Preventing Epidemic Spread Using Leaky Vaccines

Da Qi Chen
Biocomplexity Institute and Initiative
University of Virginia
Charlottesville, United States

Ann Li
Department of Computer Science
University of Virginia
Charlottesville, United States

George Li
School of Computer Science
Carnegie Mellon University
Pittsburgh, United States

Madhav Marathe
Biocomplexity Institute and Initiative
University of Virginia
Charlottesville, United States

Aravind Srinivasan
Department of Computer Science
University of Maryland
College Park, United States

Anil Vullikanti
Biocomplexity Institute and Initiative
University of Virginia
Charlottesville, United States

Jiayi Wu
Department of Computer Science
University of Maryland
College Park, United States

***Abstract*—Vaccines have proven to be extremely effective in preventing the spread of COVID-19 and potentially ending the pandemic. However, it is impossible to guarantee 100% effectiveness of a vaccine. This is not considered in previous models, which may have a significant impact on the distribution, implementation, and outcome of any preventive measures. In this paper, we study the problem of choosing a set of individuals to vaccinate given a contact network in order to minimize the expected number of infections given that vaccines are not fully effective, by sample averaging and linear programming techniques to achieve approximate solutions. We also consider this model for other variations of the problem, minimizing average and maximum degree of a network. These objectives are often used as proxies for minimizing the total infection in a network, and we show experimentally that the objectives indeed correlate well with the expected number of infections. In addition, we present a greedy approximation algorithm for the two variations of the problem. Lastly, we run experiments on a contact network for the population of Montgomery County, VA in order to investigate the effectiveness and robustness of our algorithms.**

***Index Terms*—algorithmic epidemiology, graph theory.**

I. Introduction

Immunization is a very important public health intervention for many diseases [1]–[3], and numerous studies have shown that immunization has played a very significant role in controlling the spread of the COVID-19 pandemic, e.g., [4]–[6], despite challenges in deployments and waning immunity, moreover, significant work has been done in epidemiology regarding "leaky vaccines" [7]. In the initial stages of the outbreak, or when the availability of vaccines is limited, an important question from a public health perspective is: how should the vaccines be distributed, so that their effectiveness is maximized? Analysis on mathematical and computational models has shown that optimizing vaccination strategies can have significant impact on the outbreak size [1], [8], [9].

Epidemic spread is commonly modeled as SIR type diffusion processes on a network, where nodes are in states Susceptible (S), Infectious (I) or Recovered (R), and an infected node spreads the infection to each susceptible neighbor independently with some probability p, e.g., [1], [2], [10]. Finding an optimal vaccination strategy in an epidemic model is an extremely challenging optimization problem, and despite the importance of the optimal vaccination problem, this is generally open. Even when the networks are quite constrained and the infection spread is deterministic (i.e., $p = 1$), the problem is NP-hard [11], [12] and the only constant approximation algorithms known are bicriteria. When $p < 1$, the problem is generally open, and rigorous approximation results are known only for selected random graphs and regimes [13].

Since the minimization of the expected number of infections has proven to be difficult, a number of heuristics and alternative approaches, based on node degrees, have been explored. The most common approach is to reduce the maximum degree, which is motivated by the observation that vaccination strategies which prioritize high degree nodes have been shown to be effective in many settings, e.g., [1], [8], [14]–[16]. Another strategy that has been explored quite extensively is to control the spectral radius (the largest eigenvalue of the network), since epidemic dynamics in many models including the SIS model is characterized in terms of it [17]–[19]. However, as far as our knowlegde, no prior work has addressed these problems in the context of leaky vaccines.

A. Our Contributions

A limitation in prior work on optimizing vaccination strategies in networked models is that they assume the vaccine has perfect efficacy. This has clearly not been true in the case of the COVID-19 vaccines, and we refer to this as the "leaky vaccine" setting. The focus of our paper is to develop efficient

This work was partially supported by NSF grants CCF- 1918656 and CNS-2317193, CCF-1918749 and DTRA award HDTRA1-24-R-0028, Cooperative Agreement number 6NU50CK000555-03-01 from the Centers for Disease Control and Prevention (CDC) and DCLS.

control strategies for the leaky vaccine setting, which we shall show could significantly differ from the optimal vaccination set of the non-leaky setting. Our main contributions are summarized below.

- We formalize the problem of minimizing the spread of the disease by using non-perfectly-effective, leaky, vaccines. We also introduce three variations of the problem where leaky vaccines are used to minimize (i) the average degree and (ii) the maximum degree of the network since these simpler graph metrics are often used as proxies for controlling the expected spread of the infection. We show a simple example where the leaky and non-leaky settings have qualitatively different optimal solutions, which emphasizes the importance of our model (Section II-A).
- We formulate a sample-based linear program and rounding algorithm to solve infection minimization and variations (i) and (ii), obtaining various bicriteria approximation solutions. Specifically, we obtain an $(1+\epsilon, O(\log n))$-approximation for minimizing the expected infection for a specific class of random graphs (Section IV), and $(1+\epsilon, 2)$- and $(1+\epsilon, O(\log n))$-approximation for minimizing the average degree and maximum degree, respectively (Section III). In particular, we note that our results nearly match the best known approximation guarantees even in the non-leaky setting. We also present sample-based greedy algorithms that solves the two variations of the problem.
- Leveraging the convenience of the algorithm and the fact that the theoretical results hold for general graphs, we run experiments using sample population data of Montgomery County, VA with the objective of minimizing average and maximum degree. Our experiments investigated the effectiveness of incorporating the knowledge about the leakiness of the vaccines property (Section V).

