Designing Effective and Practical Interventions to Contain Epidemics

Prathyush Sambaturu
University of Virginia
Charlottesville, VA
pks6mk@virginia.edu

Bijaya Adhikari
Virginia Tech
Blacksburg, VA
bijaya@cs.vt.edu

B. Aditya Prakash
Georgia Institute of Technology
Atlanta, GA
badityap@cc.gatech.edu

Srinivasan Venkatramanan
University of Virginia
Charlottesville, VA
srini@virginia.edu

Anil Vullikanti
University of Virginia
Charlottesville, VA
vsakumar@virginia.edu

ABSTRACT

Vaccination is a standard public health intervention for controlling the spread of epidemics. However, the supply of vaccines is typically limited, and therefore, their deployment needs to be optimized. Further, vaccines are produced over time, so the strategies have to be temporal. We study the problem EpiControl of designing vaccination strategies, within available budget constraints, to minimize the spread of an outbreak.

This is a challenging stochastic optimization problem. We design a bicriteria approximation algorithm, which combines a linear programming based rounding, along with the sample average approximation technique. Our approach also provides the empirical approximation factor for the problem instance, relative to the optimum. We find that the approximation factor is significantly better than the worst case bound, and, in practice, is a small constant factor. Further, our method shows significantly better performance than all prior heuristics for this problem. With additional pruning techniques, we are able to scale our algorithm to networks with millions of edges.

KEYWORDS

computational epidemiology, network science, optimization, computational social science

ACM Reference Format:
Prathyush Sambaturu, Bijaya Adhikari, B. Aditya Prakash, Srinivasan Venkatramanan, and Anil Vullikanti. 2020. Designing Effective and Practical Interventions to Contain Epidemics. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 9 pages.

1 INTRODUCTION

Vaccination and social distancing are the primary strategies for controlling the spread of epidemic outbreaks [4, 17, 21, 23, 25, 27–29, 38–40]. The production of vaccines is expensive and time intensive, and so there is always a shortage of vaccine supply [13]. As a result, there is a lot of interest in evaluating different kinds of interventions [17, 21], and finding optimal interventions [23].

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand. © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

The spread of epidemics is very complex, and SIR type diffusion processes (and variations) are commonly used: informally, each infected individual u (in state I) spreads the infection to each susceptible neighbor v (in state S) with transmission probability $p(u, v)$ (or p, if this is uniform for all edges); this is defined formally in Section 2. These can be modeled as a system of differential equations [23, 36, 38], or as stochastic agent based models on social contact networks, e.g., [22]. Differential equation models are small enough that they can be solved optimally by simple brute-force local search methods [23].

However, network and agent based models cannot be easily optimized this way. In this paper, we study EpiControl, which involves *designing vaccination strategies to minimize the expected outbreak size in an SIR epidemic process on a network* $G = (V, E)$. There is a lot of relevant prior work on this topic, and can be split along the following lines: (1) Optimization in the SI/SIS type models, in static or dynamic networks, e.g., [25, 27–30]. Much of this work is based on controlling spectral properties, but does not give any guaranteed bounds on the expected outbreak size; (2) Firefighter problem, which can be viewed as EpiControl on SI model with with $p = 1$, e.g., [3, 12]. Rigorous bounds are known for the number of infected and saved. However, this has not been studied much for the case where $p < 1$, except [34], which only gives rigorous algorithms for trees. A special case of this problem is with work of [4], which considers EpiControl but with the intervention specified at time 0; (3) Static interventions in SIR models, e.g., [39, 40], which also do not directly bound the expected outbreak size, (4) Heuristics for picking nodes based on degree or centrality, e.g., [8, 24], which work for all models but give no guarantees. In summary, none of the prior results directly address the EpiControl problem for the SIR models of epidemics, with $p < 1$.

Our results. We present algorithm saaRound for the EpiControl problem. Our specific contributions are described below.

- We design algorithm saaRound for selecting a set of nodes within a given budget, to vaccinate at the start of the epidemic (the 1sEpiControl problem). We show a rigorous worst case approximation guarantee on the performance of saaRound, which is logarithmic in the number of paths in sampled subgraphs of G; this is typically significantly smaller than the number of paths in G, so that in practice, saaRound has a much smaller approximation ratio. Our main technical ideas are rounding a linear programming (LP) relaxation, along with the sample average technique.

By comparing with the LP objective, we are able to obtain an empirical approximation guarantee for any instance. We show that SAAROUND is a good heuristic for the multi-stage problem as well, and gives similar guarantees as the single-stage when the disease transmission subgraphs are trees (e.g., when p is low).
- We augment SAAROUND with a sparsification step, which significantly reduces the size of the LP, and is able to scale to networks with millions of edges.
- We evaluate our algorithms on diverse real and random networks. We show that SAAROUND has an approximation ratio very close to 1, significantly better than the worst case guarantee we prove rigorously. We find that SAAROUND outperforms two of the most commonly used baselines for intervention design. We also examine the structure of solutions, and find significant differences in the characteristics of nodes picked in different stages.

At the start of every flu outbreak, and during major pandemics, there is a lot of interest from the CDC and other public health agencies on finding optimal solutions in different models, e.g., [17, 21, 23, 36]. These can be used to guide policies when there are shortages, including the logistics of where vaccines should be deployed [36]. Our methods can help in designing effective policies using agent based models, which have been found to be more useful in planning for large outbreaks, e.g., [11, 15, 17, 21].

2 NOTATION AND PROBLEMS

Network and disease model (see Table 1). Let $G = (V, E)$ be a contact graph where V is the set of people (or nodes) and $e = (u, v) \in E$ if nodes $u, v \in V$ come into direct contact, which can allow a disease to spread. Let $n = |V|$ be the number of nodes in graph G. We assume a simple SIR model of disease spread [22], in which each node is in one of the following states: susceptible (S), infectious (I) or recovered (R). The epidemic starts at one or more externally infected nodes, and spreads from an infected node u to each susceptible neighbor v with probability p. An infected node becomes recovered in the next time step. We assume s_v is the probability that v is initially infected; $\mathbf{s}$ denotes the initial infection vector. Let $\texttt{EInf}(G, \mathbf{s})$ denote the expected number of infections in this process. Let $H^{(sir)} =< I(0), \ldots, I(\tau), E' >$ denote a stochastic outcome from the SIR process in this case, in which (a) $I(t)$ denotes the set of nodes that are infected at time t, (b) $I(0)$ is the source nodes, and (c) E' is the (random) subset of edges on which the infection spread. Then, $\texttt{EInf}(G, \mathbf{s}) = E_{H^{(sir)}}[\sum_t |I(t)|]$.

Note. The SIR model generalizes the well studied *independent cascades* model [18]. Also, there are lots of variations of the SIR model, such as: varying transmission probability $p(u, v)$ on each edge (u, v), with an exposed state, varying infectious duration, etc.

Interventions and objective. We use x_{vt} as an indicator variable, which is 1 if node v gets vaccinated at time t; let $\mathbf{X}_t = \{v : x_{vt} = 1, v \in V\}$, and let B_t denote the number of vaccines available for use at time t. We assume the vaccine is immediately effective. In our discussion, we primarily focus on interventions at one or two time steps, and use the following notation:

(1) Single stage: $\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0)$ denotes the expected number of infections when the intervention is done for set $\mathbf{X}_0$ at time $t = 0$. Extending the earlier notation, let $H^{(sir)}(\mathbf{X}_0) =< I(0), \ldots, I(\tau), E' >$ denote a stochastic outcome from the SIR process in this case (so none of the nodes in $\mathbf{X}_0$ is part of any $I(t)$).

