}(X_u^t - Y_c^t = j | Z_c^t = n, s^t, a^t) \\ \times \mathbb{P}(W_c^t = k | Z_c^t = n, z^t) \mathbb{P}(Z_c^t = n | s^t, a^t), \end{aligned} \quad (13)$$

we can compute (7) by combining equations (9) – (13).

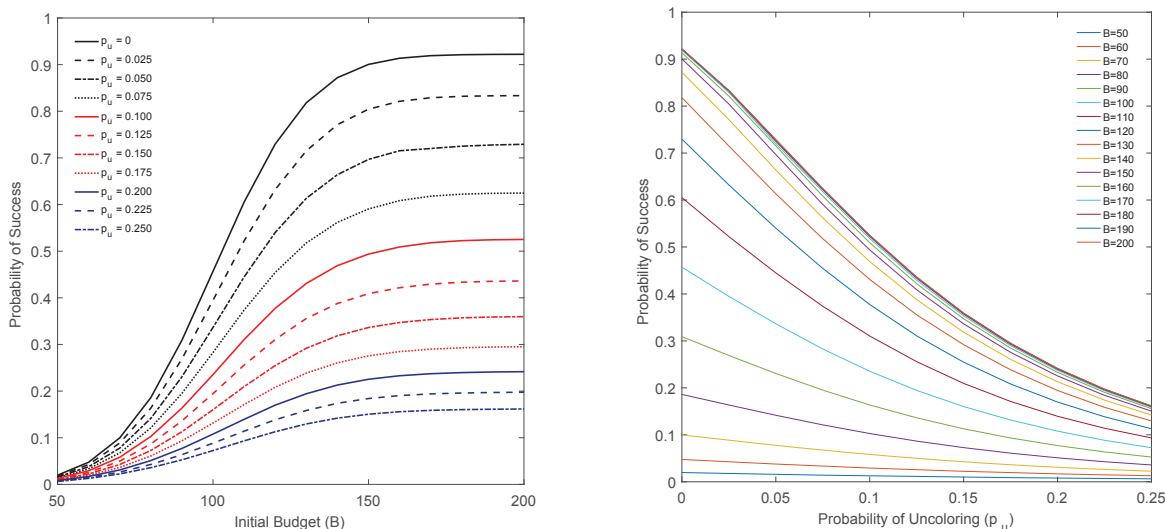
Hence, the value function can be computed by standard backwards dynamic programming. For $t < u$, by conditioning and removing all zero terms, we obtain the following form of $v(s^t)$:

$$\begin{aligned} v(s^t) = \max_{a \in A(s^t)} \left\{ \sum_{n \in \mathcal{N}} \sum_{m=0}^{y^t} \sum_{i=-w_1^t}^{n+y^t} \sum_{j \in \mathcal{J}} \sum_{k=0}^{z^t-n} v(w_1^t + i, x_1^t + j, w_2^t + k, b^t - z^t - y^t) \right. \\ \left. \times \mathbb{P}(X_c^t - X_u^t + Y_c^t = +i, X_u^t - Y_c^t = j, W_c^t = k | s^t, a^t) \right\}, \end{aligned}$$

where $\mathcal{J} = \mathbb{Z} \cap [\max\{-i, -m\}, \min\{w_1^t - m, n - i\}]$.

4.4 Numerical Illustrations and Results

As an extensive test bed, the problem with $T = [12, 24]$, $N = 25$, $M = 5$ and $p_c = 0.20$ was explored. In particular, all 176 combinations of the cases where $B \in \{50, 60, \dots, 200\}$ and $p_u \in \{0, 0.025, \dots, 0.250\}$ were considered. The simplest summary measure for each of these test cases is the probability of success. Figures 6(a) and 6(b) summarize the probability



