Who and When to Screen: Multi-Round Active Screening for Recurrent Infectious Diseases Under Uncertainty

Han-Ching Ou[1], Arunesh Sinha[2], Sze-Chuan Suen[1], and Andrew Perrault[1]
Milind Tambe[1]

[1] University of Southern California hanchino@usc.edu
[2] University of Michigan

Abstract. Controlling recurrent infectious diseases is a vital yet complicated problem. In this paper, we propose a novel active screening model (ACTS) and algorithms to facilitate active screening for recurrent diseases (no permanent immunity) under infection uncertainty. Our contributions are: (1) A new approach to modeling multi-round network-based screening/contact tracing under uncertainty, which is a common real-life practice in a variety of diseases [10, 30]; (2) Two novel algorithms, Full- and Fast-REMEDY. Full-REMEDY considers the effect of future actions and finds a policy that provides high solution quality, where Fast-REMEDY scales linearly in the size of the network; (3) We evaluate Full- and Fast-REMEDY on several real-world datasets which emulate human contact and find that they control diseases better than the baselines. To the best of our knowledge, this is the first work on multi-round active screening with uncertainty for diseases with no permanent immunity.

Keywords: Active screening · social networks · optimization

1 Introduction

Contagious diseases, such as influenza, gonorrhea, and chlamydia are critical public-health challenges that continue to threaten lives and economic productivity. While low-cost treatment programs are available, individuals may ignore symptoms and delay care, increasing transmission risk. Public health agencies may therefore engage in active screening or contact tracing efforts, where individuals in the community are asked to undergo diagnostic tests and are offered treatment if tests return positive results [10, 6].

However, in many settings, active screening/contact tracing is expensive and time consuming. Even in the United States, budget cuts in 52% of states and STD programs in 2012 has impacted the quality and quantity of the contact tracing [5]. Efficiently identifying infectious cases is therefore of vital importance.

Our *first contribution* is a model of the active screening problem (ACTS). We focus on recurrent infectious diseases that assumes contact to be in structured

networks with the SIS model of transmission [29], which is applicable for a wide range of diseases such as pertussis, syphilis and typhoid. The SIS model is the foundation of more complex models that capture more disease dynamics (such as latent states, variation in birth/death rates, or multiple treatment states). In the SIS model, individuals can be either susceptible (S) (currently healthy, but may become exposed) or infected (I). We consider diseases for which there is no means to achieve permanent immunity and prove the ACTS problem to be NP-hard. We assume that health workers are uncertain about the health state of individuals, have a small budget relative to population size for active screening and must engage in active screening over multiple rounds (time periods). To the best of our knowledge, no other models consider multi-round active screening for network SIS diseases in the AI literature.

Our *second contribution* is a novel algorithm—**RE**current screening **M**ulti-round **E**fficient **DY**namic algorithm (**REMEDY**)—to guide scalable active screening. We develop two versions of the algorithm, Full- and Fast-REMEDY. In Full-REMEDY, we consider both current and future actions simultaneously to understand the underlying disease dynamics and uncertainty of individuals' health states. Fast-REMEDY reduces the time complexity to scale to very large networks by exploiting eigendecomposition techniques. We illustrate the benefits of Full- and Fast-REMEDY via extensive testing on a set of real-world human contact networks against various baselines across a range of realistic disease parameters.

The paper is structured as follows. In Section 2, we provide background on disease modeling and active screening. In Section 3, we formalize the active screening problem (ACTS) and prove that it is NP-hard. In Section 4, we present REMEDY, our novel algorithm for ACTS. In Section 5, we empirically analyze REMEDY and compare its performance to relevant baselines.

2 Disease Model and Background

We first introduce the disease model notation for our problem. In an SIS model [1, 2], an individual can either be in state S (a healthy individual *susceptible* to disease) or I (the individual is *infected*). SIS models recurrent diseases, where permanent immunity is not possible (e.g., TB, typhoid) and not diseases such as Hepatitis A and measles, which follow a SIR or SEIR pattern where treated individuals may achieve permanent immunity by entering R state.

2.1 Disease Model

We adopt a discrete time SIS model for modeling the disease dynamics, which was earlier considered by Wang et al. [29]. Given a contact network $G(V, E)$, infection spreads via the edges in the network. There are $|V|$ individuals, and let $\delta(v)$ denote neighbors of node v in the network. Each individual (node) v in the network (at time t) is in state $\mathbf{s}_v(t) \in \{S, I\}$. Let $\mathbf{t}_v(t)$ denote the state vector that represents the true state of node v at time t where S is represented

as $[1,0]^\top$ and I as $[0,1]^\top$. Given the initial state, a infected node infects its healthy neighbors with rate α independently and recovers with probability c. The health state transition probabilities of a node is given by $P\left[s_v(t+1) = \{S, I\}\right] = \mathbf{T}_v^N(t)\mathbf{t_v}(t)$ where

$$\mathbf{T}_v^N(t) = \begin{array}{c} \\ S \\ I \end{array} \begin{array}{c} \begin{array}{cc} S & I \end{array} \\ \begin{bmatrix} 1-q_v & c \\ q_v & 1-c \end{bmatrix} \end{array}, \tag{1}$$

where $q_v = 1-(1-\alpha)^{|\{u \in \delta(v) \mid \mathbf{s_u}(t)=I\}|}$. Note that the columns denote the state of v at time t and the rows denote the state at $t+1$. The transition probabilities follow the disease dynamics described earlier. In particular, q_v captures the exact probability that node v becomes infected from its infected neighbors $\{u \in \delta(v) \mid \mathbf{s_u}(t) = I\}$ and c captures the probability that I individuals recover without active screening.

Given such transition probabilities and an initial state, if no intervention happens, the network state evolves by flipping biased coins for each node to determine their next true state in each time step. The process is repeated until the terminal step T is reached.