(2) Two stage: $\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0, \mathbf{X}_T)$ denotes the expected number of infections if the interventions are done on sets $\mathbf{X}_0$ and $\mathbf{X}_T$ at times 0 and T, respectively. In this case, $H^{(sir)}(\mathbf{X}_0, \mathbf{X}_T)$ denotes a stochastic outcome. We drop G and $\mathbf{s}$, when it is clear from the context.

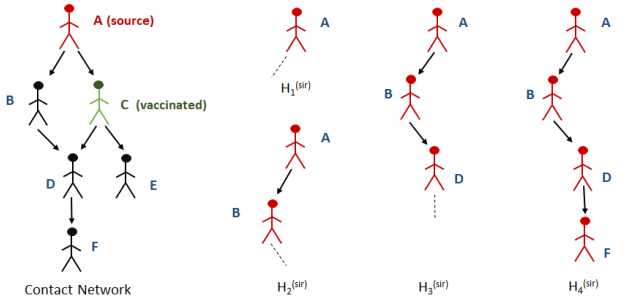

Figure 1: Example illustrating the SIR model: the contact network $G = (V, E)$ is shown in the left, with $V = \{A, B, C, D, E, F\}$ (shown in circles), and edges as solid lines.Node A is initially infected, and node C is vaccinated. The four subgraphs $H_1^{(sir)}, H_2^{(sir)}, H_3^{(sir)}, H_4^{(sir)}$ (on the right) are possible stochastic outcomes in the SIR model.

Notation	Definition
$G = (V, E)$	Graph
$\mathbf{s}$	Source distribution
p, $p(u, v)$	Transmission probability
x_{vt}	Indicator for node v vaccinated at time t
$\mathbf{X}_t$	Set of nodes vaccinated at time t
$\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0)$	Exp. #infections for single stage with intervention $\mathbf{X}_0$
$\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0, \mathbf{X}_T)$	Exp. #infections for two stage version
1sEPICONTROL	Single stage vaccination problem
2sEPICONTROL	Two stage vaccination problem
(α, β) approximation	Bicriteria approximation factor

Table 1: Summary of notation used in the paper.

Example. Figure 1 shows the SIR model and the definitions of the above quantities on a graph G with six people. Initially, A is infected, and C is vaccinated. In the SIR model, the disease spreads from an infected person to each susceptible neighbor with probability p, and does not spread with probability $1 - p$. Therefore, we have four possible stochastic outcomes $H_1^{(sir)}, \ldots, H_4^{(sir)}$, which occur with probabilities $1-p$, $p(1-p)$, $p^2(1-p)$, and p^3, respectively. Suppose we have $\mathcal{T} = \{0\}$. Then, $x_{C0} = 1$, and $\mathbf{X} = \{\mathbf{X}_0\} = \{\{C\}\}$. We have

$$\texttt{EInf}(\mathbf{X}) = (1-p) + 2p(1-p) + 3p^2(1-p) + 4p^3$$

Problem statement. Given a contact network $G = (V, E)$, and an initial infection vector $\mathbf{s}$, we consider the following problems
- **Single stage vaccination problem (1sEPICONTROL):** given budget B_0, choose $\mathbf{X}_0$ such that $|\mathbf{X}_0| \le B_0$ and $\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0)$ is minimized.
- **Two stage vaccination problem (2sEPICONTROL):** given budget B_0, B_T, choose $\mathbf{X}_0, \mathbf{X}_T$ such that $|\mathbf{X}_0| \le B_0$, $|\mathbf{X}_T| \le B_T$, and $\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0, \mathbf{X}_T)$ is minimized.

For conciseness, we refer to them as EPICONTROL problems. Our methods can be extended to more than two time steps, but we omit them to simplify the notation and discussion.

Bi-criteria approximate solution. For the 1sEpiControl problem, we say that an intervention $\mathbf{X}_0$ is an (α, β) - approximation if: (1) $|\mathbf{X}_0| \leq \alpha B_0$, and (2) $\texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0) \leq \beta \texttt{EInf}(G, \mathbf{s}, \mathbf{X}_0^*)$, where $\mathbf{X}_0^*$ is an optimal solution. We say that an algorithm is an (α, β)-approximation algorithm, if it gives an (α, β)-approximate solution. This notion also extends to the multi-stage intervention problem.

We use the following versions of the Chernoff bound.

Theorem 2.1. *(Theorem 1.1 of [10]) Let $Z = \sum_{i=1}^{n} Z_i$, where Z_i are independently distributed random variables in $[0, 1]$. Then, for any $\epsilon \in (0, 1)$, we have $\Pr[Z \notin [(1-\epsilon)E[Z], (1+\epsilon)E[Z]]] \leq 2exp(-\epsilon^2 E[Z]/3)$. Also, for any $t > 2eE[Z]$, $\Pr[Z > t] \leq 2^{-t}$.*

3 OUR APPROACH

We first present algorithm saaRound for the 1sEpiControl problem, using a linear programming rounding technique, combined with the sample average approximation technique from stochastic optimization. We then show that this can be significantly speeded up by augmenting it with a sparsification step. Finally, we discuss how to extend this approach to the multi-stage version.

3.1 Algorithm saaRound for 1sEpiControl

Algorithm 1 describes saaRound. It is based on a linear program, in which the variables x_{v0} are indicators for node v getting vaccinated at $t = 0$ as defined in Section 2. The variables y_{vj} are indicators for node v getting infected in sample H_j (i.e., there is a path from $\texttt{src}(H_j)$ to v with no node on it vaccinated). We first describe the intuition behind the algorithm, and then analyze its performance.

Algorithm 1 saaRound
Input: $G = (V, E), \mathbf{s}, B_0$
Output: X_0

1: Construct a sampled graph $H_j = (V, E_j)$, for $j = 1, \ldots, M$, by picking each edge $e \in E$ to be in E_j with probability p. Also pick a set of sources $\texttt{src}(H_j)$ by sampling from $\mathbf{s}$
2: Solve the following linear program (LP_{saa})

$$(LP_{saa}) \quad \min \frac{1}{M} \sum_j \sum_v y_{vj} \tag{1}$$

$$\forall j, \forall u \in V: \; y_{uj} \leq 1 - x_{u0} \tag{2}$$

$$\forall j, \forall u \in V, \; (w, u) \in E_j: \; y_{uj} \geq y_{wj} - x_{u0} \tag{3}$$

$$\forall j, \forall s \in src(H_j): \; y_{sj} = 1 - x_{u0} \tag{4}$$

$$\sum_{u \in V} x_{u0} \leq B_0 \tag{5}$$

$$\text{All variables} \in [0, 1] \tag{6}$$

3: Let x, y be the optimal fractional solution to (LP). We round it to an integral solution X, Y in the following manner
   (1) If $y_{vj} \in \{0, 1\}$, set $Y_{vj} = y_{vj}$. Similarly, if $x_{v0} \in \{0, 1\}$, set $X_{v0} = x_{v0}$.
   (2) Round $Y_{vj} = 1$ for each (v, j) if $y_{vj} \geq \frac{1}{2}$, otherwise set $Y_{vj} = 0$.
   (3) For each v, set $X_{v0} = 1$ with probability $\min\{1, 2x_{v0} \log(4nMN)\}$, where N is the maximum number of paths from $\texttt{src}(H_j)$ to any node v in H_j.
   (4) $X_0 = \{v : X_{v0} = 1\}$ is the set of nodes vaccinated.
4: **return** X_0

3.2 Intuition behind saaRound and its analysis

Our algorithm involves four key ideas, which are described below, along with an intuitive description of the steps of the algorithm.