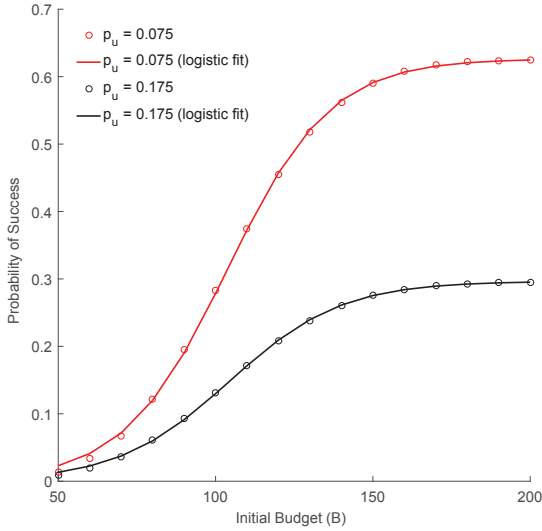
(a) $\mathbb{P}(\text{Success})$ as a function of the budget B .

(b) $\mathbb{P}(\text{Success})$ as a function of p_u .

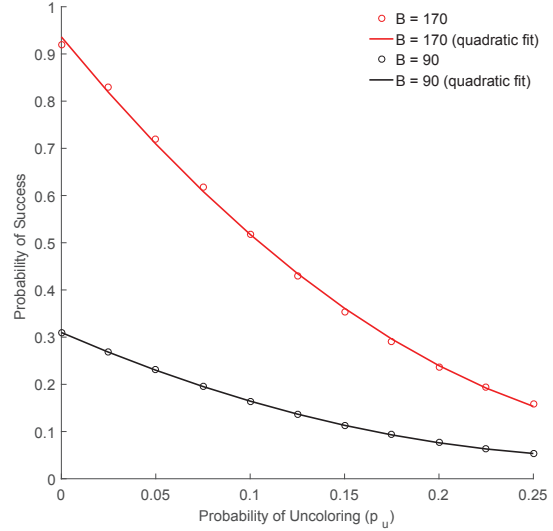
Figure 6: Behavior of the success probability as a function of B and p_u .

of success for these numerical tests. Some interesting properties are immediately evident. Specifically, it is seen in Figure 6(a) that the probability of success, as a function of the coloring budget, is a sigmoidal function (logistic functions of this type can be fit nearly exactly). This indicates that even in the case with uncoloring, there are regions of increasing, constant, and decreasing marginal returns associated with increasing the budget. Practically speaking, beyond a certain point, increasing the coloring budget is no longer an effective way of improving the success probability. Figure 6(b) demonstrates that the probability of success is decreasing, but with decreasing marginal differences, in the uncoloring probability p_u . Indeed, the least-squares fit seems to indicate that the relationship is quadratic (or very nearly quadratic).

Figures 7(a) and 7(b) depict the exact and fitted values for a few select curves. The quality of these fits, as well as the simplicity of their form, suggests that there may be some fundamental, analytic underpinning to these results. In all cases, the least-squares fit yields a coefficient of determination $R^2 > 0.999$. The accuracy of these fit curves would suggest that solving only a few instances is sufficient to accurately predict the probability of success over a large number of problem parameters. This type of analysis should be explored over additional cases where other parameters are varied.



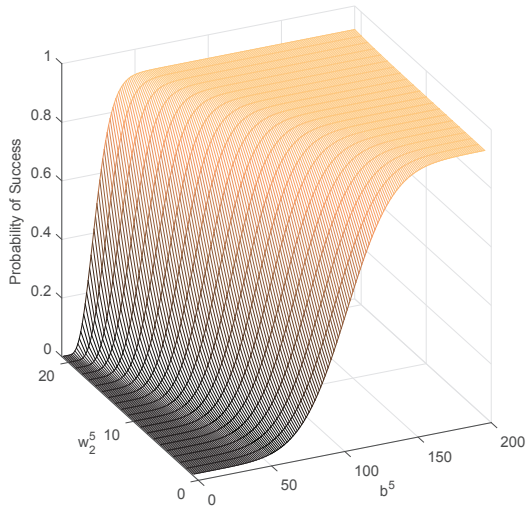
(a) Logistic function fits.