II. Preliminaries

Given a network G, let $s \in V(G)$ be the initial source of the infection. For an edge e, let p_e be the *infection probability* of the event that an infected endpoint of e will transmit the disease to the other endpoint. For a vertex v, let q_v be the *effectiveness probability* representing the probability that an administered vaccine will work on protecting v. Let $d_v(G)$ denote the degree of node v in graph G, and let $\Delta(G) = \max_v d_v$ and $\bar{d}(G) = (\sum_v d_v)/n$ denote the maximum and average degree, respectively; when G is clear from the context, we drop it from the notation. Let $\rho(G)$ denote the largest eigenvalue of the adjacency matrix of G, also referred to as its spectral radius.

Problem 1 (Minimum Infection Problem with Leaky Vaccines (MinInfLV)): Given network G with source s, infection probability p_e, effectiveness probability q_v and a budget B, find a subset $Y \subseteq V(G)$ such that $|Y| \leq B$ and the expected number of infected people is minimized after administering vaccines to those in Y.

MinInfLV is quite hard, in general, and motivated by the relationship between the degree distribution with the disease dynamics, and the fact that $\rho(G)$ is upper and lower bounded by Δ and $\bar{d}$, we also study the following two formulations for epidemic control in terms of degrees, which are much easier to address.

Problem 2 (Minimize Average Degree with Leaky Vaccines (MinAvgDegLV)): Given network G, effectiveness probability q_v and a budget B, find a subset $Y \subseteq V(G)$ such that $|Y| \leq B$ and it minimizes the expected average degree of the graph after independently removing each vertex in $v \in Y$ with probability q_v.

Problem 3 (Minimize Maximum Degree with Leaky Vaccines (MinMaxDegLV)): Given network G, effectiveness probability q_v and a budget B, find a subset $Y \subseteq V(G)$ such that $|Y| \leq B$ and it minimizes the expected maximum degree of the graph after independently removing each vertex in $v \in Y$ with probability q_v.

Remark 1: Instead of applying vaccines to vertices, we can also impose social distancing to the edges with effective probability of q_e. We point out that our models and algorithms can be easily adapted to handle this variant of the above problems as well.

In our paper, we will give bicriteria approximate solutions, defined as follows:

Definition 2.1: An (α, β)-bicriteria approximation solution is a solution whose objective value is at most α times the optimal while the budget is violated by at most a β factor.

A. Importance of leakiness

A natural question when faced with the problem of administering leaky vaccines is whether or not the problem is qualitatively different from the non-leaky setting. In this section, we will give a simple example where the optimal solutions differ for the two settings due to leakiness. In fact, we will show that the optimal solution in the non-leaky setting can't obtain a good approximation to that of the leaky setting.

Our example graph is illustrated in Figure 1. Let us consider the case where the budget is $B = 2$ and the infection process is deterministic (i.e., $p = 1$); we will compare the leaky setting where $q_v = 0.5$ for each $v \in V$ and the non-leaky setting where $q_v = 1$ for each $v \in V$. In the non-leaky setting, one may vaccinate the two neighbors of the source, which would result in 1 infection. In the leaky setting, however, this solution will cause the expected number of infections to be over 550 since the ovals will be infected each with probability $1/2$. Suppose we instead chose the two nodes along the path to the larger oval (with 1000 nodes); then the probability that those nodes would be infected would drop to $1/4$ and the smaller oval would always be infected, so the expected number of infections drops to approximately 350. By scaling the size of the larger oval to infinity, we also obtain our "inapproximability" result.

More generally, optimal solutions for vaccinating nodes with leaky vaccines will invest more of the budget on saving the larger populations because the vaccines aren't as effective.

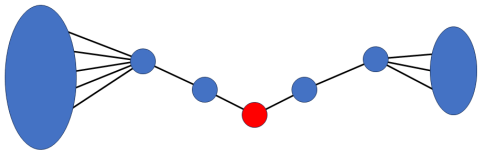

Fig. 1. A simple example illustrating that the leaky and non-leaky settings can have different optimal solutions. The ovals represent some connected graph, where the larger oval has 1000 nodes and the smaller oval has 100 nodes. The red node is the source node which is infected at the beginning.

As a result, we see that our new model is qualitatively different from previous work and merits further study.

B. Summary of our techniques

Our main tool is to use the sample-average technique where we solve a heuristic approximation of the problem using a linear programming. In particular, given $\epsilon > 0$, we generate $N = poly(n, 1/\epsilon)$ many sample instance of possible effectiveness of the disease (and possible spread of the disease for MinInfLV). Formally, for each $i = 1, ..., N$, for each vertex v, independently generate a binary variable $w_u^i \sim \text{Bern}(q_u)$. This variable represents whether administering vaccine to v will be effective or not. When solving MinInfLV, we additionally independently at random keep each edge e in sample graph G^i with probability q_e. This represents whether the disease will spread along edge e in sample i or not. This additional randomness in edges is required for MinInfLV because we care about how the disease spread in this problem while the other two problems only focuses on the effectiveness of the vaccine.

After generating these samples, we write an integer program that minimizes average of the given objective across all the generated samples. Then, we solve the LP relaxation of the program and use the appropriate rounding techniques to obtain a good approximation solution for the heuristic problem. Finally, to claim that we have an approximation algorithm, we must relate the heuristic solution to the actual objective, which we will achieve using the lemma below. For $i = 1, ..., N$, $Y \subseteq V(G)$, let $M^i(Y)$ be the objective value (e.g. average degree, max degree, or number of infection) in sample i if the Y is vaccinated. Let $M(Y)$ represent the expected objective value if we vaccinate Y in the original problem. Then, we have the following lemma, showing that the heuristic average for any solution is not too far from the its actual expected objective. Since this is a standard exercise using Chernoff-Hoeffding bounds, we defer the proof to the Appendix.

Lemma 2.1: Let $N = \frac{n^3 \log n}{\epsilon^2}$ and we generate N samples as above. Then, for any subset $Y \subseteq V(G)$, the probability that $|\frac{1}{N}\sum_{i=1}^{N} M^i(Y) - M(Y)| \geq \epsilon M(Y)$ is at most $\frac{1}{n^2}$.