- **Sample average approximation technique**: The first idea (see, e.g., [33]) is that it suffices to get a solution which minimizes the average number of infections in a set of M sampled outcomes, in order to minimize $\texttt{EInf}(\cdot)$, which is an expectation over all possible outcomes; we show that it suffices that M is bounded by a polynomial in n. Further, instead of actually using simulation outcomes, Step 1 and a breadth-first search (BFS) in each simulation from sources exploit an equivalence between an SIR process and percolation, making this much more efficient.
- **Compact integer program**: We show that using the structure of the SIR model, it suffices to work with sampled subgraphs H_j, instead of the stochastic outcomes of the SIR process. The problem is challenging even if we have to minimize the average number of infections restricted to $H_1, \ldots, H_M$. We start with an integer program (IP) which expresses the following constraints: if a node u is not infected in H_j (which is indicated by $y_{uj} = 0$), then for every path P from a node in $\texttt{src}(H_j)$ to u in H_j, there must be a node v which has been vaccinated. However, such an integer program would have exponentially many constraints (one for each path). Instead, we design a more compact program (referred to as IP_{saa}), simply based on states of nodes on an edge, as expressed in constraints (3).
- **Linear relaxation**: IP_{saa} cannot be solved in polynomial time, and we consider a linear relaxation of it (referred to as LP_{saa}), by replacing the binary constraints by (6). LP_{saa} involves minimizing a linear objective over a convex polytope, and so step 2 of saaRound can be done efficiently to compute the fractional solutions x, y. Also note that since LP_{saa} is optimizing over a larger space (specifically, the convex hull of all the feasible integral solutions), the objective value in (1) might be smaller than the integral objective value.
- **Rounding to an integral solution**: If the solution computed by LP_{saa} is integral, we are done (Step 3(1)). However, in general solution x is fractional, which poses a problem: if we have $x_{u0} \in (0, 1)$, e.g., a fractional value of 0.2, it is not clear how to construct a valid integral solution. In Step 3(2) of saaRound, we pick all the nodes with $y_{uj} \leq 1/2$ (denoted by $Y_{uj} = 0$), and pick a set of nodes to vaccinate (Step 3(2)), such that every node u with $Y_{uj} = 0$ gets disconnected from $\texttt{src}(H_j)$. Step 3(3) achieves this by rounding the fractional solution x, after appropriate scaling. This randomized rounding step ensures that the budgets are not violated by much. This also implies that any node u which gets infected in H_j has $y_{uj} \geq 1/2$, so that the average number of infections can be bounded by at most twice the fractional objective value.

3.3 Analysis of saaRound

For a sample H_j computed in Step 1 of saaRound, let $f(H_j(\mathbf{X}_0)) =< U(0), \ldots, U(\tau), E'_j >$ be defined in the following manner: (1) E'_j is the subset of E_j when nodes in $\mathbf{X}_0$ are removed, (2) $U(0) = \texttt{src}(H_j) - \mathbf{X}_0$, and (3) for all $t > 0$, $U(t)$ is the set of nodes at distance t in the subgraph induced by E'_j. We first observe that the sampling process is "equivalent" to the SIR process.

Observation 1. *For a given outcome* $O =< U(0), \ldots, U(\tau), E'_j >$, $\Pr[H^{(sir)}(\mathbf{X}_0) = O] = \Pr[f(H_j(\mathbf{X}_0)) = O]$.

For a vaccination set $\mathbf{X}$, let $Z_j(\mathbf{X})$ be the number of nodes in $H_j - \mathbf{X}$, which are still reachable from $\texttt{src}(H_j)$; note that this includes the sources themselves. From Observation 1, it follows that $Z_j(\mathbf{X})$ is equal to the number of infections in the stochastic outcome $H^{(sir)}(\mathbf{X})$ of the SIR process. Let $Z(\mathbf{X}) = \frac{1}{M}\sum_j Z_j(\mathbf{X})$, and let $\hat{X}_{opt} = \text{argmin}_{X'} Z(X')$ be the solution that achieves the minimum average number of infections in the samples. Finally, let $X_{opt} = \text{argmin}_{X'} \texttt{EInf}(X')$ be the optimal solution to the 1sEpiControl instance. The following lemma shows that the average number of infections achieved by a solution $\mathbf{X}$ restricted to the samples $H_1, \ldots, H_M$ is close to the $\texttt{EInf}$ objective.

Lemma 3.1. *Let* $Z(\cdot)$ *be as defined above. If* $M \geq 24n^2 \log n$*, with probability at least* $1 - 1/n$*, for every intervention set* $\mathbf{X}$*, we have* $Z(\mathbf{X}) \in [\frac{1}{2}\texttt{EInf}(\mathbf{X}), \frac{3}{2}\texttt{EInf}(\mathbf{X})]$.

Proof. From Observation 1, we have $E[Z(\mathbf{X})] = E[Z_j(\mathbf{X})] = \texttt{EInf}(\mathbf{X})$ for all j. The $Z_j(\mathbf{X})$ variables are independent, and $\frac{Z_j(\mathbf{X})}{n} \in [0,1]$. This implies the Chernoff bound (Theorem 2.1) can be applied to $M\frac{Z(\mathbf{X})}{n} = \sum_j \frac{Z_j(\mathbf{X})}{n}$, so that

$$\Pr\left[\frac{MZ(\mathbf{X})}{n} \notin \left[\frac{M\texttt{EInf}(\mathbf{X})}{2n}, \frac{3M\texttt{EInf}(\mathbf{X})}{2n}\right]\right] \leq 2exp(-\frac{M}{12n}\texttt{EInf}(\mathbf{X})).$$

We have $\texttt{EInf}(\mathbf{X}) \geq 1$, since there is always at least one infection. For $M = 24n^2 \log n$, this probability is at most $2e^{-2n\log n} = \frac{2}{n^n n^n}$. The number of possible intervention sets is the number of possible sets $\mathbf{X} \subseteq V$, which is at most 2^n. Therefore, the probability that there exists an intervention set $\mathbf{X}$ for which $Z(\mathbf{X}) \notin [\frac{1}{2}\texttt{EInf}(\mathbf{X}), \frac{3}{2}\texttt{EInf}(\mathbf{X})]$ is at most $2^n \cdot \frac{2}{n^n n^n} \leq \frac{1}{n}$ for $n > 1$. □

Recall that IP_{saa} is the integral version of LP_{saa}, obtained by requiring all the variables to be integral, instead of constraints (6). We first show that IP_{saa} is "valid".

