NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

BUDGET-CONSTRAINED ROBUST INFLUENCE MAXIMIZATION

by

Jeremy L. Berg

March 2023

Thesis Advisor: Louis Chen
Second Reader: Johannes O. Royset

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, 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 Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC, 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2023	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE BUDGET-CONSTRAINED ROBUST INFLUENCE MAXIMIZATION			**5. FUNDING NUMBERS**
6. AUTHOR(S) Jeremy L. Berg			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Departing from traditional combinatorial, independent cascade influence maximization, we propose a continuous, correlation-robust influence maximization model. Instead of a deterministic seeding of nodes, a budgeted selection of discounts is now used to affect the likelihood of seeding. Additionally, edge probabilities are no longer assumed to be independent and are instead coupled adversarially. This model features a combination of increased computational tractability while also providing some means to express more sophisticated edge relationships or dependencies. We provide a study of the maximization problems, and show favorable performance of its solutions as compared to those of previous works assuming independence. More precisely, we measure the relative trade-off in performance between independent cascade and adversarial models. Further, we show that this proposed model can be used for networks with variable node rewards. We conclude with experiments on real-world datasets.

14. SUBJECT TERMS influence maximization, robust, sensitivity, independent cascade			**15. NUMBER OF PAGES** 47
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

BUDGET-CONSTRAINED ROBUST INFLUENCE MAXIMIZATION

Jeremy L. Berg
Lieutenant, United States Navy
BS, North Carolina State University, 2015

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2023

Approved by: Louis Chen
Advisor

Johannes O. Royset
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Departing from traditional combinatorial, independent cascade influence maximization, we propose a continuous, correlation-robust influence maximization model. Instead of a deterministic seeding of nodes, a budgeted selection of discounts is now used to affect the likelihood of seeding. Additionally, edge probabilities are no longer assumed to be independent and are instead coupled adversarially. This model features a combination of increased computational tractability while also providing some means to express more sophisticated edge relationships or dependencies. We provide a study of the maximization problems, and show favorable performance of its solutions as compared to those of previous works assuming independence. More precisely, we measure the relative trade-off in performance between independent cascade and adversarial models. Further, we show that this proposed model can be used for networks with variable node rewards. We conclude with experiments on real-world datasets.

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

List of Acronyms and Abbreviations

BRIM Budget-constrained Robust Influence Maximization

CIM Continuous Influence Maximization

CRIM Correlation Robust Influence Maximization

IC Independent Cascade

IM Influence Maximization

RC Relative Cost

THIS PAGE INTENTIONALLY LEFT BLANK

Executive Summary

Influence maximization is the process by which adoption of ideas, knowledge, or products spreads across a network is maximized based on an initially influenced "seed" set. Formalized by Kempe et al. (2003), the goal is the find the best seed set that maximizes the expected influenced spread throughout the rest of the network. A seminal model is the Independent Cascade (IC) spreading mechanism, through which each edge, or connection between two nodes, has a random associated influence spread probability. Influence starts at the seed nodes, who have one chance to influence their neighbors, who, if influenced, also get one chance to influence their neighbors, and the process continues until no new nodes are influenced. In IC, the edge probabilities are independent from each other, which might not be the case in an adversarial network.

In this study, we take the Continuous Influence Maximization (CIM) model by Yang et al. (2020) and compare influence performance against a Correlation Robust Influence Maximization (CRIM) model by Chen et al. (2020a). Instead of direct seed selection, the CIM model utilizes an overall budget to distribute discounts to individual nodes $c \in [0, 1]$ to persuade them to become seeds, then uses the IC spreading mechanism to measure performance of those discounts. Some nodes are early adopters that have a high probability of becoming a seed at a low discount, some nodes have a linear relationship with the discount given, while other nodes are resilient adopters requiring a high discount to become a seed. The CRIM model, on the other hand, chooses the seed set directly that maximizes performance in the most pessimistic or adversarial edge correlation network through the use of a computationally efficient linear program. We develop the Budget-constrained Robust Influence Maximization (BRIM) model that incorporates the discount concept into the efficient linear program format of the CRIM model. This study then compares the CIM optimized discounts in the adversarial network, and the CRIM optimized discounts in the IC or independent network to determine the expected cost of increasing the robustness of the discounts.

We develop an algorithm for solving the BRIM linear program and compare model performance on three datasets: (1) wikivote: A node represents a user and an edge represents a vote of user i for user j to become an admin (Leskovec and Krevl 2014), (2) astro-PH: A node represents a published author and edges are created between published co-authors within the Astro Physics community (Leskovec and Krevl 2014), and (3) Oahu Census: the Hawaiian island of Oahu's census data, which contains the latitude and longitude of concentrated population as well as the population value of each node (State of Hawaii Office of Planning and Sustainable Development 2010). We compare ratios of the non-optimized discounts for a spread mechanism against the optimized discounts for that spread mechanism, and also utilize a Relative Cost metric that measures the expected drop in IC performance against the increase in BRIM performance. We show that scenarios exist where the robustness of a discount set can be improved with little or no cost in IC performance. We additionally show that the traditional IC spread mechanism may be problematic in an emergency scenario, where emergency supplies are not expected to travel far from their starting point in the geographic network.

Acknowledgments

I would like to thank my advisor, Assistant Professor Louis Chen, for his guidance through model derivations and patience throughout the entire thesis process. Thank you for driving me forward, for motivating me to write even more lines of code, and for the wisdom and advice given throughout the past year.