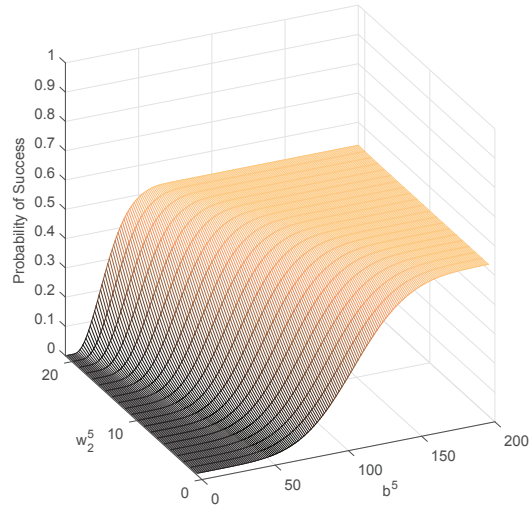
(b) Quadratic function fits.

Figure 7: Best-fit curves for best attainable success probability..

Finally, we examined the optimal actions and value functions. Figure 8 depicts the change in value function when increasing p_u from 0.0 to 0.1. As expected, the value function is monotone in b^5 and w_2^5 and never larger when there is uncoloring. Surprisingly, the shapes of the value functions are nearly identical. In both cases, the shape is sigmoidal in both dimensions and reaches a point of zero marginal returns at approximately the same point. This suggests that simple value function approximation schemes can be used in both models (with and without uncoloring).



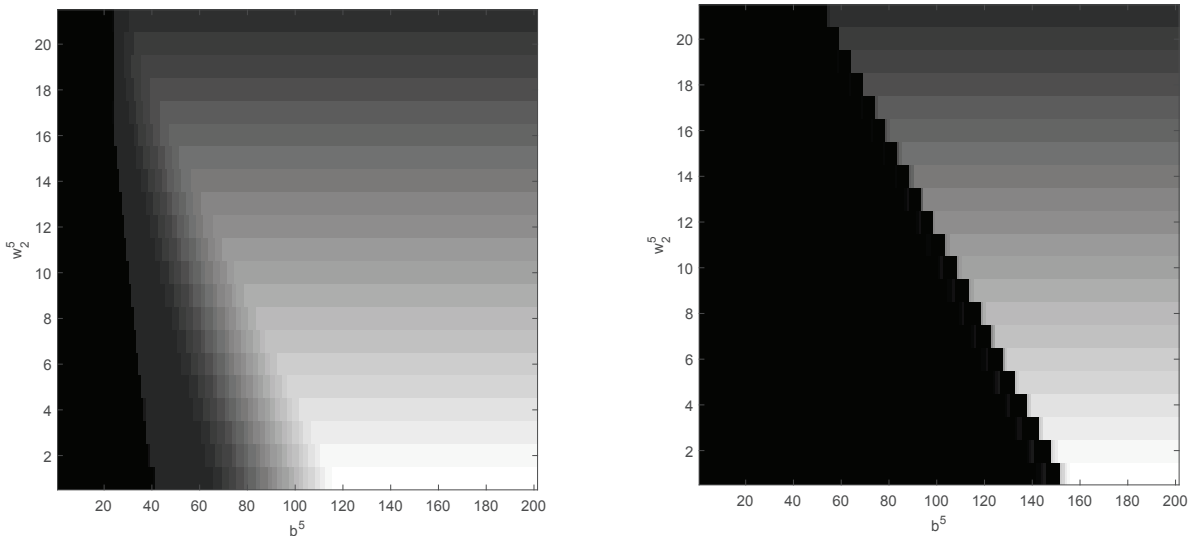
(a) Value function when $p_u = 0.0$.



(b) Value function when $p_u = 0.1$.

Figure 8: Value function comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). For both cases, the time is fixed at $t = 5$ and no nodes in S have been colored.