Lemma 3.2. *For every feasible intervention set* $\mathbf{X}$*, there exists a feasible integral solution* $\bar{x}, \bar{y}$ *to* IP_{saa}*, such that* $\frac{1}{M}\sum_j \sum_v \bar{y}_{v,j} = Z(\{v : \bar{x}_v = 1\})$*. If* $\bar{x}, \bar{y}$ *is an optimal solution to* IP_{saa}*,* $Z(\hat{X}_{opt}) = \frac{1}{M}\sum_j \sum_v \bar{y}_{v,j}$

Proof. (Sketch) First, consider a feasible intervention $\mathbf{X}$. We define $\bar{x}_v = 1$ for all $v \in \mathbf{X}$. We define $\bar{y}$ in the following manner. Let $f(H_j(\mathbf{X})) =< U_j(0), \ldots, U_j(\tau_j), E'_j >$, as defined earlier; we have $Z_j(\mathbf{X}) = \sum_t |U(t)|$. We define $y_{v,j} = 1$ if $v \in \cup_t U_j(t)$. We show that $\bar{x}, \bar{y}$ is a feasible solution to IP_{saa}. For any j, consider a node $u \in U_j(t)$ for some t. Then, there exists a path $P = u_0, u_1, \ldots, u_t = u$ with $u_i \in U_j(i)$ for $i \leq t$. By construction, for each node u_i, we have $y_{u_i,j} = 1 \geq y_{wj} - z_{u_i 0}$ for every neighbor w of u, which implies the constraint (3) is satisfied for u, and each of its neighbor w. Let $U = \cup_{t=0}^{\tau_j} U_j(t)$. Consider a node $u \notin U$. If u has a neighbor $w \in U$, it must be the case that $u \in \mathbf{X}$, else node u would be infected at time $\tau_j + 1$, and would have been in a set $U(\tau_j + 1)$. This implies $x_{u0} = 1$, and the constraint (3) holds for node u and any neighbor w. If u has no neighbor $w \in U$, then $y_{wj} = 0$, and so the constraint (3) holds for u, w. The converse follows similarly. We need the following additional property: if $y_{uj} = 1$, there is a path P from $\texttt{src}(H_j)$ with $y_{u_i j} = 1$ for all nodes $u_i \in P$; this holds due to the min objective. □

Lemma 3.3. *For any* H_j*, and any node* $v \in V$ *with* $y_{vj} < \frac{1}{2}$*,* $\Pr[v$ *is reachable from* $\texttt{src}(H_j)$ *in* $H_j[V - \mathbf{X}]] < \frac{1}{4nM}$*, where* $H_j[V - \mathbf{X}]$ *is the graph induced by removing the nodes in* $\mathbf{X}$ *from* H_j.

Proof. Let $\mathcal{P}_{vj} = \{P_1, \ldots, P_L\}$ be the set of paths to node v in H_j. For a path P, let $S(P) = \{u : u \in P\}$ be set of nodes on the path P. Node $v \in V$ is reachable from $\texttt{src}(H_j)$ in $H_j[V - \mathbf{X}]$ if and only if there exists some path $P \in \mathcal{P}_{vj}$ such that none of the nodes in $S(P)$ are vaccinated (i.e., $X_{u0} = 0,\ \forall u \in S(P)$). If there exists $u \in S(P)$ with $2x_{u0}\log(4nMN) \geq 1$, the rounding ensures that $X_{u0} = 1$; therefore, we only consider the case $2x_{u0}\log(4nMN) \leq 1$. Our rounding ensures that we have $\Pr[X_{u0} = 1] \geq 2x_{u0}\log(4nMN)$, so that $\Pr[\sum_{u\in S(P)} X_{u0} = 0]$ is upper bounded by $\prod_{u\in S(P)}\left(1 - 2x_{u0}\log(4nMN)\right) \leq e^{-\sum_{u\in S(P)} 2x_{u0}\log(4nMN)} \leq e^{-\log(4nMN)} = \frac{1}{4nMN}$, since $\sum_{u\in S(P)} x_{u0} \geq 1 - y_{vj} \geq 1/2$. Equivalently, the probability that no node from $S(P)$ is picked is at most $\frac{1}{4nMN}$; here we say a node is picked from $S(P)$ if $X_{u0} = 1$ for some $u \in S(P)$. By a union bound, the probability that there exists a path $P \in \mathcal{P}_{vj}$ such that no node from $S(P)$ is picked is at most $\frac{L}{4nMN} \leq \frac{1}{4nM}$ (since, $L \leq N$). Therefore, the lemma follows. □

Lemma 3.4. *With probability at least* $1 - 1/n$*, we have* $|X_0| \leq 12\log(4nMN)B_0$.

Proof. The expected number of nodes picked for vaccination is given by $\mu = E\left[\sum_u X_{u0}\right] \leq \sum_u 2x_{u0}\log(4nMN) \leq 2\log(4nMN)B_0$. The first inequality is by linearity of expectation and the second inequality follows from the constraint (5) of LP_{saa}. The X_{u0}'s are all rounded independently, we have

$$\begin{aligned}\Pr\left[\sum_u X_{u0} > 12\log(4nMN)B_0\right] &\leq \Pr\left[\sum_u X_{u0} \geq 6\mu\right] \\ &\leq exp(-6\log(4nMN)B_0) \\ &\leq \frac{1}{n}.\end{aligned}$$

The first inequality follows from the bound on μ. The second inequality follows from the Chernoff bound (Theorem 2.1), as $6\mu \geq 2e\mu$. The last inequality follows as $6\log(4nMN)B_0 \geq \log n$. □

Theorem 3.5. *Let* X_0 *denote the vaccination set computed by algorithm* saaRound*. Then, with probability at least* $1/2$*, we have* $\texttt{EInf}(X_0) \leq 6\texttt{EInf}(X_{opt})$*, and* $|X_0| \leq 12\log(4nMN)B_0$.

Proof. (Sketch) Let $\hat{X}_{opt}$ be as defined above. By Lemma 3.3, for any v, j, if $y_{vj} \leq 1/2$, the probability that node v is reachable from $\texttt{src}(H_j)$ is at most $\frac{1}{4nM}$. By a union bound, the probability that this holds for at least one vertex $v \in V$ (for a fixed j) is at most $\frac{1}{4M}$. This implies that with probability at least $1 - \frac{1}{4M}$,

$$Z_j(X_0) \leq |\{v : y_{vj} \geq 1/2\}| \leq \sum_{v: y_{vj}\geq 1/2} 2y_{vj} \leq \sum_v 2y_{vj}$$

By a union bound, with probability at least $1 - \frac{M}{4M} = 1 - \frac{1}{4}$, we have $Z_j(X_0) \leq 2\sum_v y_{vj}$, for all j. By definition of $\hat{X}_{opt}$, we have $\frac{1}{M}\sum_j Z_j(X_0) \leq \frac{1}{M}\sum_{v,j} 2y_{vj} \leq 2Z(\hat{X}_{opt})$, since the LP solution is also a lower bound on $Z(\hat{X}_{opt})$. By Lemma 3.4, the condition $|X_0| \leq 12\log(4nMN)B_0$ holds, in addition to $Z(X_0) \leq 2Z(\hat{X}_{opt}) \leq$

$2Z(X_{opt})$, with probability at least $1 - \frac{1}{4} - \frac{1}{n}$, since $Z(\hat{X}_{opt}) \leq Z(X_{opt})$, by definition of $\hat{X}_{opt}$.