I would also like to thank my beautiful wife Nicole, for the love and support as we tackled the thesis process. Thank you for listening to my mad ravings about iterative Python scripting, how to make graphs look pretty, and no doubt many more topics. Thank you for being my wife, and my best friend. It just keeps working.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 1:
Introduction

Influence Maximization (IM) is the process by which information spread across a network is maximized based on an initially influenced "seed" set of nodes. First mentioned by Domingos and Richardson (2001) then further studied and formalized by Kempe et al. (2003), the goal is to find the seed set that maximizes the expected influence spread throughout the rest of the network. Many stochastic models have been proposed for studying influence spread/dynamics, discussed in detail by Singh et al. (2022). A seminal model is the Independent Cascade (IC) model, which centers around a randomly generated network. Each edge, or connection between two nodes, has an associated influence spread probability, allowing the influence to spread through the network. Influence starts at an initially chosen or "seeded" set of nodes, where each seed node has one opportunity to influence each neighboring node, dependent on the edge probabilities. Any newly influenced nodes have the same single opportunity to influence their neighbors, and the influence continues to "cascade" until no new nodes are influenced and the process stops. IC requires the initial seed nodes to be selected, assumes independence between edge probabilities, and requires many thousands of Monte Carlo simulations due to its probabilistic nature.

1.1 Continuous Influence Maximization

Continuous Influence Maximization (CIM) as published by Yang et al. (2020) modifies the previously described IM process by changing the seed selection process from a binary-style selection process to a probabilistic process. Instead of selecting nodes to act as seeds, an overall budget is managed instead where nodes are given a discount between 0 and 1. The probability a node becomes a seed depends on both the level of discount given and the node's specific adoption probability function. This alteration changes the problem from a discrete namespace to a continuous namespace, as seed selection is no longer strictly binary. Yang et al. (2020) also incorporated a coordinate descent process, which works to further improve the expected performance of the discount distribution across the network. The CIM model results in an optimized discount distribution, but still assumes independence between edge probabilities and requires thousands of Monte Carlo simulations since it utilizes the IC process.

1.2 Correlation Robust Influence Maximization

Instead of maximizing the expected influence spread, Chen et al. (2020a) propose a correlation robust model that focuses on maximizing the influence spread in the worst-case edge correlation scenario. The knowledge of edge probabilities is still present, but the correlation between edges are adversarial instead of independent. While this model does still require an initial seed set selection, the Monte Carlo simulations are no longer required since this model is calculating the worst-case performance, which can be calculated directly through a linear program.

1.3 Budget-constrained Robust Influence Maximization

In this work, we propose a Budget-constrained Robust Influence Maximization (BRIM) model that incorporates the CIM budget concept and adoption probability functions into the Correlation Robust Influence Maximization (CRIM) model. This union between the two models will allow calculation of CIM's optimized solution performance in a heavily correlated network, and the drop in CIM performance once that optimized solution is instead optimized for the heavily correlated network. The goal of this study is to determine the magnitude of cost in expected performance relative to the robust performance improvement in the correlated network, which will empower leaders to make resource allocation decisions based on their individual level of risk-adverseness.

CHAPTER 2: Literature Review

After Kempe et al. (2003) formalizes the Influence Maximization problem and the $1 - 1/e$ approximation guarantee of the greedy approach, studies involving a variety of network applications and solving techniques followed. The concept was applied to a variety of networks including static networks (Mossel and Roch 2010, Chen et al. 2010, Budak et al. 2011), dynamic networks (Ohsaka et al. 2016, Yang et al. 2017, Aggarwal et al. 2012), and competing influence maximization (Bharathi et al. 2007, Tzoumas et al. 2012, Lu et al. 2013), for example. These studies all focus on the same objective- selecting the seed set that maximizes the expected influence spread with known and independent edge probabilities. The significant computational demand and subsequent runtime of the greedy / Monte Carlo solving method used by Kempe et al. (2003) led to the development of heuristics to improve one or both of these drawbacks, including restricting computations to "local influence regions" (Chen et al. 2010), solving based on node degree (Adineh and Baygi 2019), or coordinate descent improved greedy solutions (Yang et al. 2020).

The question of sensitivity or robustness is more recent, to the best of our knowledge first discussed by He and Kempe (2014), where edge probabilities are randomly drawn from a uniform distribution $p_{ij} \in U[0, 1]$ for each Independent Cascade iteration. While the edge probabilities are still independent, the objective is to determine the best seed set to maximize influence spread that is also robust to the unknown and variable edge probabilities. Additional models followed that consider the sensitivity of the Linear Threshold influence spread mechanism (Nannicini et al. 2020), variable edge probabilities (Chen et al. 2016), and across different influence spread mechanisms (He and Kempe 2016).

The CRIM model proposed by Chen et al. (2020a) instead assumes the edge probabilities are fixed or "true," but the correlation between edges is unknown and takes on the worst-case dependence. A deviation from previous works, the CRIM model can be solved exactly and efficiently, and can therefore be coupled to traditional Influence Maximization models for comparison. We are among the first to study the impact of increasing the robustness against edge correlation against the decrease in expected influence maximization performance.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 3: Models

Networks can be represented as a graph $G = \langle V, E \rangle$, where V is the set of users and E is the set of connections between users. In traditional IM, influence spreads from a seed node, $s \in S$, to its neighbors via the edges where the probability of influence spread across that edge is p_{ij}, or from node i to node j. In conventional models, the set of seed nodes, i.e., the seed set S, is the decision to be made. If the seeding process itself is instead uncertain, we may pursue the following alternative model in which now a discount factor c_i for each node i is to be decided. The discount c_i yields the likelihood $P_i(c_i)$ that node i is seeded as part of the initial seed set. Formally, we refer to $P_i(\cdot)$ as the adoption probability function for node i and $c_i \in [0, 1]$ as the discount factor given to node i. Generalizing the former notion of budget, the total amount of discounts is not to exceed a given budget K, i.e., $\sum_{i \in V} c_i \leq K$.

3.1 CRIM Model

The starting formulation for this study is the CRIM model shown in Equation 3.1, as derived by Chen et al. (2020a). CRIM is both a linear program and deterministic, so it is fast to solve and has a single optimal solution. This model does assume the edge probabilities given are exact, and the correlations between edges are chosen adversarially while remaining true to the edge probability values.

$$
\begin{aligned}
\max_{S} \min_{\pi} \quad & \sum_{i \in V} \pi_i \\
\text{s.t.} \quad & \pi_i - \pi_j \leq 1 - p_{ij} \quad \forall (i, j) \in E \\
& \pi_i = 1 \quad \forall i \in S \\
& 0 \leq \pi_i \leq 1 \quad \forall i \in V
\end{aligned}
\tag{3.1}
$$

The objective of the CRIM model is to find the seed set that maximizes the worst-case edge correlation performance. This can be done using a greedy algorithm, where nodes are sequentially added to S based on individual performance. While this method does not ensure optimality, it does carry a $(1 - 1/e)$ approximation guarantee of true optimality.