2.2 Prior Approaches to Active Screening

Most previous work on active screening deals primarily with SIR or SEIR type diseases, often referred to as the *Vaccination Problem* [3, 25, 28, 31, 12], where permanent immunization (entry into R state) can be viewed as removing nodes from the graph [23, 26]. Exploiting this idea, Saha et al. [26] and Tong et al. [23] focus on immunization ahead of an epidemic and suggest a heuristic method of removing a set of k nodes based on the eigenvalues of the adjacency matrix. Zhang and Prakash [31] consider the problem of selecting the best k nodes to immunize in a network after the disease has started to spread. These methods assume that the exact status of each node is known and deal with a single round of screening that offers permanent immunity.

However, for diseases in which there is no permanent immunity, one-time screening (cure) is not enough and, further, it may not be reasonable to quarantine patients until the disease has died out. We focus on a multi-round screening of SIS diseases that cannot be permanently cured. To the best of our knowledge, this complex setting has not been studied previously. Generally, the problem of minimizing disease spread is different from the well-studied problem of influence maximization [15, 8], where one optimizes the selection of seeds or starting nodes for maximizing influence spread, as opposed to optimizing the selection of nodes on which to intervene in order to minimize disease spread.

3 The Active Screening (ACTS) Problem

Motivated by active screening/contact tracing campaigns that has been practiced since the 1980s [6] and applied in various forms/diseases [18, 5], we propose the

Active Screening (ACTS) Problem. Given the SIS model in previous section, an active screening agent seeks to determine the best node sets $C_a(t) \subset V$ to actively screen and cure with a limited budget of $|C_a(t)| \leq k$ at each time step t. The agent does not know the the ground truth health state of all individuals. The agent knows the ground truth of the network structure $G(V, E)$, the infection probability α and recovery probability c. In addition, the agent observes the *naturally cured* node set $C_n(t)$ at time t—because this set of patients come to the clinic voluntarily. The active screening happens after the agent acquires information about $C_n(t)$. Let $C_a(t)$ be the set of nodes that are actively screened at time t. A node $v \in C_n(t) \cup C_a(t)$ becomes cured at time $t+1$. Thus, the transition matrix for a node $v \in C_n(t) \cup C_a(t)$ is $P\left[\mathbf{t_v}(t+1) = \{S, I\}\right] = \mathbf{T}_v^A(t)\mathbf{s_v}(t)$, where

$$\mathbf{T}_v^A(t) = \begin{array}{c} \\ S \\ I \end{array} \begin{array}{c} \begin{array}{cc} S & I \end{array} \\ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \end{array}. \tag{2}$$

It is worth noting that the action the agent takes at time t does not affect the transition matrix $\mathbf{T}_v^N(t)$ of the nodes not involved in active screening.

Unlike most of the previous work which focused on minimizing the budget spent until the long-term disease eradication is achieved, we focus on maximizing the health quality of each individual at each time step. We treat each individual and time step equally, although it is easy to modify the costs and values to be weighted. The objective of the ACTS problem is:

$$\min_{C_a(0),\ldots,C_a(T)} \mathbb{E}\left[\sum\nolimits_{t=0}^{T} \sum\nolimits_{v \in V} |\mathbf{s_v}(t) = I|\right]. \tag{3}$$

Problem Statement *(ACTS Problem) Given a contact network $G(V, E)$, the disease and active screening model, find an active screening policy such that the expectation of $\sum_{t=0}^{T} \sum_{v \in V} |\mathbf{s_v}(t) = I|$ is minimized.*

Even assuming we know the ground truth infected state for each node, ACTS is NP-hard.

Theorem 1. *The ACTS Problem is NP-hard.*

Proof. We reduce the VertexCover to the decision problem "Does there exist a curing strategy of objective function equals $5|V| - 2k$ with budget of k each round of the constructed ACTS problem?"

Given a VertexCover decision problem with graph $G = (V, E)$ and budget k, we construct a new graph $G^* = (V_0^* \cup V_1^* \cup V_2^*, E^*)$ as follows: First, for each node $v \in V$, create three nodes v_0, v_1 and v_2 in G^*. Second, for each node $v \in V$, create an edge (v_0, v_1) in G^*. Finally, for each edge $(u, v) \in E$ create two edges, (u_1, v_2) and (u_2, v_1) in G^*. We set the parameters of the ACTS problem to be $(\alpha, c) = (1, 0)$ and $T = 2$ with budget of k in each round. The initial state of the graphs are $s_v(0) = I \ \forall \ v \in V_0^*$ and $s_v(0) = S \ \forall \ v \in V_1^* \cup V_2^*$. Fig. 1 shows a simple example.

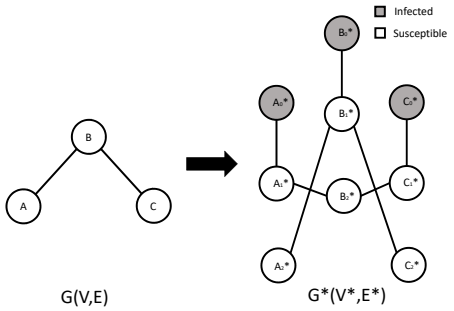


Fig. 1. A simple example of graph transformation for problem deduction.