By Lemma 3.1, with probability at least $1-\frac{1}{n}$, we have $Z(X_{opt}) \leq \frac{3}{2}\texttt{EInf}(X_{opt})$, and $\frac{1}{2}\texttt{EInf}(X_0) \leq Z(X_0)$. This gives us

$$\texttt{EInf}(X_0) \leq 2Z(X_0) \leq 4Z(X_{opt}) \leq 6\texttt{EInf}(X_{opt})$$

Therefore, all the conditions of the theorem hold with probability $\geq 1 - \frac{1}{4} - \frac{2}{n} \geq \frac{1}{2}$. □

3.4 Extension to the 2sEpiControl problem

We discuss the changes to be made to LP_{saa} to adapt it for the two-stage version (this can be similarly extended to multiple stages). We have B_0, B_T as inputs. After step (1) of saaRound, run breadth first search (BFS) in each H_j from the nodes in $\texttt{src}(H_j)$. Let $V_{j,t}$ denote the set of all nodes at level t in the BFS tree in H_j (with the nodes in $\texttt{src}(H_j)$ at level 0); let $V_{j,\geq t} = \cup_{t' \geq t} V_{j,t}$ denote the set of all nodes at level t or more.

Constraint (2) is modified in the following manner: for all nodes u in the set $V_{j,\geq T} - \texttt{src}(H_j)$ in each sample H_j, we have

$$\forall j, u \in V_{j,\geq T} - \texttt{src}(H_j), \forall t: \; y_{uj} \leq 1 - x_{ut}.$$

The Constraint (3) is changed to

$$\forall j, \forall u \in V, \; (w,u) \in E_j: \; y_{uj} \geq y_{wj} - \sum_{t: u \in V_{j,\geq t}} x_{ut}.$$

We add the constraint

$$\sum_u x_{uT} \leq B_T.$$

We refer to this linear program as LP^e_{saa} The rounding procedure is same for the x and y variables as in saaRound. The algorithm returns X_0, X_T as the solution.

Analysis: If the sampled subgraphs are trees (which is typical for low transmission probability), LP^e_{saa} is valid, and we can show the same guarantees as Theorem 3.5. In general, however, LP^e_{saa} may not be valid, and the solution might not have these guarantees, due to the following reason: suppose there is a node u which is at level $< T$ in a sampled subgraph H_j before the first stage of intervention is done at time 0. After a set X_0 is picked (and removed from the graph), the distance of u from $\texttt{src}(H_j)$ might increase, and it could be vaccinated at time T. However, our algorithm will not pick such nodes, and thus optimizes over a smaller decision space.

3.5 Improving performance and speeding up saaRound

Improved approximation factor. The worst case approximation is most impacted by the scaling we do in step 4(3) of saaRound, which is needed for the application of the Chernoff bound in Lemma 3.4. However, as we discuss later, we find that LP_{saa} computes near-integral solutions, in which most (and sometimes all) variables are integral. Step 4(1) handles integral variables separately. We also modify Step 4(3) by using a smaller scaling factor, depending on the fractional value.

Better scaling. The main bottleneck in saaRound is the solution of LP_{saa}, which has: nM variables of the form y_{uj}, $n|\mathcal{T}|$ variables of the form x_{ut}, and $\sum_j |E(H_j)|$ constraints (3). The worst case dependence of the running time of LP solvers is super-quadratic in these parameters (though we find the Gurobi solver [16] scales very well in practice, as we discuss later). In order to improve the scaling of saaRound to larger instances, we use the following methods.

- Reduced number of samples: the rigorous bound on the number of samples needed in the worst case comes from Lemma 3.1, as a result of the Chernoff bound. In practice, we find that there is concentration even with $O(\sqrt{n})$ samples, and so we use fewer samples in our experiments. This can be estimated in a statistically rigorous manner by picking the smallest number of samples such that the variance is within a factor δ.
- Reducing the number of variables: we define the *vulnerability* of node u, denoted by y_u, as probability that it gets infected (when no interventions are done). This can be estimated as the fraction of samples H_j in which u is connected to $\texttt{src}(H_j)$, i.e., $y_u = \frac{1}{M}|\{j : u \text{ is connected to } \texttt{src}(H_j)\}$. For a parameter γ, we restrict the interventions to nodes in $V_\gamma = \{v : y_v > 1-\gamma\}$; in other words, we can set $x_{vt} = 0$ for nodes with vulnerability at most γ. The intuition is that such nodes are likely to have low x_{vt} values in LP_{saa}, and so it is safe to remove them and reduce the size of the LP. This is borne out from our experiments.

4 EXPERIMENTS

We study the following questions:

- **Scaling**: how well does saaRound scale to large networks? How effective are the techniques for choosing the number of samples and pruning?
- **Approximation Guarantees**: what is the approximation factor of saaRound in practice? How does it compare with the other baselines?
- **Effect of multiple stages**: how does the effectiveness of the solution vary with the number of stages and the budget in each?
- **Characteristics of the solutions**: what kinds of nodes are picked in the solutions at each stage?

4.1 Dataset and Methods

Datasets. We experiment with three different classes of networks (total of six), in order to fully explore the effect of network structure on the the results. We consider two random networks, namely the small world [19], and the preferential attachment [5] models. The parameters used in generation of the random networks is presented in the full version. [2]. We study the results on the CA-GrQc collaboration network [20], since it is a type of social network. We also consider synthetic agent based populations for Montgomery County, VA, and Portland, OR, constructed by a first principles approach by [6, 11]. This has been used in several public health studies, e.g., [32]. This network has a rich set of demographic attributes for each node, e.g., age, gender, and income. The datasets are summarized in Table 2.

Choosing parameters. There is a large space of model parameters over which the analysis could be done. Due to the space limits, and in order to get the most insights, we choose a subset as described here. We choose the source distribution **s** so that about 10 initial infections are picked. Following standard practice in public health, e.g., [17], we choose three values for the transmission probability p based on the expected number of infections (referred to as the "attack rate") that result when there are no interventions: we choose

a probability p_{low} if the attack rate is $< 10\%$ (low), p_{med} if the attack rate is in $[10\%, 20\%]$ (medium), and p_{high} if the attack rate is $> 20\%$. The specific probability values depend on the datasets. The full version of the paper [2] shows how the attack rate varies with the probability, and the specific probability values which were chosen.

Dataset	Nodes	Edges
Montgomery	70729	198138
Portland	1409197	8307767
CA-GrQc	5242	14496
Small World (SW)	2500	14833
Preferential1 (PA1)	1000	1996
Preferential2 (PA2)	100000	199996

Table 2: Description of datasets

Methods. We only focus on one stage (1sEpiControl) or two stage (2sEpiControl) versions of EpiControl here, and use saaRound to find interventions. We use the Gurobi solver [16] to implement saaRound. For 1sEpiControl, we consider the following baselines, which select B nodes based on two criteria:

- Nodes with the highest degree (top-B degree), which has been a popular approach in a number of papers [5, 31]
- Nodes with the highest eigenvector centrality (top-B EVC).

No prior results are known for 2sEpiControl. Therefore, we adapt the above baselines and pick B_0 and B_T nodes in the order of the above scores. The top-B EVC does not give insights on the performance of the spectral approaches[27–29, 35, 38–40]. A more detailed comparison of our approach with the spectral methods is an important future direction.