3.2 Adoption Probability Functions

As discussed and utilized by Yang et al. (2020), instead of directly selecting the nodes that will be in the seed set S, discounts are given to individual nodes to "persuade" them to become seeds. The probability of becoming a seed is dependent on the level of discount given, where nodes given no discount have no chance of becoming a seed and nodes given the maximum discount of 1 become a seed with 100% probability. To represent the fact that some individuals are eager to adopt new technologies and ideas while others are more resilient to change and will fight adoption, 3 adoption probability functions are used shown in Figure 3.1.

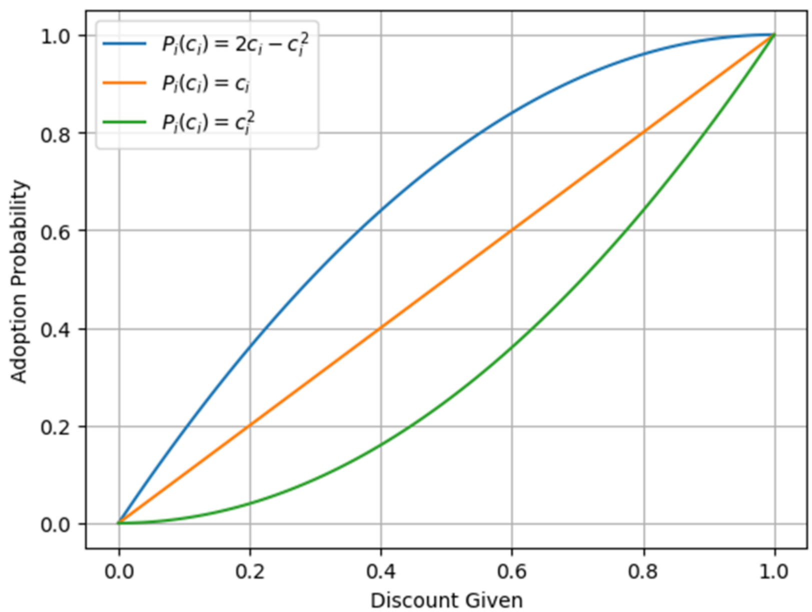

Figure 3.1. Adoption Probability Functions. Adapted from Yang et al. (2020).

Unless otherwise noted, 25% of nodes are assigned as early adopters $\left(P_i(c_i) = 2c_i - {c_i}^2\right)$, 50% as neither early adopters or resistant $(P_i(c_i) = c_i)$, and the remaining 25% of nodes are assigned as resistant $\left(P_i(c_i) = {c_i}^2\right)$.

3.3 BRIM Model

We first incorporate the discounts and the adoption probability functions into the CRIM model, shown in Equation 3.2. Instead of being given a seed set S, we are now giving the discount distribution $c \in [0,1]^V$ across the network that satisfies the budget $\sum_{i \in V} c_i \leq K$. We have also added the concept of node weights, w_i, which represent the relative value of node i. In the traditional IM scenario, $w_i = 1$ across all nodes, but in the event nodes represent entire apartment buildings or city blocks, w_i can instead be set to that relative value.

Theorem 1 *Let $P_i : [0,1] \rightarrow [0,1]$ adoption probability functions be given, along with edge probabilities $\{p_{ij}\}_{(i,j) \in E}$. Then, given any discount policy $c \in [0,1]^V$, the smallest (across all couplings of P_i and p_{ij}) expected number of influenced nodes is given by the following linear program*

$$\begin{aligned} \min_{\pi} \quad & \sum_{i \in V} w_i \cdot \pi_i \\ \text{s.t.} \quad & \pi_i - \pi_j \leq 1 - p_{ij} \quad \forall (i,j) \in E \\ & -\pi_i \leq -P_i(c_i) \quad \forall i \in V \\ & \pi_i \in [0,1] \quad \forall i \in V \end{aligned} \tag{3.2}$$

Proof: We refer the reader to Appendix A for the proof.

Maximizing Equation 3.2 across discounts $c \in [0,1]^V$ yields the Budget-Constrained Robust Influence Maximization (BRIM) model, shown in Equation 3.3.

$$\begin{aligned} \max_{c} \quad & Corr(c) \\ \text{s.t.} \quad & \sum_{i \in V} c_i \leq K \\ & c_i \in [0,1] \quad \forall i \in V \end{aligned} \tag{3.3}$$

Because c_i only appears in the constraints of Equation 3.3, the dual is taken with respect to π_i and further reduced. This changes the formulation from a $max - min$ problem to a

$max - max$ problem and includes c_i in the objective function, in the full formulation shown in Equation 3.4. We define the set S to be the first three constraints containing α_{ij} and β_i in Equation 3.4, and set D to be the last two constraints containing c_i in Equation 3.4. These sets are used in 4 to create subproblems for solving purposes.

$$\begin{aligned} \max_{\alpha,\beta} \max_{c} \quad & \left[L\left((\alpha,\beta), c\right) := -\sum_{(i,j)\in E} (1-p_{ij}) \cdot \alpha_{ij} + \sum_{i\in V} P_i(c_i) \cdot \beta_i \right] \\ \text{s.t.} \quad & (\alpha,\beta) \in S := \left\{ (\alpha \in R^E, \beta \in R^V) : -\sum_{j|i,j\in E} \alpha_{ij} + \sum_{j|j,i\in E} \alpha_{ji} + \beta_i \leq w_i \quad \forall i \in V \right\} \\ & \alpha_{ij} \geq 0 \quad \forall (i,j) \in E \\ & \beta_i \geq 0 \quad \forall i \in V \\ & c \in D := \{c \in [0,1]^V : \sum_{i\in V} c_i \leq K\} \end{aligned} \tag{3.4}$$

Convexity is not guaranteed in this form as the adoption probability functions can be non-convex and non-linearity exists in the $P_i(c_i) \cdot \beta_i$ term. This form is tractable however, as the adoption probability functions are individually convex or concave within the feasible discount range.