III. Solving heuristic objectives

In this section, we present our algorithms and results for MinMaxDegLV and MinAvgDegLV. Due to space constraints, we defer all proofs to the Appendix.

A. Minimizing Average Degree

If the vaccines are perfectly effective, this problem can be recast as a vertex cover-type problem: choose B nodes which are incident to the maximum number of edges. To approximately solve this problem (in the bicriteria sense), we first guess the optimal number of edges, OPT, which can be covered by B nodes. Our problem then reduces to choosing the fewest number of nodes to remove such that OPT edges are covered. This is a partial vertex cover problem, which has a 2-approximation algorithm [20]. Since B vaccines suffices to remove OPT edges, applying this algorithm guarantees that (a) the number of vaccines used is at most $2B$ and (b) the number of edges removed is at least OPT. In other words, we have a $(1, 2)$-bicriteria approximation algorithm for minimizing the average degree of a graph.

We extend the above idea to the setting of leaky vaccines via the sample average framework for stochastic optimization. We generate N samples according to Section II-B. Our goal is now to choose B nodes to vaccinate such that the average degree (over all nodes and all samples) is minimized. Here, the average degree is computed based on removing nodes in sample i only if the node is vaccinated and vaccines are functional for the node in sample i. Since the samples are fixed, this becomes a deterministic problem and we can formulate it as a linear program:

$$\begin{aligned}
\min \quad & \textstyle\sum_{i=1}^{N} \sum_{e \in E} (1 - y_e^i) && && (1)\\
\text{s.t.} \quad & \textstyle\sum_{u \in V} x_u \leq B && && (2)\\
& y_e^i \leq w_u^i x_u + w_v^i x_v && \forall i \in [N], e \in E && (3)\\
& x_u, y_e^i \in [0, 1] && \forall u \in V, i \in [N], e \in E &&
\end{aligned}$$

In the integral version of the above LP, the variable x_u represent whether node $u \in V$ will be vaccinated. Constraint (1) enforces the budget. The variable y_e^i represents whether edge e will be removed in the resulting induced graph based on sample i. Constraint (3) enforces that y_e^i is removed (equals 1) only if one of its endpoint is administered an effective vaccine in sample i. Finally, the objective (1) is clearly equivalent to maximizing the average degree on the samples, so the above LP is a fractional relaxation of our problem. Consider the following algorithm:

Require: An instance of MinAvgDegLV and $\epsilon > 0$
Ensure: An $(1 + \epsilon, 2)$-bicriteria approximation solution

1: Generate N samples and solve the above LP.
2: Let $\lambda = 2(1 - \epsilon)$ and $S_1 = \{u \in V : x_u \geq \lambda\}$
3: For each $u \in V - S_1$, include u into set S_2 independently with probability λx_u.
4: Output $S_1 \cup S_2$

Theorem 3.1: Given $\epsilon > 0$, the algorithm described above is an $(1 + \epsilon, 2)$-bicriteria approximation for minimizing average degree using leaky vaccines with high probability.

B. Minimizing Maximum Degree

Given a graph $G = (V, E)$, suppose we know the minimum possible maximum degree of the graph G after removing

B nodes; let us denote this by OPT. Consider the reverse problem where we wish to remove the fewest number of nodes in order to reduce the maximum degree of the graph to OPT. In order to guarantee that the maximum degree is at most OPT after removing nodes, we need to either remove at least $d_G(u) - \text{OPT}$ neighbors of each node u or remove the node itself. This can be formulated as a multi-set multi-cover problem, where the elements are the nodes in the graph and the multi-sets are induced by removing nodes in the graph. That is, the multi-set corresponding to node u consists of n copies of u and one copy of each neighbor of u in G. Finally, we set the (multi-)covering requirement of each node u to be $\max\{0, d_G(u) - \text{OPT}\}$. Via this reduction, it's clear that our problem is a special case of multi-set multi-cover. Since multi-set multi-cover has an H_n-approximation, where $H_n \leq \ln(n) + 1$ is the n^{th} harmonic number, we can guess the value OPT via binary search to obtain a $(1, H_n)$-bicriteria approximation algorithm for minimizing the maximum degree of a graph.

We will now extend the ideas above to obtain a similar approximation guarantee for the setting of leaky vaccines, again using the sample average framework. Generate N samples according to Section II-B. Our goal is now to choose B nodes to vaccinate such that the average maximum degree over all the samples is minimized. We can write this as a linear program:

$$\begin{aligned}
\min \quad & \textstyle\sum_{i=1}^{N} z^i && (4)\\
\text{s.t.} \quad & \textstyle\sum_{v \in V} x_v \leq B && (5)\\
& y_e^i \leq w_u^i x_u + w_v^i x_v \quad \forall e = uv \in E, i \in [N] && (6)\\
& \textstyle\sum_{u:e=uv \in E} (1 - y_e^i) \leq z_v^i \quad \forall v \in V, i \in [N] && (7)\\
& z_v^i \leq z^i \quad \forall v \in V, i \in [N] && (8)\\
& x_v, y_e^i \in [0,1] \quad \forall v \in V, e \in E, i \in [N]
\end{aligned}$$

In the above LP, z_i represents fractional maximum degree in the induced subgraph of sample i, z_i^v represents the fractional degree of vertex v in the induced subgraph of sample i, x_v represents fractional extent to which node v is chosen for vaccination, and y_e^i represents the fractional extent to which edge e is removed in the induced subgraph of sample i. Constraint (5) enforces that the number of people chosen for vaccination is at most B. Constraint (6) enforces that y_e^i is indicated to be removed only when one of the nodes is chosen for vaccination and the vaccine works in sample i. Constraint (7) enforces that z_v^i represents the degree of node v in sample i and constraint (8) enforces that z^i represents the maximum degree in sample i. It is now clear that the objective in (4) is exactly the empirical expected maximum degree which we wish to minimize.