We now argue that G has a vertex cover of size k if and only if the ACTS problem of the above setting has the objective function of $5|V|-2k$. In the above setting, we get to act twice. Acting at $t = 0$ allows us to force k nodes into S state at $t = 1$. Denote the objective function we get at time t as $Score(t)$, no matter what nodes we chose at $t = 0$, our sum of score in the first two rounds is always going to be $Score(0) = |V|$, $Score(1) = 2|V| - k$ and for the action we take at $t = 1$ will only reduce $Score(2)$ by amount of k, as long as we pick nodes in I state since it has no chance to propagate. Thus the only action matters is the action on $t = 0$ toward $Score(2)$. Picking the copy of vertex cover set of G in V_1^* results in $|V| + (|V| - k) + k$ of I nodes in $t = 2$, which are all the nodes in V_0^*, all the nodes in V_1^* except vertex cover copy and the vertex cover in V_2^*. We argue that this is the optimal strategy as picking anything that is not vertex cover results more than k infected nodes in V_2^*. Then we pick arbitrary k nodes as our action in $t = 1$ and results a score of $Score(2) = 2|v| - k$. The objective function we gain is $5|V| - 2k$ if and only if the vertex cover exist. Thus we have proven the ACTS problem to be NP-hard.

4 Algorithm for the ACTS problem

We introduce REMEDY, an algorithm for selecting nodes to actively screen in the ACTS problem. REMEDY, shown in Algorithm 1, contains two parts: (i) a marginal belief state update that we use for reasoning about the infected status of nodes, and (ii) an algorithm for selecting which nodes to actively screen based on the marginal belief state and an upper bound of the ACTS objective.

4.1 Belief State Update

Tracking the exact probability that a node is infected in ACTS requires storing $O(2^{|V|})$ values, which is computationally intractable for reasonably sized graphs. Thus, REMEDY maintains a belief state based on the *marginal* probability that each node is infected, requiring only $O(|V|)$ values. In the action choice

algorithm, we form an upper bound on the ACTS objective that accounts for the imprecision of the marginal belief state.

The marginal belief update is lines 1–7 and 9–15 of Algorithm 1. At each time step $t \in \{0, \ldots, T-1\}$, we acquire perfect information about the infected state of any I node when it recovers without active screening with probability c. Otherwise its state remains unknown. This naturally recovered node set $C_n(t)$ is given as nodes such that $s(t) = I$ and $s(t+1) = S$.

Let $x_v(t) \in [0, 1]$ be a random variable indicating whether node v is in state I at time t and let $\mathbf{b_v}(t) = [1 - x_v(t), x_v(t)]^\top$ be the marginal belief vector. For each node, we update an intermediate belief state $\bar{\mathbf{b}}_\mathbf{v}(t) = [1 - \bar{x}_v(t), \bar{x}_v(t)]^\top$ in which $\bar{x}_v(t) = 1$ for $v \in C_n(t)$ and $\bar{x}_v(t) = \frac{(1-c)x_v(t)}{(1-x_v(t))+(1-c)x_v(t)}$ for the remaining nodes $v \in V \setminus C_n(t)$. These update steps are in lines 1–7 of Algorithm 1. This intermediate belief state is then exploited by the action choice subroutine to select $C_a(t)$, the node set we actively cure (line 8). After that, we calculate the marginal belief state of the next time step as $\mathbf{b_v}(t+1) = \mathbf{B}_v^N(t)\bar{\mathbf{b}}_\mathbf{v}(t)$ and $\mathbf{b_v}(t+1) = \mathbf{B}_v^A(t)\bar{\mathbf{b}}_\mathbf{v}(t)$ for $v \in V \setminus (C_n(t) \cup C_a(t))$ and $v \in C_n(t) \cup C_a(t)$ respectively where

$$\mathbf{B}_v^N(t) = \begin{array}{c} \\ S \\ I \end{array}\!\!\begin{array}{c} \begin{array}{cc} S & I \end{array} \\ \begin{bmatrix} 1-p_v & 0 \\ p_v & 1 \end{bmatrix} \end{array}, \mathbf{B}_v^A(t) = \begin{array}{c} \\ S \\ I \end{array}\!\!\begin{array}{c} \begin{array}{cc} S & I \end{array} \\ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \end{array} \tag{4}$$

and $p_v = 1 - \prod_{u \in \delta(v)}(1 - \alpha\bar{x}_u(t))$. These are shown in lines 9-15 of Algorithm 1. The transition matrix $\mathbf{B}^N$ does not contain parameter c because each node in the I state that did not naturally recover will remain in I state with probability 1. It is worth noting that, intuitively, to update the marginal belief state for node v, one has to calculate the probability of all possible sets of infected neighbors of v. However, we show below in Lemma 1 that the approach adopted by Eq. 4 to calculate p_v yields the exact probability of v becoming infected by its neighbors given it is currently in S, which saves a great amount of computational time.

Lemma 1. *The exact marginal probabilities of $P[s(t+1) = I | s(t) = S]$ can be calculated by p_v without listing the probability associated with each possible set of infected neighbors.*

Proof. The theorem can be proved by induction. For the base case where there is only one neighbor, the probability that node v is infected in the next time step given it is currently in S is $p_{v,1} = \bar{x}_u(1-(1-\alpha)^1)+(1-\bar{x}_u)(1-(1-\alpha)^0) = \alpha\bar{x}_u$. Assume $p_{v,k} = 1 - \prod_{u \in \delta(v)} (1 - \alpha\bar{x}_u^t)$ for $|\delta(v) \leq k|$ is true, for $|\delta(v) = k+1|$, where

Algorithm 1 REMEDY

Input: $\mathbf{A}$, $\mathbf{b}(t)$, α, c, $C_n(t), t, T, k$
Output: $C_a(t)$, $\mathbf{b}(t+1)$