CHAPTER 4: Methods and Applications

In this chapter, we discuss the different solving approaches used to counter the non-linearity of Equation 3.4. These approaches are general, and can be used for any adoption probability function whose maximum or minimum occur within the $c_i \in [0, 1]$ discount constraint or at the extreme values of $c_i = 0$ or $c_i = 1$.

4.1 Combined Model Execution

To counter the additional solving complexity that the non-linear $P_i(c_i) \cdot \beta_i$ term presents, we break $L((\alpha_{ij}, \beta_i), c_i)$ from Equation 3.4 into two subproblems, $\max_{(\alpha,\beta)\in S} L((\alpha_{ij}, \beta_i), c)$ for a given c_i solution and $\max_{c\in D} L((\alpha_{ij}, \beta_i), c_i)$ for a given (α_{ij}, β_i) solution. Given an initial c_i solution that can be obtained from a greedy algorithm or other traditional CIM solving methods, $\max_{(\alpha,\beta)\in S} L((\alpha_{ij}, \beta_i), c)$ is a linear problem and can be solved efficiently. The linearity / non-linearity of $\max_{c\in D} L((\alpha_{ij}, \beta_i), c_i)$ now depends solely on the form of the adoption probability functions $P_i(c_i)$: if all nodes have a linear adoption probability function, then $\max_{c\in D} L((\alpha_{ij}, \beta_i), c_i)$ is also linear and can be solved efficiently. The adoption probability functions used in this study are non-linear in nature, and require the use of a non-linear problem solver.

For this study, we implement the iterative approach laid out in the $CoordAscent$ Algorithm across a variety of budgets, K, broken into increments of 10 for the social networks, or in increments of 1 for smaller networks. The Gurobi Optimization, LLC (2023) linear solver is used to solve $L(\alpha_{ij}, \beta_i), S$ while the IPOPT non-linear solver by Wächter and Biegler (2006) is used to solve $arg \max_{c\in D} L((\alpha^*_{ij}, \beta^*_i), c_i)$. This algorithm is also applied across varied adoption probability function distributions. In addition to the standard $P_i(c_i)$ distribution of 25% early adopters, 50% neither early adopter nor resilient, and 25% resilient, the solving algorithm is applied to the network when all nodes are early adopters and have non-convex adoption probability functions.

CoordAscent Algorithm $CoordinateAscent(\epsilon, K)$

Input: Tolerance $\epsilon > 0$, integer K
Output: c^R

Initialize $c^* = 0$
Generate CIM optimized c^{IC} distribution solution from Yang et al. (2020)
Hold the CIM solution c^{IC} values constant and solve

$$(\alpha^*_{ij}, \beta^*_i) < -arg \max_{(\alpha,\beta)\in S} L((\alpha_{ij}, \beta_i), c^{IC})$$

repeat

$$c^* \leftarrow arg \max_{c \in D} L((\alpha^*_{ij}, \beta^*_i), c_i), \qquad C^* \leftarrow \max_{c\in D} L((\alpha^*_{ij}, \beta^*_i), c_i)$$

$$(\alpha^*_{ij}, \beta^*_i) < -arg \max_{(\alpha,\beta)\in S} L((\alpha_{ij}, \beta_i), c^*), \qquad AB^* < - \max_{(\alpha,\beta)\in S} L((\alpha_{ij}, \beta_i), c^*)$$

until $|AB^* - C^*| < \epsilon$
$c^R \leftarrow c^*$

Upon termination of $CoordAscent$, we can compute the expected CIM performance (Yang et al. 2020) of c^R for comparison against the performance in the worst case, $Corr(c^R)$. We elaborate on these comparisons in the numerical experiments of Chapter 5.

4.2 Datasets

Our experiments were performed on three publicly available datasets (1) wikivote: A node represents a user and an edge represents a vote of user i for user j to become an admin (Leskovec and Krevl 2014), (2) astro-PH: A node represents a published author and edges are created between published co-authors within the Astro Physics community (Leskovec and Krevl 2014), and (3) Oahu Census: the Hawaiian island of Oahu's census data, which contains the latitude and longitude of concentrated population as well as the population value of each node (State of Hawaii Office of Planning and Sustainable Development 2010). Edges can be created a variety of ways based on existing roadways, city blocks, or in terms of geographical proximity in the event existing transportation infrastructure is destroyed by a natural disaster. This paper assumes the latter, where edge probabilities are created using a diminished-distance method up to a maximum travel distance of one mile. Pairwise distances are calculated between all nodes, and undirected edges are created between any two nodes that have less than one mile distance between them. The edge probability is a function of distance and is reduced with greater distance, shown in Equation 4.1, where d_{ij} is the distance from node i to node j. The node weights, w_i, are set to the population value at each node.

$$p_{ij} = 1 - \frac{1}{d_{ij}} \tag{4.1}$$

To ensure the variable node weights can be properly interpreted and utilized by the CIM model used by Yang et al. (2020) where each node only rewards a value of 1, the Oahu network is "extended." For example, a node with a population of 750 would be "extended" by creating and chaining 749 nodes from the parent node, all with edge probabilities of 1. This shift will significantly increase the number of nodes and edges present in the network and heavily reduce the average degree closer to 1, as shown in Table 4.1.

Table 4.1. Numerical Summary of Datasets

Network	Nodes	Edges	Average Degree
wiki-Vote	7,115	103,689	14.6
astro-PH	18,772	396,220	22.1
Oahu Census	653	4,802	7.35
"Extended" Oahu Census	953,207	957,356	1.004

4.3 Implementation

For the wiki-Vote and astro-PH networks, edge probabilities were assigned as $p_{ij} = 1/\nu_j$, where ν_j is the in-degree of node j. For the Oahu network, edge probabilities were assigned based on distance in Equation 4.1. We used the pyomo Python package for the problem formulation, the Gurobi solver when the problem was linear, and the IPOPT solver when the problem was non-linear. For IC solutions, Yang et al. (2020) and this paper used 80,000 Monte Carlo runs coded in C# for the expected influence spread. All calculations were run on a desktop with the AMD 5900X 12-core, 24-hyperthread 3.70 GHz processor with 32 GB of DDR4 RAM.