Given the above linear program, our algorithm proceeds as follows. We solve the linear program in polynomial time to obtain a fractional solution $x \in [0,1]^V$ to the problem, where x_v represents the fractional extent to which we should vaccinate node $v \in V$. To round the fractional solution to an integral one, we follow the randomized rounding approach for multi-set multi-cover given in Chapter of [21]. For each node $v \in V$, we flip a coin with probability x_v a total of $2 \log n$ times and include node v in our solution if any of the coin tosses lands heads. Our algorithm has the following approximation guarantee:

Theorem 3.2: The algorithm described above obtains an $(1+\epsilon, O(\log n))$-bicriteria approximation for minimizing maximum degree using leaky vaccines with high probability.

C. Greedy Algorithms

In addition to the above LP solutions, we also present a greedy algorithm for minimizing the heuristics. Again, we generate N samples as in Section II-B, with $w_v^i \sim \text{Bern}(q_v)$ being the indicator that on sample i, if vertex v were to be vaccinated, then it would be successful. All of the following are applications of Lemma 2.1 on standard results (such as partial set cover and multi-set multi-cover).

1) Minimizing Average Degree: This problem can be reduced to parital set cover. More specifically, the ground set is the set of all edges in all samples. Vaccinating vertex v would cover all edges incident to it in sample i if $w_v^i = 1$. In this case, minimizing the average degree over all samples is equivalent to covering the maximum number of edges. We can then use the greedy partial set cover algorithm to obtain a $\ln mN$ approximation, by iteratively selecting the vertices that cover the most number of the remaining edges.

2) Minimizing Maximum Degree: We use the reduction to multi-set multi-cover in Section III-B. Vaccinating vertex v would cover v^i $d_G(v)$ times and all of its neighbors u^i once if $w_v^i = 1$ for each sample i. If the minimum possible maximum degree for all samples after vaccinating B vertices is known to be OPT, consider the reverse problem of multi-set multi-cover with the covering requirement for each vertex in each sample being $\max 0, d_G(u^i) - OPT$. Since multi-set multi-cover has an H_k-approximation, k being the size of the largest set. Given that each multiset has size at most $nN = n^4 \log n/\epsilon^2$, we may obtain a $(1, 4\log n - 2\log \epsilon)$ approximation for MinMaxDegLV by guessing the value of OPT using binary search and applying mutli-set multi-cover.

IV. Algorithm For Minimizing Infections

In this section, we give our main algorithm for the problem of minimizing expected number of infections. Because no approximation algorithms are known for general graphs even with non-leaky vaccines, we focus on graphs where the number of simple paths of length k is at most $\text{poly}(n, 2^k)$. For example, it has been shown that a family of social network-like random graphs, namely Chung-Lu random graphs with power law degree distributions for parameter $\beta > 3$, satisfies this condition with high probability [13]. We remark that this is the only regime where approximation results are known for the non-leaky version of the problem as well.

Given $\epsilon > 0$, we generate $N = n^3 \log n/\epsilon^2$ samples as described in Section II-B. Then, we write the problem of minimizing the average number of infections over the samples as the following linear program:

$$\min_{x,y,z} \quad \sum_{i=1}^{N} \sum_{v \in V} (1 - y_v^i) \tag{9}$$

$$s.t. \quad y_u^i + x_e^i \geq y_v^i \qquad \forall e = uv \in E^i, i \in [N] \tag{10}$$

$$x_e^i \leq w_u^i z_u + w_v^i z_v \quad \forall e = uv \in E^i, i \in [N] \tag{11}$$

$$\sum_{v \in V} z_v \leq B \tag{12}$$

$$x_e^i, y_v^i, z_v \in [0,1] \qquad \forall e \in E^i, v \in V, i \in [N] \tag{13}$$

This LP attempts to minimize infection across all N samples. As an integer program, x_e^i is 1 if and only if infection cannot spread along edge e in sample i. The variable y_v^i is 1if and only if v is not infected in sample i. Lastly, z_v represents whether vaccine is administered to v or not. The objective hence is to minimize the total (or equivalently average) number of infected across all samples. Constraint 10 says v must be infected if u is infected and x_e^i is 0 in sample i. Constraint 11 limits x_e^i, allowing it to be 1 only if at least one of its endpoint received an effective administered vaccine. Constraint 12 is the budget constraint. Along with the last binary constraints, the integral solution clearly solves the sampled version of MinInfLV. Let Γ be the expected number of paths in a random sample due to percolation of the disease. Given $\epsilon > 0$, let $\epsilon\gamma = 2(4 + \log_n \Gamma)$.

Theorem 4.1: Given an instance of MinInfLV, a constant ϵ and define γ defined as above.. Then, there exists a polytime randomized algorithm that vaccinate a set of nodes V such that $|V| \leq \gamma \log n \times B$ and the expected number of infected is at most $\frac{1+\epsilon}{(1-\epsilon)^2}$ times the optimal with probability $1 - O(n^{-\gamma}) - O(1/n)$.

Require: An instance of MinInfLV and $\epsilon > 0$

Ensure: A subset $Y \subseteq V(G)$ to vaccinate such that the cost is at most $O(\log n) \cdot B$ and the expected number of infection is at most $(1 + O(\epsilon)) \times OPT$

1: Generate N samples and solve the above LP

2: For every $v \in V(G)$, put v in Y with probability $\frac{\gamma}{\epsilon} \log n \cdot z_v$ where $\gamma = 2(4 + \log_n \Gamma)$

3: Output Y.