1: **for** $v \in V$ **do**
2: **if** $v \in C_n(t)$ **then**
3: $\bar{\mathbf{b}}_\mathbf{v}(t) \leftarrow [0,1]^\top$
4: **else**
5: $\bar{\mathbf{b}}_\mathbf{v}(t) \leftarrow \frac{[(1-x_v(t),(1-c)x(t)]^\top}{((1-x_v(t))+(1-c)x_v(t))}$
6: **end if**
7: **end for**
8: $C_a(t) \leftarrow$ ActionChoice$(\mathbf{A}, \bar{\mathbf{b}}(t), \alpha, c, C_n(t), t, T, k)$
9: **for** $v \in V$ **do**
10: **if** $v \in V \setminus C_n(t) \cup C_a(t)$ **then**
11: $\mathbf{b}_\mathbf{v}(t+1) \leftarrow \mathbf{B}_v^N(t)\bar{\mathbf{b}}_\mathbf{v}(t)$
12: **else**
13: $\mathbf{b}_\mathbf{v}(t+1) \leftarrow \mathbf{B}_v^A(t)\bar{\mathbf{b}}_\mathbf{v}(t)$
14: **end if**
15: **end for**
16: **return** $C_a(t)$, $\mathbf{b}(t+1)$

w denotes the newly added neighbor, we have:

$$\begin{aligned} p_{v,k+1} &= p_{v,k} + \bar{x}_w\alpha - p_{v,k}\bar{x}_w\alpha \\ &= (1-\bar{x}_w\alpha)(1 - \prod_{u\in\delta(v)\setminus w}(1-\alpha\bar{x}_u)) + \bar{x}_w\alpha \\ &= 1 - \prod_{u\in\delta(v)}(1-\alpha\bar{x}_u) \end{aligned}$$

Thus we proved that p_v evaluates the exact probability of $P[s_v(t+1) = I | s_v(t) = S]$.

4.2 Action Choice Algorithm

We now turn our attention to selecting the set of nodes to actively screen, i.e., line 8 in Algorithm 1. One fast yet naive approach to this problem is to select the node set with maximum marginal belief to be in I state. This approach can be in $O(|V|)$, but it does not take the network structure and future infection probabilities into account. Another approach is to choose nodes to cure in order to minimize the largest eigenvalue of the network that results from deleting or permanently actively screen the same set of nodes [20]. This approach guarantees that the infection is eradicated in the long term if the largest eigenvalue can be reduced below $\frac{c}{\alpha}$ for sufficient budget k. However, it does not take the belief state into account, nor does it consider how many people become infected before the disease is eradicated. These methods are examined in the experiments as

our baselines. We seek to do better by minimizing an upper bound of the ACTS objective directly.

We develop two different algorithms for action choice: Full-ActionChoice that looks ahead through all future actions and Fast-ActionChoice, a less computationally intensive variant, that considers only the current action, allowing it to exploit eigenvalue decomposition. We refer to REMEDY with Full-ActionChoice as Full-REMEDY and REMEDY with Fast-ActionChoice as Fast-REMEDY. Noted that in both Full-REMEDY and Fast-REMEDY, we change the action based on the observation $C_n(t)$ in each round.

We begin by deriving an upper bound for the ACTS objective starting with some preliminary notation. To encapsulate the effect of active-screening toward our objective function, we define the $|V| \times |V|$ diagonal action matrix $\mathbf{R}_a(t)$ at time t as $\mathbf{R}_a(t)_{v,v} = 1$ if and only if $v \in C_a(t)$ and 0 otherwise. For the current round, say t_0, we observe the nodes that are cured and need to decide the nodes to actively screen. We define the *naturally cured matrix* $\mathbf{R}_n(t_0)$ as $\mathbf{R}_n(t_0)_{v,v} = 1$ if and only if $v \in C_n(t_0)$, which encapsulates the knowledge we gain from natural recovery in the current round. Let vector $\mathbf{x}(t)$ represent $x_v(t)$ for all v. To bound $\mathbf{x}(t)$ across all time steps given the actions we take, let $\mathbf{M}' = \alpha\mathbf{A} + \mathbf{I}$, where $\mathbf{A}$ is the adjacency matrix and $\mathbf{I}$ is the identity matrix, define the *upper bound transition matrix* for the current round as $\mathbf{M}_a(t_0) = (\mathbf{I} - \mathbf{R}_a(t_0) - \mathbf{R}_n(t_0))\mathbf{M}'$, and as $\mathbf{M}_a(t) = (\mathbf{I} - \mathbf{R}_a(t))\mathbf{M}$ for future rounds $t > t_0$, where $\mathbf{M} = \alpha\mathbf{A} + (1-c)\mathbf{I}$.

Theorem 2. *Let the current time be t_0. $\mathbf{M}_a$ is defined as above for t_0 and $t > t_0$. The ACTS objective (Eq. 3) that the actions affect is bounded above as follows:*

$$\mathbb{E}[\sum_{t=t_0}^{T} \sum_{v \in V} |s_v(t) = I|] \leq F(\mathbf{R}_a(t_0), ..., \mathbf{R}_a(T)), \tag{5}$$

$$\textit{where } F = \mathbb{1}^\top \sum_{t=t_0}^{T} \prod_{\tau=t_0}^{t} \mathbf{M}_a(\tau)\mathbf{x}(t_0) \tag{6}$$

$$\textit{and } \prod_{\tau=t_0}^{t} \mathbf{M}_a(\tau) = \mathbf{M}_a(t)\mathbf{M}_a(t-1)...\mathbf{M}_a(t_0). \tag{7}$$

Proof. Observe that the conditional probability $P[\mathbf{s_v}(t+1) = I | \mathbf{s_v}(t) = S]$ is bounded by

$$P[\mathbf{s_v}(t+1) = I | \mathbf{s_v}(t) = S] \leq 1 - (1-\alpha)^{\sum_{u \in \delta(v)} x_u}.$$