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 5: Experiments

We now discuss the experiments comparing IC and BRIM models. We also introduce the Relative Cost (RC) measure of effectiveness, shown in Equation 5.1, where $CIM(c^R)$ is the IC performance of c^R, $CIM(c^{IM})$ is the IC performance of c^{IM}, $Corr(c^R)$ is the BRIM performance of c^R, and $Corr(c^{IM})$ is the BRIM performance of C^{IM}. A high RC represents a large drop in IC performance, a small gain in BRIM, or both, indicating a high relative cost of increasing the robustness of the initial c^{IM} solution. Alternatively, a low RC represents a small drop in IC performance, a large gain in BRIM, or both, indicating a low relative cost of increasing the robustness of the initial c^{IM} solution. To avoid instances where increases in the robust network cannot be found and $Corr(c^R) - Corr(c^{IM}) = 0$, RC is not calculated when both the CIM and $Corr$ values change my less than 5%.

$$RC = \frac{CIM(c^R) - CIM(c^{IM})}{Corr(c^R) - Corr(c^{IM})}, \tag{5.1}$$

Due to the design of the $CoordAscent$ Algorithm, the denominator of Equation 5.1 cannot be negative, therefore a positive RC would indicate an improvement in both the IC and BRIM models. It is improbable that a solution can be improved in both influence models, thus RC is expected to be negative. The highest performing solution would increase robust performance with minimal or no detriment to expected IC performance, resulting in RC being close to zero.

5.1 Social Networks

In Figure 5.1, we show the performance of c^R and c^{IM} under both the IC and BRIM models by way of the ratios $\frac{CIM(c^R)}{CIM(c^{IM})}$ and $\frac{Corr(c^{IM})}{Corr(c^R)}$. When changing the discount policy from c^{IM} to c^R, these two ratios express, respectively, the performance loss under objective CIM and the performance gain under objective $Corr$. A ratio close to 1 indicates little difference between the solutions, while a large ratio indicates significant difference between the solution sets for that model. For example, a low $\frac{Corr(c^{IM})}{Corr(c^R)}$ ratio shows large improvement in the BRIM model. Figures 5.1 and 5.2 show the difference in ratios and the RC for the wiki-Vote and astro-PH networks across varying K, respectively.

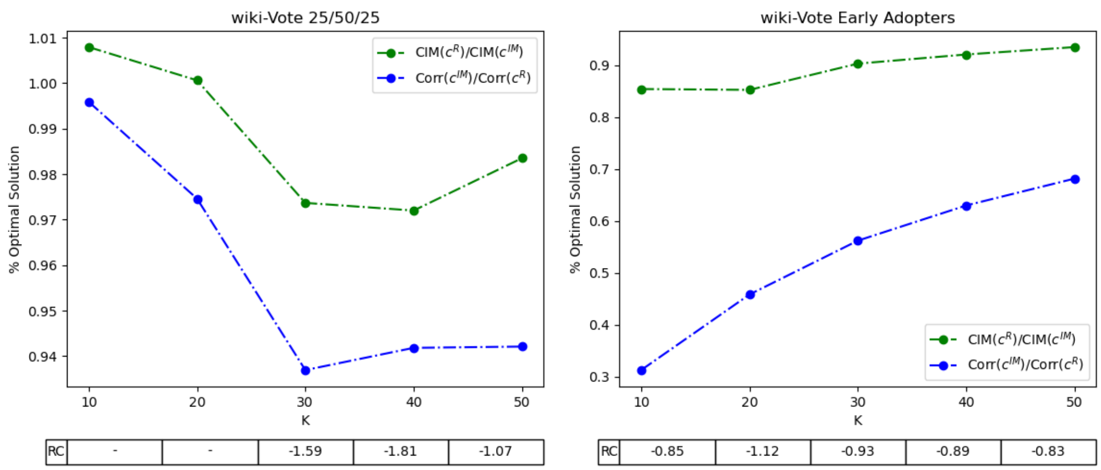

Figure 5.1. wiki-Vote IC and BRIM Optimized Solution Ratios

For lower budgets on the wiki-Vote network, the performance in the adversarial network can be improved with little to no loss in the IC model indicating we can increase the robustness of the c_i distribution. This is particularly evident in a network full of Early Adopters, as shown in the large gap between $\frac{CIM(c^R)}{CIM(c^{IM})}$ and $\frac{Corr(c^{IM})}{Corr(c^R)}$ at $K = 10$. The RC values are also closer to zero in the Early Adopters scenario, with a single-unit increase in BRIM performance reducing IC performance by less a single-unit (-0.85).