Figure 9 depicts the optimal actions and reveals several interesting insights: (i) the optimal actions are monotone nondecreasing in the remaining budget; (ii) when $p_u = 0.1$, the optimal action is either to do nothing or attempt to color all remaining nodes; and (iii) the optimal action when $p_u = 0$ is always at least as large as the optimal action when $p_u = 0.1$. The first point suggests a structural result that can be proven and exploited to improve solution time, while the second point suggests the optimality of a threshold-type policy. Regarding the third point, it may seem intuitive that, because uncoloring complicates the problem, we should color more aggressively early in the time horizon. However, if too many nodes in S are successfully colored early, some of those may uncolor before the start of the time horizon. It should be noted that these optimal policies are generally more aggressive than the baseline policy of subsection 4.1.

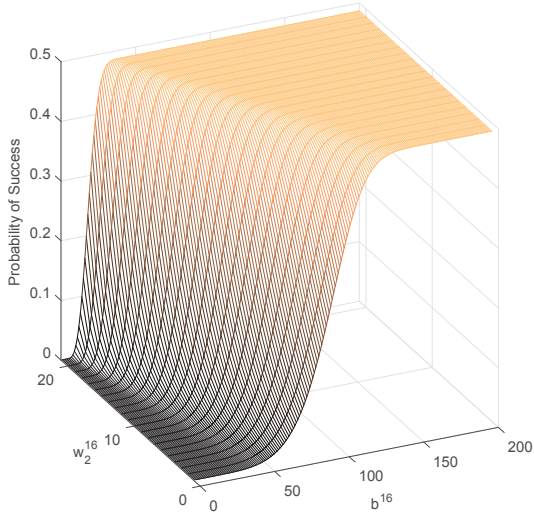


(a) Optimal actions without uncoloring.

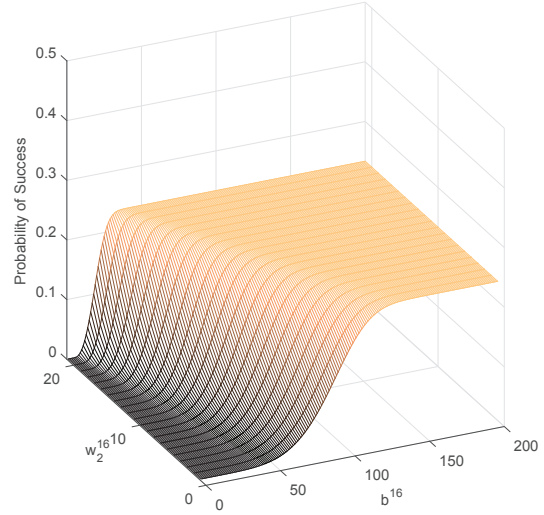
(b) Optimal actions when $p_u = 0.10$.

Figure 9: Comparing optimal actions with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). Brighter color indicates attempting to color a larger number of nodes. In both cases, the time is fixed at $t = 5$ and no nodes in S have been colored.

Figure 10 gives the value function for the example problems, but now at time $t = 16$. Once again the shapes of the functions are sigmoidal and nearly identical. In fact, if we let V_a denote the matrix of values plotted in Figure 10(a) and V_b denote the matrix of values plotted in Figure 10(b) then $\|0.5V_a - V_b\|_{\max} = 0.0077$. Hence, the shapes of the curves are nearly identical up to a multiplicative constant.



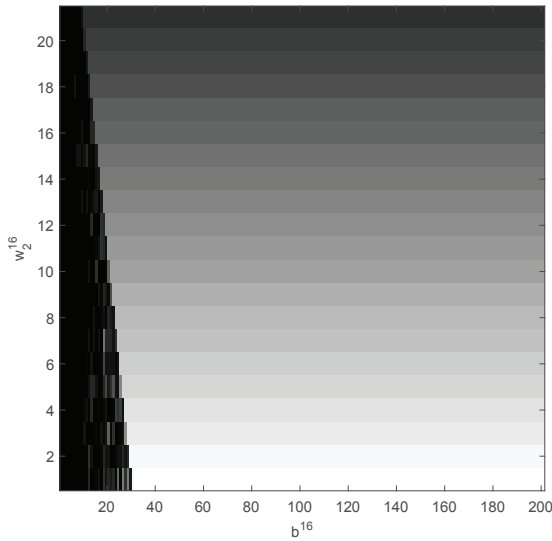
(a) Value function when $p_u = 0.0$.