4.2 Structure of solutions

Very surprisingly, we find that the LP gives solutions with a lot of integral or half-integral variables, i.e., with values in $\{0, 1/2, 1\}$. In fact, in 25 out of 40 instances, the LP solution was optimal! Understanding the specific problem structure leading to this property is an interesting open problem.

4.3 Scaling

We find that saaRound easily scales to all the networks we consider. The two strategies for speeding up have a significant impact on the scaling.

- **Number of samples needed**: We find the number of samples sufficient to get reasonable variance, as shown in Figure 2, to be less than the worst case bound of $\Theta(n \log n)$ from Lemma 3.1. We observe that the number of samples needed for convergence in some cases to be $O(\sqrt{n})$. The number of samples needed is lower when the transmission probability is medium or high, and when the budget is not too high, since this has better convergence. We note that these are typically the regimes of maximum concern in public health.
- **Impact of pruning**: The pruning of low vulnerability nodes has a very significant impact on the running time, as shown in Figure 3, which shows the running time with and without pruning. When the number of samples used is low, the difference is negligible, but when the number of samples increases to the range needed for low variance, we find the difference in running times is in several orders of magnitude. The objective value differs by less than 5% with and without pruning for PA1 network. Similar trend is observed for Portland network. This can be seen in Figure 4. This implies that our scaling strategies give good solutions on very large networks.

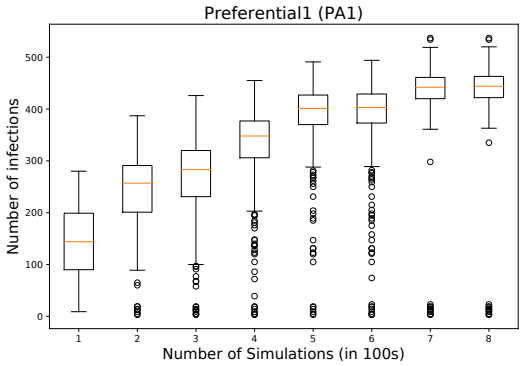

Figure 2: Number of samples needed.

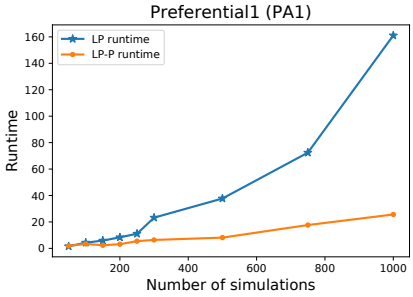

Figure 3: Comparison of runtimes of Linear Programs with (LP-P) and without pruning (LP).

4.4 Performance Guarantees and comparison with baselines

4.4.1 Comparison to baselines. Figures 5 (a-c) show the objective value for varying budgets for the 1sEpiControl problem. The baselines use exactly the same budget as the solution to saaRound, for fair comparison. We observe that saaRound significantly outperforms both the baselines. For social contact networks (Fig 5) (c), which are relatively dense, the objective value from the eigenscore and degree baselines are over seven and three times that from saaRound, respectively, over the entire budget range. saaRound outperforms eigenscore by a similar factor in Figure 5 (b) as well.

4.4.2 Approximation Ratio. We observe that the approximation ratio with respect to the objective value is close to 1. Figures 5 shows that for all the three networks, the objective value of LP optimal solution almost coincides with that of saaRound. Also,

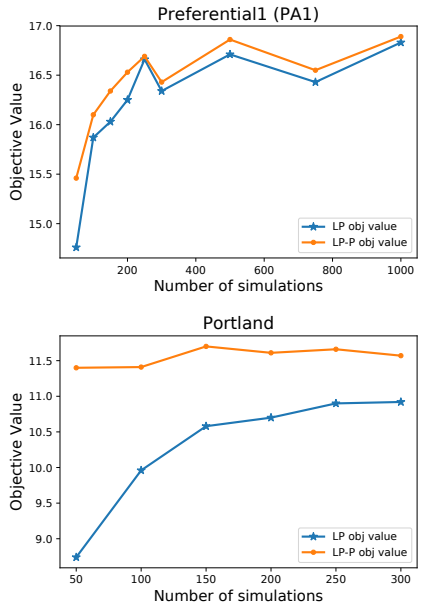

Figure 4: Comparison of objective values of Linear Programs with (LP-P) and without pruning (LP).

the approximation ratio with respect to budget violation (the ratio of number of interventions after rounding to the budget given as input — original budget) is close to 1, in practice, shown in Figure 6.

4.5 Two stage intervention and Structure of solution

We study the 2sEpiControl. In Figure 7, we examine how the objective value `EInf` increases with T in a two stage intervention, where T is the time of the second stage, while the first intervention is performed at time step 0. We observe that `EInf` increases very rapidly with T.

Next, we examine the structure of the sets picked in each stage. Figure 8 shows a scatter plot of the node degree and age of the solution to 2sEpiControl with $T = 4$. We observe that there are slight differences between the sets X_0 and X_4: X_0 has slightly higher degree nodes, whereas X_4 has slightly lower age nodes. But more importantly, *it is not the case that all high degree nodes are used in* X_0.

5 RELATED WORK

5.1 Public health policy planning and use of diffusion models

Public health policy analysis relies heavily on mathematical models of SIR type processes, e.g., [1, 21]. As discussed earlier, there are two broad classes of models. The first involves using a system of coupled differential equations to represent the dynamics, e.g., [23, 36, 38]. These do not have any closed form solutions, in general, and when the system is not very large, it can be solved by brute force local search methods [23]. For some types of models, greedy strategies have been used [36, 38]. The second is network based, and uses a stochastic diffusion model for the spread of the disease [11, 15, 17, 21, 22]. Such models have been found to be more powerful and useful for epidemic spread on large heterogeneous populations, where the complete mixing assumptions of differential equation models are not valid.

During any large outbreak, public health agencies solve a variety of models, and make plans and guidelines based on the results, e.g., during the 2009 swine flu [23], and the 2014 Ebola outbreak [21]. Such studies typically explore the space of different possible interventions, within given resource constraints. Therefore, there is a lot of interest in the design of optimal or near-optimal interventions.

5.2 Resource optimization to control epidemic spread

There is a lot of work on this topic, and we summarize the main research directions relevant to EpiControl.

Firefighter problems, e.g., [3, 7, 12]. This is most closely related to EpiControl. The basic version of the problem is to determine a temporal intervention $\mathbf{X} = \{X_t : t = 1, \ldots, n\}$, such that $|\mathbf{X}_t| \leq B$, and the number of nodes not infected (i.e., saved) is maximized (this is referred to as the Max-Save version). The disease model is a SI process with $p = 1$ (so this is a deterministic model). The Max-Save version cannot be approximated within an $\Omega(n^{\epsilon})$ factor for any $\epsilon < 1$, in general graphs. A related problem is the Min-Budget version, for which an $O(\sqrt{n})$ approximation is possible [7]. This work corresponds to *non-spreading* interventions. Better approximations are possible for the setting in which the vaccination is also a *spreading* process. A special case of this problem is with work of [4], which considers EpiControl but with the intervention specified at time 0, and $p = 1$, and gave the first rigorous approximation guarantee for this problem. We refer to Finbow et al. [12] for an extensive survey on the firefighter problems. One of the few works on the firefighter problem with a stochastic disease model (i.e., $p < 1$) is by Tennenholtz et al. [34], who formalize the problem as a Markov Decision Process (MDP), and compute an optimal solution for trees. One of the main differences between this and our paper is that the MDP formulation of [34] makes the problem *adaptive*, i.e., it is possible to use information about which nodes are infected before time t, in designing the intervention $\mathbf{X}_t$. Drakopoulos et al. [9] consider the problem of designing dynamic policies to contain contagion spread based on SIS epidemic model. But the results in this paper consider graphs with bounded degree and bounded CutWidth. Also, their dynamic policy has a bottleneck — computing CutWidth, which is NP-Complete. In contrast to both these works, EpiControl only considers a non-adaptive setting, and the intervention has to be determined ahead of time.