We show this as follows: $P[\mathbf{s_v}(t+1) = S | \mathbf{s_v}(t) = S] = \sum_{m=0}^{|\delta(v)|} p_m (1-\alpha)^m$ where m denotes number of infected neighbors of v and p_m is probability of m neighbors infected. Then, $\sum_{m=0}^{|\delta(v)|} p_m (1-\alpha)^m = \mathbb{E}[(1-\alpha)^m] \geq (1-\alpha)^{\mathbb{E}[m]} = (1-\alpha)^{\sum_{u \in \delta(v)} x_u}$ yields the result by applying Jensen's inequality. We approximate the right hand side with a first-order Taylor series expansion as

$\alpha \sum_{u \in \delta(v)} x_u(t)$, yielding

$$x_v(t+1) \leq (1 - x_v(t))\alpha \sum_{u \in \delta(v)} x_u(t) + x_v(t)(1-c).$$

Using a vector $\mathbf{x}(t)$ to represent $x_v(t)$ for all v, the above yields the following inequality in vector form:

$$\mathbf{x}(t+1) \leq \mathbf{M}\mathbf{x}(t) - diag(\alpha \mathbf{A}\mathbf{x}(t))\mathbf{x}(t), \tag{8}$$

where $\mathbf{M} = \alpha\mathbf{A} + (1-c)\mathbf{I}$ and $\mathbf{A}$ is the adjacency matrix. We drop the negative term $\text{diag}(\alpha\mathbf{A}\mathbf{x}(t))\mathbf{x}(t)$, and use $\mathbf{M}\mathbf{x}(t)$ as an upper bound.

While the above holds without intervention, we need to use the transition matrix with given interventions $R_a(t_0), \ldots, R_a(T)$ and knowledge of $C_n(t_0)$. This matrix is precisely $M_a(t_0)$ for the current time t_0 and $M_a(t)$ for $t > t_0$. Then, $\mathbf{x}(t+1) \leq \mathbf{M}_a(t)\mathbf{x}(t)$ for all $t \geq t_0$ and $\mathbb{E}[\sum_{t=t_0}^{T} \sum_{v \in V} |s_v(t) = I|] = \sum_{t=t_0}^{T} \mathbb{1}^\top \mathbf{x}(t)$ yields the desired result.

Given that the function F upper bounds our objective function, we next describe the method we use to select the action matrix $R_a(t)$ that minimize F for every time step. Distinct from previous literature, our objective takes into account the number of infected nodes at each time step. We also have the flexibility to change the action we take based on the observation we make in each round. Such flexibility results in an even larger solution space, $\binom{n}{k}^T$. Since our problem is NP-hard, we apply a Frank-Wolfe style method to attempt to optimize F [11].

Frank-Wolfe is a gradient-base algorithm runs for some number L steps by starting with some arbitrary feasible points and updates it in two steps: (i)computes the gradient of the objective at the current point (ii)find the point which optimizes the gradient over the feasible set and step toward it. For the first step, we need the gradient of F w.r.t. the available action choices. We relax our optimization to a continuous problem by allowing $\mathbf{R}_a(t)_{v,v}$ to take real values between 0 and 1 (instead of binary $0, 1$), which can be interpret as the probability of choosing node v. As a consequence, the feasible solution space is the convex hull of the binary $\mathbf{R}_a(t)$. We denote this convex hull Ψ. By taking the derivative of F, the gradient w.r.t. action at each time t is

$$\frac{\partial F}{\partial \mathbf{R}_a(t)} = -\sum_{t'=t+1}^{T} \prod_{\tau=t'}^{t+1} \mathbf{M}_a^\top(\tau) \mathbb{1}\mathbf{x}^\top(t_0) \prod_{\tau=t-1}^{t_0} \mathbf{M}_a^\top(\tau), \tag{9}$$

The above gradient is a matrix $\Delta(t)$, where the diagonal elements $\Delta(t)_{v,v}$ represents the gradient w.r.t. choice of node v to actively screen at time t. Given the gradient and current information, an approximately optimal action for all times in the continuous relaxation can be obtained through a projected gradient descent or a Frank-Wolfe style algorithm, yielding Full-ActionChoice (Algorithm 2).

Algorithm 2 FULL-ACTIONCHOICE

Input: $\mathbf{A}$, $\bar{\mathbf{b}}(t_0)$, α, c, T, t_0, k
Output: $C_a(t_0)$
1: $\mathbf{R}_a^0(t) \leftarrow \mathbf{0} \quad \forall t$
2: **for** $l = 1...L$ **do**
3: **for** $t = t_0...T$ **do**
4: $\boldsymbol{\Delta}(t) \leftarrow$ GRADIENTORACLE($\mathbf{R}_a^{l-1}$)
5: $\mathbf{R}_a^*(t) \leftarrow$ PROJECTFEASIBLE(Δ, k)
6: $\mathbf{R}_a^l(t) \leftarrow \gamma_l \mathbf{R}_a^{l-1}(t) + (1-\gamma_l)\mathbf{R}_a^*(t)$
7: **end for**
8: **end for**
9: $C_a(t_0) \leftarrow \arg\max_k \mathbf{R}_a^L(t_0)$
10: **return** $C_a(t_0)$

In the FULL-ACTIONCHOICE algorithm, we first set an arbitrary feasible point in Ψ for each time step, say $\mathbf{R}_a^0(t) = \mathbf{0}$ for iteration 0. In each iteration, first we calculate the gradient for the current feasible point using Eq. 9 as our GRADIENTORACLE in line 4. Second, we project the resultant point toward the current best solution on the gradient hyper-plane for every time step simultaneously. We do so simply by greedily selecting k nodes with largest $\Delta(t)_{v,v}$ as our current best solution $\mathbf{R}_a^*(t)$ in line 5. Third, we set the initial point $\mathbf{R}_a^l(t)$ of the next iteration in line 6, in which $\gamma_l = 2/(l+2)$ is the step size of Frank-Wolfe algorithm. Since Ψ is convex and $\mathbf{R}_a^l(t)$ is the convex combination of two feasible points, it is guaranteed that it will remain in the convex hull Ψ by the update. After the iteration finishes, we output our action in the current round by greedily selecting k nodes of the relaxed $\mathbf{R}_a^L(t_0)$ of the final iteration and wait for new information from the next round to arrive.