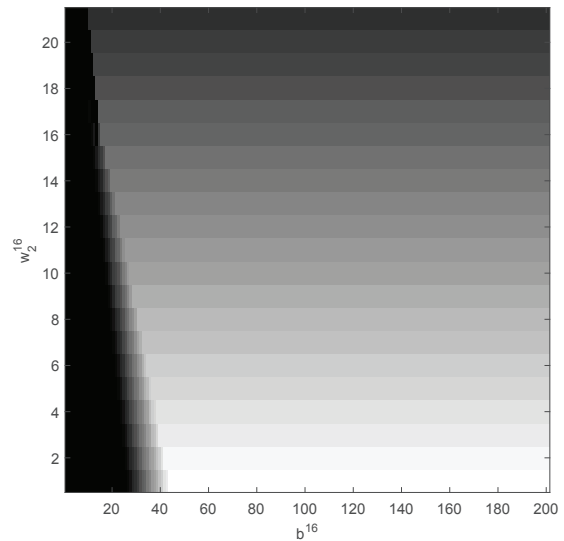
(b) Value function when $p_u = 0.1$.

Figure 10: Value function comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). In both cases, the time is fixed at $t = 16$ and no nodes in S have been colored.

Finally, Figure 11 depicts the optimal actions at $t = 16$. In this case, we still have that the optimal action is larger when there is no uncoloring, but the optimal policies are much closer to each other than during the earlier time period ($t = 5$).



(a) Optimal actions when $p_u = 0.0$.



(b) Optimal actions when $p_u = 0.10$.

Figure 11: Optimal action comparison with and without uncoloring ($T = [12, 24]$, $N = 25$, $M = 5$, $B = 200$, and $p_c = 0.2$). Brighter color indicates attempting to color a larger number of nodes. In both cases, the time is fixed at $t = 16$ and no nodes in S have been colored.

5.0 CONCLUSIONS

In this research project, we considered the problem of optimally locating and “coloring” a subset of nodes of interest within a graph. Within the context of our problem, coloring does not refer to the classical node-coloring problem of graph theory, but rather implies the altering of the binary status of a node. The main objective is to maximize the probability that all nodes of interest are successfully located and colored within a prescribed window, subject to a fixed coloring budget and coloring/uncoloring probabilities. We examined cases in which the nodes can and cannot uncolor independently after having been successfully colored.

The most interesting results stem from the third stochastic optimization model in which we considered a fully-connected graph within which the nodes can uncolor independently. While preliminary findings indicate extremely interesting problem structure – namely that the probability of success appears to assume an analytical form that might be predictable by knowing only a few problem parameters – further numerical experimentation and analytical work is needed to make definitive assertions about these relationships. Nonetheless, the apparent trends in our data are very strong and appear unlikely to be spurious.

The initial (baseline) model employs some simplifying assumptions but provides a framework for approaching the problem when more complex dynamics are introduced. Indeed, the baseline policy can be seen as one that provides a performance limit for other approaches. Identifying policies that perform well when the baseline policy fails is crucial. To this end, in future research, we will leverage the solutions to existing graph-theoretic problems (e.g., the stochastic Steiner tree problem) to devise more sophisticated policies. Additionally, exploiting the structural insights from our models, we will devise custom *reinforcement learning* algorithms for systematically improving existing policies and generating novel, near-optimal policies.

LIST OF ACRONYMS & GLOSSARY

DARPA – Defense Advanced Research Projects Agency

MDP – Markov Decision Process