V. Experiments

a) Experimental Setup.: To examine the experimental efficacy of our proposed algorithms, we simulate the process of leaky vaccination on the population of Montgomery County, Virginia. Our experiments use a synthetic social contact network for this population [22], [23], which has also been used for COVID-related simulations in several other works [15], [16]. Our algorithms are run on a subsets of the population, since our methods are intended to work on smaller communities at the outbreak of an epidemic: where mapping contact networks may be feasible and beneficial. Unless otherwise stated, our default experimental parameters are listed below in Table I.

After obtaining the rounded vaccination recommendation set, all algorithm performances are evaluated using the same set of samples that simulate effective/ineffective vaccinations. This set of test samples is separate from the set of samples used to formulate the leaky algorithms. We also simulate the spread of a disease over the social network using 15 randomly generated initial infection sets. Each infection set contains 15 vertices selected at random from the social network. The vertices in this set uniformly transmit the infection with a probability of 0.1 to their direct neighbors. When the infection has stopped propagating, we calculate the total number of people infected in the network to evaluate the infection spread.

TABLE I
Default Algorithm Experiment Parameters

Parameter	Value
Vaccine Leak Probability	0.2
Budget (Number of Vaccinations)	200
Sample Size for LP Formulation	250
Sample Size for Algorithm Evaluation	250
Number of Algorithm Trials	15
Rounding Approach	No Repetition
Initial Infection Set Size	15
Disease Transmission Rate	0.1
Number of Infection Trials	15

We utilize Gurobi Optimizer to formulate and solve our linear programs. The complete code for all experiments is available through GitHub.

b) Implementation Details.: In our experiments, we will be evaluating the performance of our algorithms for decreasing the average and maximum degrees of a network from Section III. In our proofs in Section III, we require a large polynomial number of samples N in order to give our approximation guarantees. In practice, however, this is not computationally feasible and it has been observed that a few hundred samples suffices for good performance in sample average-type algorithms [24]. As a result, we use $N = 250$ samples for each instance of our algorithms. To accommodate for the reduction in the number of samples, we run 15 independent instances of our algorithms. As indicated by the confidence intervals in our figures, the variance in the resulting objectives is very low.

c) Evaluation.: To assess the performance of our algorithms, we generate 250 samples simulating different scenarios of vaccine leakiness for a given leak probability. After applying our algorithm's recommended vaccination set, we calculate metrics of the vaccinated social network for each sample. We also simulate the spread of a disease over the social network, and measure the total number of people infected in the network using 15 random samples.

TABLE II
Montgomery Neighborhood Network Properties

Subpopulation	Nodes	Edges	AvgDeg	MaxDeg	Infection Spread
1	1000	6581	13.162	57	315.62
2	1000	7006	14.012	68	443.40
3	1000	6925	13.850	70	340.66

A. Comparison with Non-Leaky Formulation

In Sections 4.1 and 4.2, our algorithms for the degree-based objectives using leaky vaccines were inspired by some

simple algorithms in the non-leaky setting, described in the beginnings of the respective subsections. These algorithms can be seen as a special case of our algorithms with the assumption that $q_v = 1$ for all $v \in V$ so no sampling is needed. The goal of the experiments in this subsection is to investigate whether or not our algorithms which use information about the leaky probabilities q_v actually improve upon these baseline algorithms. For our experiments here, we will study leakiness probability values which range up to a probability of 0.3 to mirror that of COVID-19 [25]. In Figure 2, we evaluate the vaccination recommendations proposed by the baseline algorithms and our leaky algorithms on the three Montgomery subpopulations.

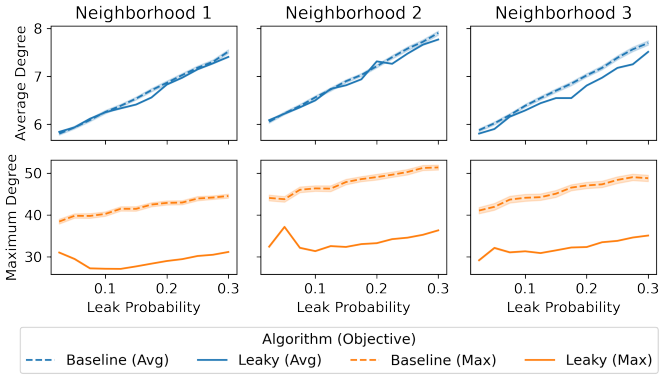

Fig. 2. Degree of Leak v. Non-Leak Algorithms

In the figure above, we observe that our algorithms perform much better than the baseline for the max degree objective. This is in sharp contrast to the average degree objective, where the two algorithms have nearly the same performance. We believe that this is because the average degree objective has nice linearity properties so when the leaky probabilities are uniform, minimizing the expected average degree is similar for both settings. In contrast, it is much harder to reduce the maximum degree of a graph. For example, if only one of the two max-degree nodes were successfully vaccinated, it decreases the average degree but fails to change the max degree. Thus, there are more considerations which needs to be taken into account when minimizing the max degree. Factoring in leakiness will allow the algorithm to provide a more diverse recommendation set that can reduce the maximum degree even in the case that the maximum-degree node itself cannot be successfully vaccinated. Therefore, leakiness is particularly influential when considering super-spreaders in an infection network.