The FULL-REMEDY algorithm considers future actions simultaneously and has time complexity of $O(T^2|V|^\omega)$, where ω an exponent arising from matrix multiplication. However, calculating such solutions for a very large network—which is often the case for contact tracing—could be time consuming for low resource regions. To reduce time complexity, we further simplify the upper-bound function by assuming that no actions are taken in the future rounds and ignore their effect on the current decision making to derive FAST-ACTIONCHOICE (Algorithm 3). By ignoring future actions, the action matrix $M_a(t)$ in FULL-REMEDY is simplified to constant M. The contribution of actively screening each node can be written as the following vector form:

$$\mathbb{1}^\top \sum\nolimits_{\tau=0}^{T-t_0-1} \mathbf{M}^\tau diag(\mathbf{M}_n \mathbf{x}(t_0)), \tag{10}$$

where $\mathbf{M}_n = (\mathbf{I} - \mathbf{R}_n(t_0))\mathbf{M}'$. Now, since $\mathbf{M}$ is the same for every future round, $\mathbf{M}$ can be decomposed as $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$ ahead of time, where $\mathbf{Q}$ is a matrix comprised of the eigenvectors of $\mathbf{M}$, and $\boldsymbol{\Lambda}$ a diagonal matrix comprised of the eigenvalues along the diagonal. Such a matrix can be approximated by calculating only the top m largest eigenvalues and their eigenvectors using the Lanczos algorithm

Algorithm 3 FAST-ACTIONCHOICE

Input: $\mathbf{A}$, $\bar{\mathbf{b}}(t_0)$, α, c, T, t_0, k
Output: $C_a(t_0)$
1: **if** $t_0 = 0$ **then**
2: $\quad \mathbf{M} \leftarrow \alpha\mathbf{A} + (1-c)\mathbf{I}$
3: $\quad \mathbf{Q}_m, \mathbf{\Lambda}_m \leftarrow$ LANCZOS$(\mathbf{M}, m)$
4: **end if**
5: **Scores** $\leftarrow \mathbb{1}^\top \mathbf{Q}_m (\sum_{\tau=0}^{T-t_0-1} \mathbf{\Lambda}_m^\tau) \mathbf{Q}_m^\top diag(\mathbf{M}_n \mathbf{x}(t_0))$
6: $C_a(t) \leftarrow k$ nodes with highest scores in vector **Scores**

([17]) that has a complexity of $O(|E|)$ (assuming the large network is sparse), yielding the FAST-ACTIONCHOICE shown in Algorithm 3. The approximate $\mathbf{M}$ is given by $\mathbf{Q}_m\mathbf{\Lambda}_m\mathbf{Q}_m^\top$, where these matrices are computed in line 3. In line 5, the well-known result $(\mathbf{Q}_m\mathbf{\Lambda}_m\mathbf{Q}_m^\top)^\tau = \mathbf{Q}_m\mathbf{\Lambda}_m^\tau\mathbf{Q}_m^\top$ is used to approximate $\mathbf{M}^\tau$. The time complexity of FAST-REMEDY is $O(|V|^2)$ assuming constant m.

5 Experiments

We perform experiments comparing FAST- and FULL-REMEDY to baselines on a variety of real-world, publicly available datasets. Table 1 lists all the networks and their properties. Most of the networks were collected in actual human contact settings. The networks used here have varied sizes ($|V|$), average degrees (d), assortativities (ρ_D), and epidemic thresholds ($1/\lambda_A$).

Network	$\lvert V\rvert$	$\frac{1}{\lambda_A^*}$	d	ρ_D	$\alpha = 0.1, c = 0.1$					
					Random	Max-Degree	Eigenvalue	Max-Belief	Fast-REMEDY	Full-REMEDY
Hospital [27]	75	0.027	15.19	-0.18	144	150	151	150	**156**	**160**
India [4]	202	0.095	3.43	0.02	605	470	420	636	**890**	**901**
Face-to-face [14]	410	0.042	6.74	0.23	809	843	745	1057	**1297**	**1409**
Flu [24]	788	0.003	150.12	0.05	1336	1421	1431	1438	**1443**	**1446**
Irvine [19]	1899	0.021	7.29	-0.18	4630	5741	3692	4957	**6676**	**7821**
Escorts [21]	16730	0.032	2.33	-0.03	27400	30167	TLE	29493	**46549**	TLE

Table 1. Improvement of the objective function over **None** (The larger the better). TLE implies time limit of 24 hours for all rounds exceeded.

Setting. In all simulations, we assume the budget k allows for screening and treatment of 20% of the total population $|V| = n$ per round. All results are averages over 30 simulation runs.

Setup. In the real world, active screening is performed only after conducting initial surveys on the prevalence and incidence of the disease. To simulate this, we run our experiments in two stages.

Stage 1 (Survey Stage). This stage starts at $t = 0$ with 25% of individuals in I and ends at $t = 10$. No active screening is done and the disease evolves

naturally. The initial belief $\mathbf{b}(\mathbf{0})$ for all nodes is assumed to be $[0.5, 0.5]^\top$ since we have no prior information. However beliefs are updated when individuals come to the clinic voluntarily (with probability c). This belief update requires knowledge of α and c. There is a rich literature of how to estimate the disease parameters (α and c) in this stage and these methods have been tested on real world scenarios [16, 22, 9]. Here, we assume that such parameters are known.