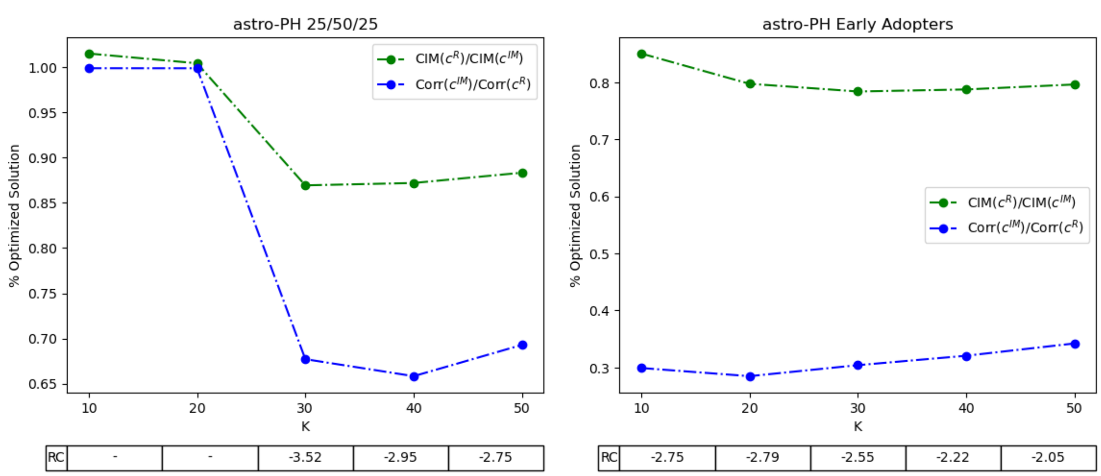

Figure 5.2. astro-PH IC and BRIM Optimized Solution Ratios

The astro-PH network shows that for lower budgets, the optimized IC solution could not be improved in the adversarial network. This may indicate that the IC solution is resilient to adversarial effects in this network, or could be a shortcoming of the iterative solving approach taken. We do see the same improvements in the Early Adopters scenario, nearly doubling the BRIM performance across all K values. The $RC \in [-2.79, -2.05]$ values of 2-3 indicate the single-unit robust performance increase comes at a reduction of ~2.5 units in the expected case. If the astro-PH network is thought to be even slightly adversarial, the decision maker would likely use the optimized BRIM discount distribution.

5.2 Oahu Network

The model ratios for the Oahu network are more intertwined than the social networks. This is likely due to the nature of the edge probabilities- where the social networks used $1/v_i$, where v_i is the indegree of node i, the edge probabilities of Oahu were made solely on geographic proximity. This allows nodes that are highly connected to still have high edge probabilities, increasing the chance for IC to propagate through much of the network, a characteristic the BRIM model does not share. We do see the ability to increase the BRIM performance when $K = 1$, after which the BRIM model chooses additional nodes in highly dense areas that perform poorly in the IC model.

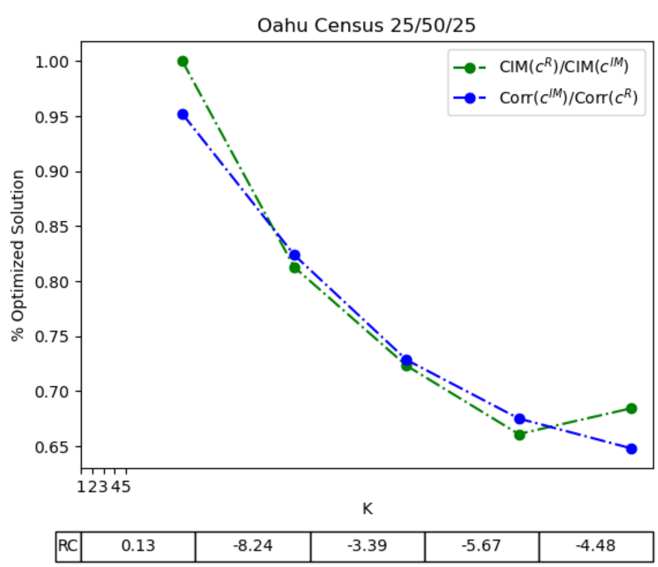

Figure 5.3. Oahu Census IC and BRIM Optimized Solution Comparison

THIS PAGE INTENTIONALLY LEFT BLANK

CHAPTER 6: Conclusions and Extensions

We have developed a model for IM where, given edge probabilities and an overall budget, discounts are chosen to select seeds that are robust in a network with adversarially selection edge correlations. The influence of the discount set can be efficiently computed in a linear program either independently using a simple greedy algorithm, or by using the optimized solution from a different influence spread mechanism. Using the RC metric and plotting the ratios between CIM and BRIM optimized solutions, we show instances where the edge correlation robustness of an initial CIM discount solution can be improved with little to no loss in the expected IC performance. Finally we show characteristics of real datasets that may not be conducive to the IC influence spread mechanism.

6.1 Broader Impact

The aim of this study is to address the edge independence assumption used in IC IM. In a network where an adversary decides correlation between edges, this independence assumption is shown to be costly in some scenarios. The methodology produced in this study allows comparison between IM performance assuming edge independence against the most-pessimistic edge correlations. This direct comparison allows the decision maker to determine the level of robust importance in a discount distribution, depending on how adversarial the network is predicted to be. This opens the opportunity for the development of a blended IC and Correlation Robust model where the decision maker can determine the level of network adversity to receive a combined optimized discount set. This can potentially offset the specific disadvantages of optimizing in an entirely independent or adversarial network.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX: Derivation and Data

A.1 BRIM Derivation

In the BRIM problem, the expected influence reduces to an expected max-flow value through the following network setup: flow moves from an artificial source, s, to all nodes in the network with edges taking an arbitrarily large capacity M with the adoption probability $P_i(c_i)$, or a capacity of 0 with probability $(1 - P_i(c_i))$. Edges within the network, $(i, j) \in E$, also have an arbitrarily large capacity M with probability 1. Flow similarly moves from all nodes in the network to the artificial sink t, along edges with a capacity of 1 with probability 1.

Using the Correlation Robust Maximum Flow formulation (Chen et al. 2020b) shown in Equation A.1, arc capacities are 0 or take one of K possible positive values, sorted $0 = c_o < c_1 < \ldots < c_K$. Each arc $(i, j) \in E$ has a randomly-initialized capacity given by a random variable $\tilde{u}_{ij}$ which has a CDF $F_{ij}(c_k) = Pr[\tilde{u}_{ij} \leq c_k]$ $\forall k \in K$ ($F_{ij}(c_K)$ is always 1), and PMF $f_{ij} = Pr[\tilde{u}_{ij} = c_k]$ $\forall k \in K$.