B. Infection Spread Baseline Comparison

We have determined that our leaky algorithms minimize their respective degree objectives, but how well do they reduce the spread of an infection? In Figure 3, we visualize the efficacy of the vaccination recommendations in reducing the infection spread on the three Montgomery neighborhoods. We observe that a reduction in the average degree objective

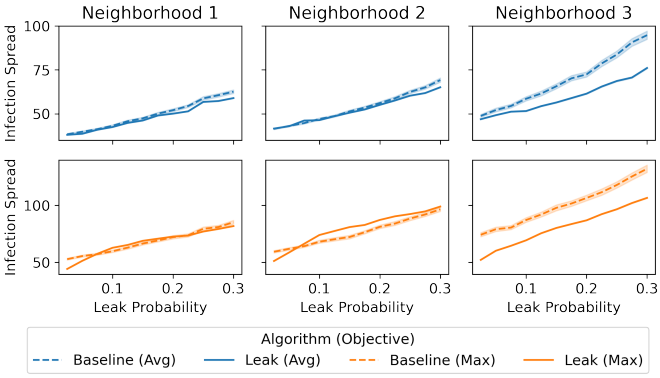

Fig. 3. Infection Spread of Leak v. Non-Leak Algorithms

translates to a reduction in infection spread. Therefore, incorporating leakiness in the formulation of vaccine distribution reduces the spread of the infection.

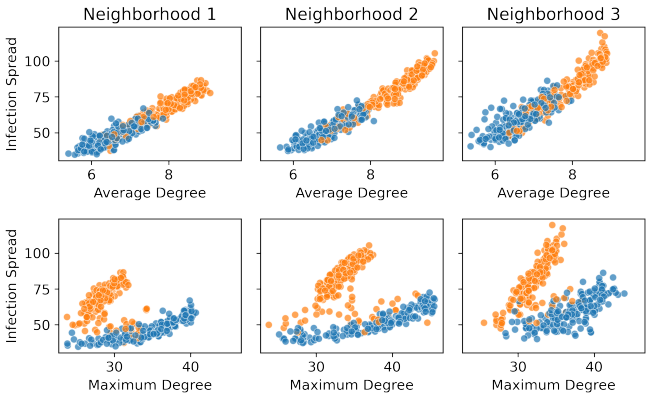

Fig. 4. Correlation between Infection Spread and Heuristics

In Figure 4, we plot the infection spread on a vaccinated network with respect to the network's average and maximum degree. These vaccinated networks are produced with the vaccination recommendation sets and leak probabilities with the same parameters as Section **??**. Across all three of our Montgomery neighborhood networks, we notice that the average degree objective shows a stronger correlation with the infection spread, regardless of the algorithm used to obtain that post-vaccination network. On the other hand, we see two distinct trends when evaluating infection spread with respect to the maximum degree of the vaccinated network. Even if two vaccinated networks have the same maximum degree, the average degree objective vaccination recommendation will inhibit infection spread more than the maximum degree objective vaccination recommendation. Therefore, it appears that the heuristic of minimizing the average degree of a network is better aligned with the objective of reducing infection spread.

VI. Conclusion

Our work introduces a new class of intervention problems for epidemics based on the observation that vaccines aren't

always 100% effective. We show that for a variety of objectives, we can obtain strong approximation algorithms for the leaky version of the problem. Finally, we show experimentally that our algorithms improve upon baselines which don't take leakiness into account.

There are limitations to our work. We make the assumption of uniform vaccine efficacy and transmission probability, which may vary (constitution of individual, duration of exposure etc). These subtle differences may be difficult to measure in practice especially during earlier stages of epidemics, hence we did not include these variations in our experiments.

VII. Appendix

A. Proof of Theorem 4.1

Proof 1 (Proof of 4.1): Given an instance of MinInfLV, apply Algorithm IV to obtain a solution Y. Let x, y, z be the optimal LP solution that achieves an LP optimal value of OPT_{LP}. Recall $M^i(Y)$ refers to the number of infected if we vaccinate Y in sample i. Let $M^{emp}(Y) = \frac{1}{N}\sum_{i=1}^{N} M^i(Y)$. We first show that $M^{emp}(Y)$ is comparable to OPT_{LP}.

Let $V_\epsilon^i := \{v \in V(G) : y_v^i \geq \epsilon\}$. Let $\mathcal{B}$ be the event that there exists some v, i such that $v \in V_\epsilon^i$ and v is infected in sample i. We want to show that this event occurs with low probability. Consider a particular vertex v and a sample i. The vertex will be infected in the sample i if there exists some sv-path where none of the vertices on the path with an effective vaccine was chosen in Y. Let $\mathcal{P}_v^i$ be the set of all sv-paths in sample i. Then,

$$\begin{aligned}
&\Pr[v \text{ is infected in sample } i] \\
&\leq \textstyle\sum_{P\in\mathcal{P}_v^i} \Pr[v \text{ is infected via path } P] \\
&\leq \textstyle\sum_{P\in\mathcal{P}_v^i} \prod_{v\in V(P)} (1-z_v)^{\gamma \log n} \\
&\leq \textstyle\sum_{P\in\mathcal{P}_v^i} n^{-\gamma \sum_{v\in V(P)} z_v}
\end{aligned}$$

The first inequality is by union bound. Note that:

$$\begin{aligned}
\textstyle\sum_{v\in P} z_v &\geq (1/2) \textstyle\sum_{e=uv\in E(P)} z_u + z_{u'} \\
&\geq (1/2) \textstyle\sum_{e=uu'\in E(P)} w_u^i z_u + w_{u'}^i z_{u'} \\
&\geq (1/2) \textstyle\sum_{e\in E(P)} x_e^i \\
&\geq (1/2) y_v^i
\end{aligned}$$

The first inequality is achieved by double-counting every vertex along the path. The second inequality is true because w^i's are binary. The third and last inequality follows from Constraints 11 and 10 respectively. Then for any v where $y_v^i \geq \epsilon$, we have

$$\begin{aligned}
\Pr[v \text{ is infected in sample } i] &\leq \textstyle\sum_{P\in\mathcal{P}_v^i} n^{-\gamma y_v^i/2} \\
&\leq |\mathcal{P}_v^i| n^{-\gamma\epsilon/2}
\end{aligned}$$