Such parameters can vary from disease to disease. For example, the transmission rate of Pertussis can be as high as 0.47 for certain age groups in [13] and as low as 0.035 for Syphilis [22]. The cure rate also depends on how resourceful are the target regions. We fix the value of our parameter to a reasonable value $(\alpha, c) = (0.1, 0.1)$ first for comparison and evaluate a wide range of values afterwards.

Stage 2 (ACTS Stage). Here, we consider various screening algorithms. We perform active screening from $t = 11$ to $t = T = 20$ to represent 5 years of time (each round is 6 months [7]). Beliefs are updated according to the belief update scheme presented in Section 4.1.

5.1 Metrics

In Table 1, we compare the outcomes of these screening strategies compared to no intervention (**None**) based on the total infected number over time. In **None**, the evolution of the health states is based on disease dynamics with no active screening for all T timesteps.

Comparison with Baselines. Given the lack of previous algorithms for our problem setting, we measure the performance of REMEDY against baselines:

(1a) **Random**: Randomly select nodes for active screening.
(1b) **MaxDegree**: Greedily choose nodes of the largest degree until the budget is reached.
(1c) **Eigen**: Greedily choose nodes that reduce the largest eigenvalue of A the most until the budget is reached.
(1d) **MaxBelief**: Greedily choose nodes with maximum probability of being in the I state.

We test these algorithms on the following realistic contact networks collected from a diversity of sources and methods. The networks are carefully selected to have rich variety of densities, structures and sizes (ranging from 75 to 16730 nodes).

(2a) **Hospital** [27]: A contact network collected in a university hospital in order to study path of disease spread.
(2b) **India** [4]: A human contact network collected from a rural village in India where active screening with limited budget may take place.
(2c) **Face-to-face** [14]: A network describing face-to-face contact in which influenza might spread through the close contact of individuals.

(2d) **Flu** [24]: This network captures close proximity interactions in an American high school. The network is highly-dense ($\lambda_A > 300$) with small-world properties and a relatively homogeneous distribution.
(2e) **Irvine** [19]: A friendship network, which is a representative network used to study rumors modeled as epidemic spread.
(2f) **Escort** [21]: A large sexual contacts between escorts and sex buyers collected for a six-year period, in which STD may be spread.

In Table 1 higher numbers indicate a larger improvement against **None**. In most cases, both versions of REMEDY make substantial improvements over all baselines, and as expected Full-REMEDY has better solution quality than Fast-REMEDY. A typical number of infected node number over time result is shown in Fig. 2. In the **Hospital** and **Flu** network, due to the network size or homogeneity, respectively, it is difficult for any algorithm to provide any improvement from a random intervention. In the **Escort** network, which is the largest network analyzed, we assume that the size of the network makes any algorithm slower than $O(|V|^2)$ impractical. However, Fast-REMEDY continues to perform better than the baselines that could be run. We examined the performance of REMEDY algorithm on a variety range of α and c. Fast- and Full-REMEDY continue to perform better than their closest competitor. (see Fig. 3, which shows selected scenarios).

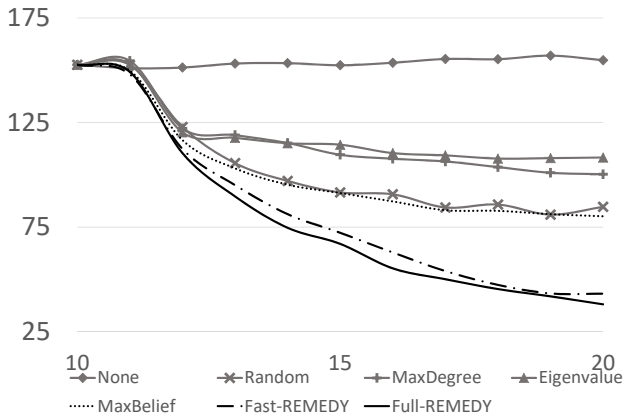
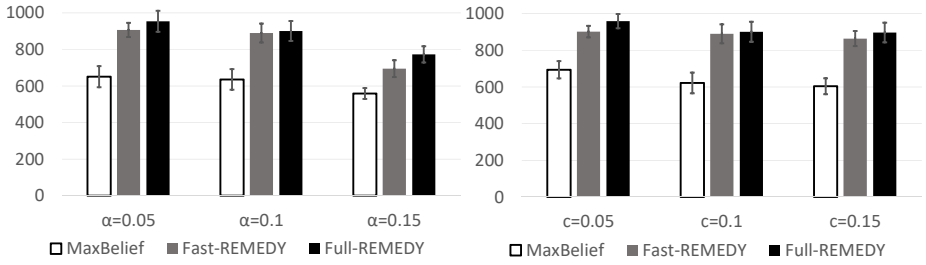


Fig. 2. Number of infected nodes vs. time in **India** network.

5.2 Impact of Budget

Determining the improvement an intervention can achieve with various budgets is critical when informing health policy. We therefore find the improvement pos-



(a) c=0.1 (b) a=0.1

Fig. 3. Improvement over **None** under different parameter sets for **India** network. Both FULL-REMEDY and FAST-REMEDY outperform MaxBelief. FULL-REMEDY outperforms FAST-REMEDY especially for larger α.

sible over different budget values for two realistically modeled diseases: influenza and syphilis.