Optimization of spectral properties. A key result in epidemic modeling is a characterization of an outbreak in terms of spectral properties, namely the first eigenvalue of the adjacency matrix (also referred to as the spectral radius, and denoted by λ_1), and the eigenvalues of the Laplacian [14, 26, 37]. An important implication of this is that the epidemic dies out if λ_1 is reduced, and this has formed the basis of a lot of work on epidemic control, e.g., [25, 27–30, 39, 40]. However, this does not lead to direct bounds on the `EInf` objective. The top-B EVC baseline doesn't optimize the

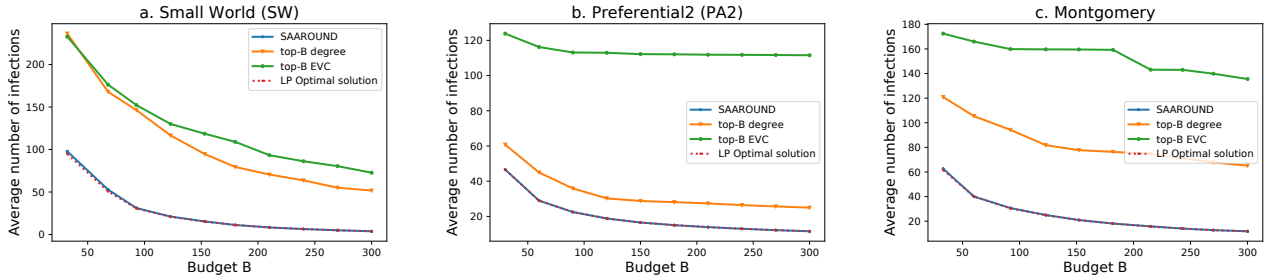

Figure 5: Objective value (y-axis) vs budget (x-axis) for SAAROUND, and the degree and eigenscore baselines for four networks.

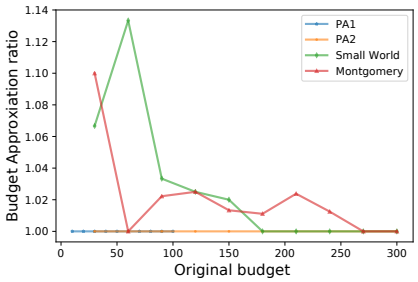

Figure 6: Budget Violation

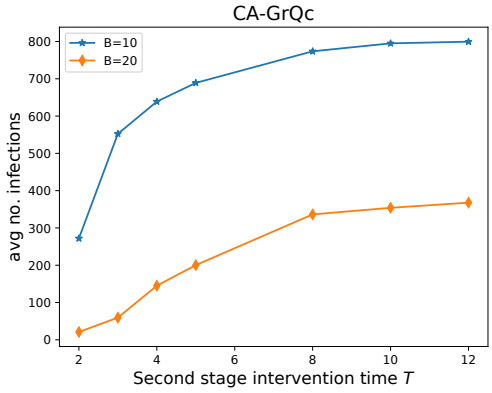

Figure 7: Two-stage Intervention.

spectrum directly, and so doesn't provide insights on the empirical performance of spectral methods on the `EInf` objective. A thorough comparison of our approach with the spectral methods will be an interesting future work.

Heuristics based on centrality and local structure. Finally, heuristics have been proposed based on local structure, e.g., degree, and centrality, e.g., [5, 8, 24]. These do reasonably well in practice, and are especially useful when the graph structure is not known well. However, they do not lead to any rigorous bounds on the `EInf` objective. Further, as our experimental results show, the degree based heuristic works better than EVC, but can still be significantly worse than SAAROUND in some networks.

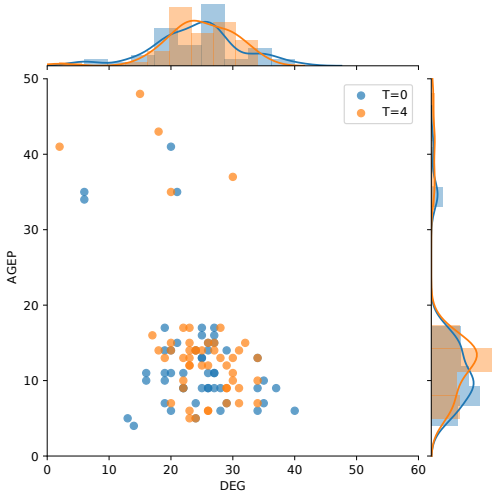

Figure 8: Montgomery Graph: scatter plot of age and degree of nodes of the sets X_0 and X_4 in a solution to the 2-stage EPICONTROL with budgets $B_0 = B_4 = 25$.

6 CONCLUSIONS

Our results show that linear programming based rounding and the sample average approximation technique are quite effective in giving solutions with good approximation guarantees in practice. Our pruning method allows scaling to pretty large networks. Our results are the first to examine multi-stage interventions, and we find that the temporal dimension leads to significant changes in the solution quality and structure. Improving the approximation guarantees by better rounding techniques is an important open problem. Our methods can help in public health policy planning and response to large outbreaks.

7 ACKNOWLEDGMENTS

This work has been partially supported by the following grants: NSF CRISP 2.0 1832587, DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), NSF DIBBS ACI-1443054, NSF EAGER CMMI-1745207, NSF BIG DATA IIS-1633028, NSF CAREER IIS-1750407, NEH HG229283-15, ORNL and a Facebook faculty gift.

REFERENCES

[1] R.M. Anderson and R.M. May. 1991. *Infectious Diseases of Humans.* Oxford University Press, Oxford.

[2] Anonymous. 2019. Designing Optimal Temporal Interventions for Epidemics. https://tinyurl.com/yy7wztzs. In *Technical Report.*

[3] Elliot Anshelevich, Deeparnab Chakrabarty, Ameya Hate, and Chaitanya Swamy. 2009. Approximation Algorithms for the Firefighter Problem: Cuts over Time and Submodularity. In *Algorithms and Computation*, Yingfei Dong, Ding-Zhu Du, and Oscar Ibarra (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 974–983.

[4] James Aspnes, Kevin Chang, and Aleksandr Yampolskiy. 2005. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '05).* 43–52.

[5] Albert-László Barabási and Réka Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286, 5439 (1999), 509–512. https://doi.org/10.1126/science.286.5439.509 arXiv:https://science.sciencemag.org/content/286/5439/509.full.pdf

[6] Christopher L Barrett, Richard J Beckman, Maleq Khan, V. S. Anil Kumar, Madhav V Marathe, Paula E Stretz, Tridib Dutta, and Bryan Lewis. 2009. Generation and analysis of large synthetic social contact networks. In *Winter Simulation Conference.* Winter Simulation Conference, 1003–1014.