$$
\begin{aligned}
\min_{\pi,\lambda} \quad & \sum_{(i,j)\in E} \sum_{k=1}^{K} (c_k - c_{k-1}) \cdot \lambda_{ij}^k \\
\text{s.t.} \quad & 1 + \pi_t - \pi_s = 0 \\
& -\sum_{\tau=1}^{k-1} f_{ij}(c_\tau) + \pi_i - \pi_j \leq \lambda_{ij}^k \qquad \forall (i,j) \in E, k \in [K] \\
& \pi_t - \pi_i \geq -1; \qquad \forall i \in N \setminus \{s,t\} \\
& \pi_i - \pi_t \geq 0; \qquad \forall i \in N \setminus \{s,t\} \\
& \lambda_{ij}^k \geq 0; \qquad \forall (i,j) \in E, k \in [K] \\
& \pi_i \ free \qquad \forall i \in N
\end{aligned}
\tag{A.1}
$$

Applying the adoption probability functions to Equation A.1 yields Equation A.2, shown below.

$$\min_{\pi\in\mathbb{R}^V,\lambda^0,\lambda^1\in\mathbb{R}^{E'}} \quad \sum_{(i,j)\in E} 0\cdot\lambda^0_{ij} + M\cdot\lambda^1_{ij} + \sum_{i\in S} 0\cdot\lambda^0_{si} + M\cdot\lambda^1_{si} + \sum_{i\in N\setminus S} 0\cdot\lambda^0_{it} + \lambda^1_{it}$$
$$\begin{aligned}
\text{s.t.} \quad & \pi_i - \pi_j \le \lambda^0_{ij} \quad \forall (i,j)\in E \\
& \pi_i - \pi_j - (1-p_{ij}) \le \lambda^1_{ij} \quad \forall (i,j)\in E \\
& \pi_s - \pi_i \le \lambda^0_{si} \quad \forall i\in S \\
& \pi_s - \pi_i - (1-P_i(c_i)) \le \lambda^1_{si} \quad \forall i\in S \\
& \pi_i - \pi_t \le \lambda^0_{it} \quad \forall i\in V\setminus S \\
& \pi_i - \pi_t \le \lambda^1_{it} \quad \forall i\in V\setminus S \\
& 0\le \pi_i \le 1; \quad \forall i\in V \\
& \lambda^0_{ij},\lambda^1_{ij} \ge 0 \quad \forall (i,j)\in E' \\
& \pi_s = 1, \pi_t = 0
\end{aligned} \tag{A.2}$$

We can enforce $\lambda^1_{ij} = \lambda^1_{sj} = 0$ such that the constraint $\pi_i - \pi_j - (1-p_{ij}) \le \lambda^1_{ij}$ becomes $\pi_i - \pi_j \le (1-p_{ij})\ \forall (i,j)\in E,\ \pi_s - \pi_i - (1-P_i(c_i)) \le \lambda^1_{si}$ becomes $\pi_i \ge P_i(c_i)\ \forall i\in V$, and since $\lambda^0_{ij},\lambda^0_{si},\lambda^0_{it}$ are all unbounded, the three constraints of $\pi_i - \pi_j \le \lambda^0_{ij},\ \pi_s - \pi_i \le \lambda^0_{si},\ \pi_i - \pi_t \le \lambda^0_{it}$ can be removed. Substituting λ^1_{it} with its constraint equivalent of $\pi_i - \pi_t = \pi_i\ \forall i\in V\setminus S$ yields the unweighted formulation shown below in Equation A.3.

$$\begin{aligned}
\max_c \min_\pi \quad & \sum_{i\in V} \pi_i \\
\text{s.t.} \quad & \pi_i - \pi_j \le 1 - p_{ij} \quad \forall (i,j)\in E \\
& -\pi_i \le -P_i(c_i) \quad \forall i\in V \\
& 0\le \pi_i \le 1 \quad \forall i\in V
\end{aligned} \tag{A.3}$$

Should a node provide a variable integer reward w greater than 1, $(w-1)$ many artificial nodes can be chained to the initial node i with edge probabilities $p_{ij} = 1$. In this case $\pi_i = \pi_j$, where the objective function can be changed from $\max_c \min_\pi \sum_{i\in V} \pi_i$ to $\max_c \min_\pi \sum_{i\in V} w_i\cdot\pi_i$, where w_i is the reward value for node i. In the case of traditional IM where all nodes reward a value of 1, $w_i = 1\ \forall i\in V$ and the problem is unchanged.

A.2 Data Tables

The following are tables of the BRIM and IC influence spread mechanisms applied to their respective datasets, as well as the RC metric and percent difference between the IC solution c^{IM} and BRIM solution c^{R}.

Table A.1. wiki-Vote Network with 25/50/25 $P_i(c_i)$ Split

BRIM Performance					
K	10	20	30	40	50
Corr(c^{IM})	82.77	110.89	130.01	151.37	168.25
Corr(c^{R})	83.11	113.78	138.77	160.73	178.6
%Diff	~0%	2.6%	6.7%	6.2%	6.2%
IC Performance					
CIM(c^{IM})	292.97	428.72	527.39	605.36	671.68
CIM(c^{R})	295.30	428.99	513.50	588.40	660.60
%Diff	~0%	~0%	-2.6%	-2.8%	-1.7%
RC	-	-	-1.59	-1.81	-1.07

Table A.2. wiki-Vote Network with 100% Early Adopters

BRIM Performance					
K	10	20	30	40	50
Corr(c^{IM})	27.12	55.83	81.73	106.48	129.97
Corr(c^{R})	86.83	121..9	145.51	169.16	190.78
%Diff	220.2%	118.3%	78.0%	58.9%	46.8%
IC Performance					
CIM(c^{IM})	346.51	501.42	610.08	697.82	767.87
CIM(c^{R})	295.83	427.27	550.71	642.20	717.63
%Diff	-14.6%	-14.8%	-9.7%	-8.0%	-6.5%
RC	-0.85	-1.12	-0.93	-0.89	-0.83