Using this, we obtain that

$$\begin{aligned}
\Pr[\mathcal{B}] &\leq \textstyle\sum_{i=1}^{N} \sum_{v\in V_\epsilon^i} \Pr[v \text{ is infected in sample } i] \\
&\leq \textstyle\sum_{=1}^{N} \sum_{v\in V_\epsilon^i} |\mathcal{P}_v^i| n^{-\gamma\epsilon/2}
\end{aligned}$$

Note that the total expected number of paths in any sample is Γ. By Markov's inequality, the probability that $\sum_{=1}^{N}\sum_{v\in V_\epsilon^i} |\mathcal{P}_v^i|$ is greater than $N\Gamma n$ is at most $1/n$. Thus, due to our choice of γ, it follows that

$$\Pr[\mathcal{B}] \leq 1/n + N\Gamma n n^{-\gamma\epsilon/2} \leq O(1/n).$$

Now, assume event $\mathcal{B}$ does not occur. Then we know that every vertex in V_ϵ^i will not be infected. Thus, we have

$$\begin{aligned}
M^{emp}(Y) &\leq \textstyle\sum_{i=1}^{N} \sum_{v\in V_\epsilon^i} 1 \\
&\leq \textstyle\sum_{i=1}^{N} \sum_{v\in V_\epsilon^i} (1-y_v)/(1-\epsilon) \\
&\leq \text{OPT}_{LP}/(1-\epsilon).
\end{aligned}$$

Let Y^* be the optimal integral solution to MinInfLV. By Lemma 2.1, we know that with probability $1 - 2/n^2 - 1/n \geq 1 - O(1/n)$, event $\mathcal{B}$ does not happen so that $(1-\epsilon)M(Y) \leq M^{emp}(Y)$ and $M^{emp}(Y^*) \leq (1+\epsilon)M(Y^*)$. As a result, we have the following with high probability:

$$\begin{aligned}
M(Y) &\leq \frac{1}{1-\epsilon} M^{emp}(Y) \\
&\leq \frac{1}{(1-\epsilon)^2} \text{OPT}_{LP} \\
&\leq \frac{1}{(1-\epsilon)^2} M^{emp}(Y^*) \\
&\leq \frac{1+\epsilon}{(1-\epsilon)^2} M(Y^*).
\end{aligned}$$

Lastly, note that each node v is chosen independently. Therefore, the expected cost of Y is $\gamma \log n \cdot B$. Using Markov's inequality, the probability that the number of nodes chosen is more than $(1+\epsilon)\gamma \log n \cdot B$ is at most $1/(1+\epsilon)$, completing our proof.

B. Proof of Lemma 2.1

Proof 2: Given any $Y \subseteq V(G)$, when generating the N samples, let $M^{emp}(Y) = \frac{1}{N}\sum_{i=1}^{N} M^i(Y)$. Note that $\mathbb{E}[M^{emp}(Y)] = \mathbb{E}[M^i(Y)] = M(Y)$ for any $i = 1, ..., N$ by linearity of expectation and definition. Each random variable $M^i(Y)$ is also bounded between 1 and n, regardless the objective (average degree, max degree or number of infection). Then,

$$\begin{aligned}
&Pr\left[|M^{emp}(Y) - M(Y)| > \epsilon M(Y)\right] \\
=&Pr\left[\left|\sum_{i=1}^{N} M^i(Y) - N \cdot M(Y)\right| > \epsilon N \cdot M(Y)\right] \\
\leq& 2\exp\left(-\frac{2(N\cdot M(Y))^2}{\epsilon N \cdot n^2}\right) \\
\leq& 2\exp\left(-\frac{2\epsilon^2 N}{n^2}\right) \leq \frac{1}{n^2 2^n}
\end{aligned}$$

The first inequality is by Hoeffding's Inequality. The second is due to $M(Y) \geq 1$. Since this holds for any subset Y and there are at most 2^n choices of subsets, the lemma follows from a simple union bound.

C. Proof of Theorem 3.1 and 3.2:

Proof 3: We will give a sketch of this proof since it relies heavily on existing results for randomized rounding algorithms for partial vertex cover and multiset multicover respectively. Recall by Lemma 2.1, we know that with high probability, the sample average objective and the true (expected) objective are within a multiplicative factor of $1 \pm \epsilon$ of each other. We will condition on this event for the remainder of the proof; our results will still hold with high probability by the union bound. Since we are conditioning on this event, it suffices to show that the algorithm's output satisfies the following: the empirical objective matches that of the LP and the number of nodes chosen is at most $2B$ for MinAvgDegLV and $B \cdot O(\log n)$ for MinMaxDegLV. This would imply the claimed result.

Let OPT_{avg} and OPT_{max} denote the optimal value of the linear program in Section III-A and III-B. Observe that the optimal solution x^* and $x^{*\prime}$ to the linear program in Section III-A and III-B is also a feasible fractional solution to the partial coverage version of the sample average problem, where the goal is to use the fewest number of (fractional) vaccines so that the average (or maximum) degree is at most OPT_{avg} and OPT_{max}. For MinAvgDegLV, since this is a partial coverage problem, we know from that the rounding algorithm has the following guarantees: the average degree will be at most OPT_{LP} and the number of nodes used will be at most $2 \cdot \sum_{v \in V} x_v^* \leq 2B$, as desired. See [26] and references therein for a proof of the above statement. In the case of MinMaxDegLV, this linear program is no longer an instance of a multi-set multi-cover linear program so we cannot directly claim the desired results by [21]. Fortunately, the proof ideas still apply: we can still claim that the elements which need to be covered are indeed covered by the rounded solution with high probability. By a union bound, all the required elements are covered (still with high probability). As a result, we know that the rounded solution has (average) maximum degree at most that of the fractional solution. Finally, the same argument as for multi-set multi-cover shows that with high probability, the rounded solution uses at most $O(B \cdot \log n)$ vaccines.