[7] Parinya Chalermsook and Julia Chuzhoy. 2010. Resource Minimization for Fire Containment. In *Proceedings of the Twenty-first Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '10).* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1334–1349. http://dl.acm.org/citation.cfm?id=1873601.1873709

[8] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. 2003. Efficient Immunization Strategies for Computer Networks and Populations. *Phys. Rev. Lett.* 91 (Dec 2003), 247901. Issue 24. https://doi.org/10.1103/PhysRevLett.91.247901

[9] Kimon Drakopoulos, Asuman Ozdaglar, and John N Tsitsiklis. 2014. An efficient curing policy for epidemics on graphs. *IEEE Transactions on Network Science and Engineering* 1, 2 (2014), 67–75.

[10] Devdatt P. Dubhashi and Alessandro Panconesi. 2009. *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press. I–XIV, 1–196 pages.

[11] S. Eubank, H. Guclu, V. S. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429 (2004), 180–184. Issue 6988.

[12] Stephen Finbow and Gary MacGillivray. 2009. The Firefighter Problem: a survey of results, directions and questions. *Australasian J. Combinatorics* 43 (2009), 57–78.

[13] Centers for Disease Control (CDC). [n. d.]. Seasonal Influenza Vaccine Supply & Distribution. ([n. d.]). https://www.cdc.gov/flu/prevent/vaccine-supply-distribution.htm.

[14] A. Ganesh, L. Massoulie, and D. Towsley. 2005. The effect of network topology on the spread of epidemics. *Proceedings of INFOCOM* (2005).

[15] T.C. Germann, K. Kadau, I.M. Longini Jr, and C.A. Macken. 2006. Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences* 103, 15 (2006), 5935–5940.

[16] Gurobi 2017. Reference Manual for Gurobi Optimizer. Gurobi.http://www.gurobi.com/.. (2017).

[17] M. Elizabeth Halloran, Neil M. Ferguson, Stephen Eubank, Ira M. Longini, Derek A. T. Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy C. Germann, Diane Wagener, Richard Beckman, Kai Kadau, Chris Barrett, Catherine A. Macken, Donald S. Burke, and Philip Cooley. 2008. Modeling targeted layered containment of an influenza pandemic in the United States. In *Proceedings of the National Academy of Sciences (PNAS).* 4639–4644.

[18] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proc. 9th KDD.* 137–146.

[19] Jon Kleinberg. 2000. The Small-World Phenomenon: An Algorithmic Perspective. In *in Proceedings of the 32nd ACM Symposium on Theory of Computing.* 163–170.

[20] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2007. Graph Evolution: Densification and Shrinking Diameters. *ACM Trans. Knowl. Discov. Data* 1, 1, Article 2 (March 2007). https://doi.org/10.1145/1217299.1217301

[21] Eric Lofgren, M. Elizabeth Halloran, C. M. Rivers, J. M. Drake, Travis C. Porco, Bryan Lewis, Wan Yang, Alessandro Vespignani, Jeffrey Shaman, Joseph N. S. Eisenberg, M. C. Eisenberg, Madhav Marathe, Samuel V. Scarpino, Kathleen A. Alexander, Rafael Meza, Matthew J. Ferrari, James M. Hyman, Lauren A. Meyers, and Stephen Eubank. 2014. Opinion: Mathematical models: A key tool for outbreak response. *PNAS* (2014), 18095–18096.

[22] M. Marathe and A. Vullikanti. 2013. Computational Epidemiology. *Commun. ACM* 56, 7 (2013), 88–96.

[23] J. Medlock and A. P. Galvani. 2009. Optimizing influenza vaccine distribution. *Science* 325, 5948 (2009), 1705–1708.

[24] Joel C. Miller and James Mac Hyman. 2007. Effective vaccination strategies for realistic social networks. *Physica A: Statistical Mechanics and its Applications*, 780–785.

[25] Masaki Ogura and Victor M. Preciado. 2017. *Optimal Containment of Epidemics in Temporal and Adaptive Networks.* Springer Singapore, Singapore, 241–266.

[26] B. Aditya Prakash, Deepayan Chakrabarti, Michalis Faloutsos, Nicholas Valler, and Christos Faloutsos. 2011. Threshold Conditions for Arbitrary Cascade Models on Arbitrary Networks. In *ICDM.*

[27] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. 2013. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks.. In *IEEE Conference on Decision and Control.* IEEE.

[28] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. 2014. Optimal Resource Allocation for Network Protection Against Spreading Processes.. In *IEEE Transactions on Control of Network Systems.* IEEE, 99 – 108.

[29] Victor M. Preciado, Michael Zargham, and David Sun. 2014. A convex framework to control spreading processes in directed networks.. In *Annual Conference on Information Sciences and Systems (CISS).* IEEE.

[30] Sudip Saha, Abhijin Adiga, B. Aditya Prakash, and Anil Vullikanti. 2015. Approximation Algorithms for Reducing the Spectral Radius to Control Epidemic Spread. In *SIAM SDM.*

[31] Marcel Salathe and James H. Jones. 2012. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Computational Biology* (2012).

[32] Meghendra Singh, Prasenjit Sarkhel, Gloria J. Kang, Achla Marathe, Kevin Boyle, Pamela Murray-Tuite, Kaja M. Abbas, and Samarth Swarup. 2019. Impact of demographic disparities in social distancing and vaccination on influenza epidemics in urban and rural regions of the United States. *BMC Infectious Diseases* 19 (12 2019). https://doi.org/10.1186/s12879-019-3703-2

[33] Chaitanya Swamy and David B. Shmoys. 2006. Approximation Algorithms for 2-Stage Stochastic Optimization Problems. *SIGACT News* 37, 1 (March 2006), 33–46. https://doi.org/10.1145/1122480.1122493

[34] Guy Tennenholtz, Constantine Caramanis, and Shie Mannor. 2017. The Stochastic Firefighter Problem. *CoRR* abs/1711.08237 (2017). arXiv:1711.08237 http://arxiv.org/abs/1711.08237

[35] H. Tong, B. Aditya Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. 2012. Gelling, and Melting, Large Graphs by Edge Manipulation. In *CIKM.*

[36] S. Venkatramanan, J. Chen, S. Gupta, B. Lewis, M. Marathe, H. Mortveit, and A. Vullikanti. 2017. Spatio-Temporal Optimization of Seasonal Vaccination Using a Metapopulation Model of Influenza. In *2017 IEEE International Conference on Healthcare Informatics (ICHI).* 134–143.

[37] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. 2003. Epidemic spreading in real networks: An eigenvalue viewpoint. *Proceedings of SRDS* (2003).

[38] Bryan Wilder, Sze-Chuan Suen, and Milind Tambe. 2018. Preventing Infectious Disease in Dynamic Populations Under Uncertainty. (2018). https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16714/15766

[39] Yao Zhang, Abhijin Adiga, Anil Vullikanti, and B Aditya Prakash. 2015. Controlling propagation at group scale on networks. In *Data Mining (ICDM), 2015 IEEE International Conference on.* IEEE, 619–628.

[40] Yao Zhang and B. Aditya Prakash. 2014. DAVA: Distributing Vaccines over Networks under Prior Information. In *Proceedings of the SIAM Data Mining Conference (SDM '14).*