Table A.3. astro-PH Network with 25/50/25 $P_i(c_i)$ Split

BRIM Performance					
K	10	20	30	40	50
Corr(c^{IM})	101.23	157.72	149.51	184.94	234.28
Corr(c^{R})	101.36	157.93	220.8	280.93	338.11
%Diff	~0%	~0%	47.7%	51.9%	44.3%
IC Performance					
CIM(c^{IM})	1034.25	1487.14	1918.41	2210.75	2446.52
CIM(c^{R})	1040.74	1493.36	1667.56	1927.45	2160.85
%Diff	~0%	~0%	-13.1%	-12.8%	-11.7%
RC	-	-	-3.52	-2.95	-2.75

Table A.4. astro-PH Network with 100% Early Adopters

BRIM Performance					
K	10	20	30	40	50
Corr(c^{IM})	29.54	52.70	81.38	112.79	142.44
Corr(c^{R})	98.78	185.1	267.66	351.83	416.34
%Diff	234.4%	251.2%	228.9%	211.9%	192.3%
IC Performance					
CIM(c^{IM})	1270.32	1819.92	2193.02	2497.79	2758.12
CIM(c^{R})	1080.15	1450.88	1718.74	1966.55	2195.56
%Diff	-15.0%	-20.3%	-21.6%	-21.3%	-20.4%
RC	-2.75	-2.79	-2.55	-2.22	-2.05

Table A.5. Oahu Census Network with 25/50/25 $P_i(c_i)$ Split

BRIM Performance					
K	1	2	3	4	5
Corr(c^{IM})	28173.0	42669.7	52487.8	61497.6	69462.3
Corr(c^{R})	29579.4	51806.9	72072.9	91126.6	107234.4
%Diff	5.00%	21.4%	37.3%	48.2%	54.4%
IC Performance					
CIM(c^{IM})	325435.5	401697.1	451408.3	495406.2	535940.1
CIM(c^{R})	325617.4	326424.4	326346.5	327312.4	366616.9
%Diff	~0%	-18.7%	-27.7%	-33.9%	-31.6%
RC	0.13	-8.24	-3.39	-5.67	-4.48

List of References

Adineh M, Baygi MN (2019) High quality degree based heuristics for the influence maximization problem. URL http://arxiv.org/abs/1904.12164.

Aggarwal CC, Lin S, Yu PS (2012) On influential node discovery in dynamic social networks. *Proceedings of the 2012 SIAM International Conference on Data Mining*, 636–647 (SIAM).

Bharathi S, Kempe D, Salek M (2007) Competitive influence maximization in social networks. *Internet and Network Economics: Third International Workshop, WINE 2007, San Diego, CA, USA, December 12-14, 2007. Proceedings 3*, 306–311 (Springer).

Budak C, Agrawal D, El Abbadi A (2011) Limiting the spread of misinformation in social networks. *Proceedings of the 20th International Conference on World Wide Web*, 665–674.

Chen L, Padmanabhan D, Lim CC, Natarajan K (2020a) Correlation robust influence maximization. *Advances in Neural Information Processing Systems* 33:7078–7089.

Chen LL, Ma W, Orlin JB, Simchi-Levi D (2020b) *Distributionally Robust Max Flows*, 81–90. URL http://dx.doi.org/10.1137/1.9781611976014.12.

Chen W, Lin T, Tan Z, Zhao M, Zhou X (2016) Robust influence maximization. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 795–804, KDD '16 (New York, NY, USA: Association for Computing Machinery), URL http://dx.doi.org/10.1145/2939672.2939745.

Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1029–1038.

Domingos P, Richardson M (2001) Mining the network value of customers. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 57–66 (San Francisco California: ACM), URL http://dx.doi.org/10.1145/502512.502525.

Gurobi Optimization, LLC (2023) Gurobi Optimizer Reference Manual. URL https://www.gurobi.com.

He X, Kempe D (2014) Stability of influence maximization. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1256–1265.

He X, Kempe D (2016) Robust influence maximization. URL http://dx.doi.org/10.48550/ARXIV.1602.05240.

Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146, KDD '03 (New York, NY, USA: Association for Computing Machinery), URL http://dx.doi.org/10.1145/956750.956769.

Leskovec J, Krevl A (2014) SNAP Datasets: Stanford large network dataset collection. http://snap.stanford.edu/data.

Lu W, Bonchi F, Goyal A, Lakshmanan LV (2013) The bang for the buck: fair competitive viral marketing from the host perspective. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 928–936.

Mossel E, Roch S (2010) Submodularity of influence in social networks: From local to global. *SIAM Journal on Computing* 39(6):2176–2188.

Nannicini G, Sartor G, Traversi E, Wolfler Calvo R (2020) An exact algorithm for robust influence maximization. *Mathematical Programming* 183(1):419–453, URL http://dx.doi.org/10.1007/s10107-020-01507-z.

Ohsaka N, Akiba T, Yoshida Y, Kawarabayashi Ki (2016) Dynamic influence analysis in evolving networks. *Proceedings of the VLDB Endowment* 9(12):1077–1088.

Singh SS, Srivastva D, Verma M, Singh J (2022) Influence maximization frameworks, performance, challenges and directions on social network: A theoretical study. *Journal of King Saud University - Computer and Information Sciences* 34(9):7570–7603, URL http://dx.doi.org/https://doi.org/10.1016/j.jksuci.2021.08.009.

State of Hawaii Office of Planning and Sustainable Development (2010) Hawaii Statewide GIS Program. https://planning.hawaii.gov/gis/download-gis-data-expanded/.

Tzoumas V, Amanatidis C, Markakis E (2012) A game-theoretic analysis of a competitive diffusion process over social networks. *Internet and Network Economics: 8th International Workshop, WINE 2012, Liverpool, UK, December 10-12, 2012. Proceedings 8*, 1–14 (Springer).

Wächter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106:25–57.

Yang Y, Mao X, Pei J, He X (2020) Continuous influence maximization. *ACM Trans. Knowl. Discov. Data* 14(3), URL http://dx.doi.org/10.1145/3380928.

Yang Y, Wang Z, Pei J, Chen E (2017) Tracking influential individuals in dynamic networks. *IEEE Transactions on Knowledge and Data Engineering* 29(11):2615–2628.

THIS PAGE INTENTIONALLY LEFT BLANK

Initial Distribution List

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California