REFERENCES

[1] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Reviews of modern physics*, vol. 87, no. 3, p. 925, 2015.

[2] M. Marathe and A. Vullikanti, "Computational epidemiology," *Communications of the ACM*, vol. 56, no. 7, pp. 88–96, 2013.

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

[4] S. M. Moghadas, T. N. Vilches, K. Zhang, C. R. Wells, A. Shoukat, B. H. Singer, L. A. Meyers, K. M. Neuzil, J. M. Langley, M. C. Fitzpatrick, *et al.*, "The impact of vaccination on coronavirus disease 2019 (covid-19) outbreaks in the united states," *Clinical Infectious Diseases*, vol. 73, no. 12, pp. 2257–2264, 2021.

[5] L. Matrajt, J. Eaton, T. Leung, D. Dimitrov, J. T. Schiffer, D. A. Swan, and H. Janes, "Optimizing vaccine allocation for covid-19 vaccines shows the potential role of single-dose vaccination," *Nature communications*, vol. 12, no. 1, p. 3449, 2021.

[6] C. E. Wagner, C. M. Saad-Roy, and B. T. Grenfell, "Modelling vaccination strategies for covid-19," *Nature Reviews Immunology*, vol. 22, no. 3, pp. 139–141, 2022.

[7] M. E. Halloran and C. J. Struchiner, "Thirty-five years of leaky vaccines," *American Journal of Epidemiology*, vol. 194, pp. 918–920, 10 2024.

[8] F. Sartori, M. Turchetto, M. Bellingeri, F. Scotognella, R. Alfieri, N.-K.-K. Nguyen, T.-T. Le, Q. Nguyen, and D. Cassi, "A comparison of node vaccination strategies to halt sir epidemic spreading in real-world complex networks," *Scientific Reports*, vol. 12, no. 1, p. 21355, 2022.

[9] M. Salathé and J. H. Jones, "Dynamics and control of diseases in networks with community structure," *PLoS computational biology*, vol. 6, no. 4, p. e1000736, 2010.

[10] M. E. Newman, "Spread of epidemic disease on networks," *Physical review E*, vol. 66, no. 1, p. 016128, 2002.

[11] A. Hayrapetyan, D. Kempe, M. Pál, and Z. Svitkina, "Unbalanced graph cuts," in *ESA*, pp. 191–202, 2005.

[12] S. G. Eubank, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, and N. Wang, "Structural and algorithmic aspects of massive social networks," in *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004* (J. I. Munro, ed.), pp. 718–727, SIAM, 2004.

[13] A. E. Babay, M. Dinitz, A. Srinivasan, L. Tsepenekas, and A. Vullikanti, "Controlling epidemic spread using probabilistic diffusion models on networks," in *International Conference on Artificial Intelligence and Statistics*, pp. 11641–11654, PMLR, 2022.

[14] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[15] G. Z. Li, A. Haddadan, A. Li, M. V. Marathe, A. Srinivasan, A. Vullikanti, and Z. Zhao, "Theoretical models and preliminary results for contact tracing and isolation," in *21st International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2022, Auckland, New Zealand, May 9-13, 2022* (P. Faliszewski, V. Mascardi, C. Pelachaud, and M. E. Taylor, eds.), pp. 1672–1674, International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2022.

[16] J. Chen, S. Hoops, A. Marathe, H. Mortveit, B. Lewis, S. Venkatramanan, A. Haddadan, P. Bhattacharya, A. Adiga, A. Vullikanti, A. Srinivasan, M. L. Wilson, G. Ehrlich, M. Fenster, S. Eubank, C. Barrett, and M. Marathe, "Effective social network-based allocation of covid-19 vaccines," in *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, (New York, NY, USA), p. 4675–4683, Association for Computing Machinery, 2022.

[17] S. Saha, A. Adiga, B. A. Prakash, and A. Vullikanti, "Approximation algorithms for reducing the spectral radius to control epidemic spread," in *SIAM SDM*, 2015.

[18] Y. Zhang, A. Adiga, A. Vullikanti, and B. A. Prakash, "Controlling propagation at group scale on networks," in *Data Mining (ICDM), 2015 IEEE International Conference on*, pp. 619–628, IEEE, 2015.

[19] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," in *2011 IEEE 11th International Conference on Data Mining*, IEEE, Dec. 2011.

[20] R. Gandhi, S. Khuller, and A. Srinivasan, "Approximation algorithms for partial covering problems," *J. Algorithms*, vol. 53, no. 1, pp. 55–84, 2004.

[21] V. V. Vazirani, *Approximation algorithms*. Springer, 2001.

[22] C. L. Barrett, R. J. Beckman, M. Khan, *et al.*, "Generation and analysis of large synthetic social contact networks," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pp. 1003–1014, 2009.

[23] S. Eubank, H. Guclu, V. S. A. Kumar, *et al.*, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, p. 180—184, May 2004.

[24] P. Sambaturu, M. Minutoli, M. Halappanavar, A. Kalyanaraman, and A. Vullikanti, "Scalable and memory-efficient algorithms for controlling networked epidemic processes using multiplicative weights update method," in *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022* (L. D. Raedt, ed.), pp. 5164–5170, ijcai.org, 2022.

[25] N. Wu, K. Joyal-Desmarais, P. A. Ribeiro, A. M. Vieira, J. Stojanovic, C. Sanuade, D. Yip, and S. L. Bacon, "Long-term effectiveness of covid-19 vaccines against infections, hospitalisations, and mortality in adults: Findings from a rapid living systematic evidence synthesis and meta-analysis up to december, 2022," *The Lancet Respiratory Medicine*, vol. 11, p. 439–452, May 2023.

[26] A. Srinivasan, "Distributions on level-sets with applications to approximation algorithms," in *42nd Annual Symposium on Foundations*

of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA, pp. 588–597, IEEE Computer Society, 2001.