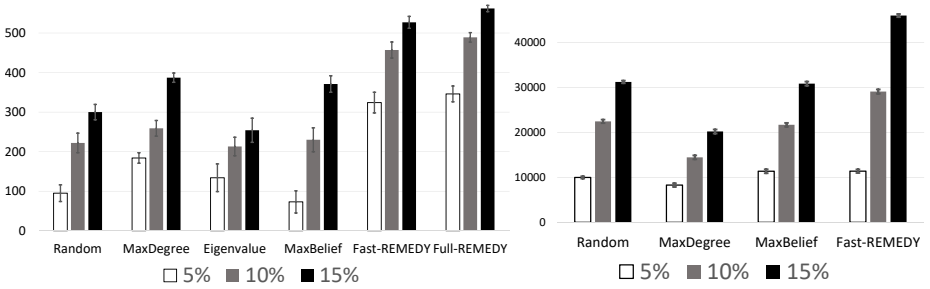
Influenza. For influenza, we use the parameters that previous literature estimated through a continuous survey administered in a student residence hall community [9]. The transition rate is estimated to be $\alpha = 0.024$ per infectious neighbor and the self cure rate is estimated to be $c = 0.3$. We test the algorithms on the **Face-to-face** network, since this network is used to study the dynamics of SIS-type epidemic spread in its original paper [14].

Fig. 4 (a) shows that both FAST-REMEDY and FULL-REMEDY outperform other baselines under realistic settings. The difference between algorithms grows larger as the budget increases. According to Prakash et al. [20], such a network requires at least $k/n \geq \alpha\lambda_A = 57\%$ for random screening to fully eradicate the disease. However, the epidemic dies out at the end of 20th round when FULL-REMEDY is deployed with a budge of only $k/n = 15\%$.

Syphilis. Saad-Roy et al. [22] provides empirically-derived syphilis parameters for our model. The natural cure rate is estimated to be $c = 0.01$ and transmission rate $\alpha = 0.035$. The network used here is the **Escort** network with 16730 nodes, which is a STD contact network. Because the network is large, we show only the algorithms that does not exceed running time due to time complexity, which are Random, Max-Degree, Max-Belief and FAST-REMEDY. Fig.4 (b) shows that FAST-REMEDY achieves significantly better results than all other baselines. On average, it saves 1140, 2900, and 4600 people from becoming infected every six months for 5%, 10% and 15% budgets, respectively.

6 Conclusion

We proposed a novel active screening model (ACTS) to facilitate multi-round active screening problem of SIS recurrent diseases with network structure. We



(a) Influenza (b) STD

Fig. 4. Improvement over **None** under specific disease parameter of different budget (5%, 10%, 15% of total population respectively).

introduced two algorithms, Full-REMEDY and Fast-REMEDY with solution quality and time complexity trade-off and tested on various realistic disease to show their effectiveness.

References

1. Anderson, R.M., May, R.M.: Infectious diseases of humans: dynamics and control. Oxford University Press (1992)
2. Bailey, N.T.: The mathematical theory of infectious diseases and its applications. Charles Griffin & Company Ltd (1975)
3. Ball, F.G., Knock, E.S., ONeill, P.D.: Stochastic epidemic models featuring contact tracing with delays. Math Biosci **266** (2015)
4. Banerjee, A., Chandrasekhar, A.G., Duflo, E., Jackson, M.O.: The diffusion of microfinance. Science **341** (2013)
5. Braxton, J., Davis, D.W., Emerson, B., Flagg, E.W., Grey, J., Grier, L., Harvey, A., Kidd, S., Kim, J., Kreisel, K., et al.: Sexually transmitted disease surveillance 2016. CDC (2017)
6. Cadman, D., Chambers, L., Feldman, W., Sackett, D.: Assessing the Effectiveness of Community Screening Programs. JAMA **251** (1984)
7. CDC: Tuberculosis: General information. MMWR. Recommendations and reports: Morbidity and mortality weekly report. (2011), https://www.cdc.gov/tb/publications/factsheets/general/tb.pdf
8. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD. ACM (2009)
9. Dong, W., Heller, K., Pentland, A.S.: Modeling infection with multi-agent dynamics. In: SBP-BRiMS. Springer (2012)
10. Eames, K.T., Keeling, M.J.: Contact tracing and disease control. Proc. R. Soc. Lond., B, Biol. Sci. **270** (2003)
11. Frank, M., Wolfe, P.: An algorithm for quadratic programming. NRL **3**(1-2) (1956)
12. Ganesh, A., Massoulié, L., Towsley, D.: The effect of network topology on the spread of epidemics. In: INFOCOM 2005. 24th Annual Joint Conference of the IEEE. vol. 2. IEEE (2005)

13. Hethcote, H.W.: An age-structured model for pertussis transmission. Math Biosci **145** (1997)
14. Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.F., Van den Broeck, W.: What's in a crowd? analysis of face-to-face behavioral networks. J. Theor. Biol. **271** (2011)
15. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM (2003)
16. Kirkeby, C., Halasa, T., Gussmann, M., Toft, N., Græsbøll, K.: Methods for estimating disease transmission rates: Evaluating the precision of poisson regression and two novel methods. Sci. Rep. **7** (2017)
17. Lanczos, C.: An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. United States Governm. Press Office Los Angeles, CA (1950)
18. Namukose, E., Bowah, C., Cole, I., et al.: Active Case Finding for Improved Ebola Virus Disease Case Detection in Nimba County, Liberia, 2014/2015: Lessons Learned. Advances in Public Health **2018** (2018)
19. Panzarasa, P., Opsahl, T., Carley, K.M.: Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. J. Assoc. Inf. Sci. Technol. **60** (2009)
20. Prakash, B.A., Chakrabarti, D., Valler, N.C., Faloutsos, M., Faloutsos, C.: Threshold conditions for arbitrary cascade models on arbitrary networks. KAIS **33** (2012)
21. Rocha, L.E., Liljeros, F., Holme, P.: Information dynamics shape the sexual networks of internet-mediated prostitution. Proc. Natl. Acad. Sci. **107** (2010)
22. Saad-Roy, C., Shuai, Z., van den Driessche, P.: A mathematical model of syphilis transmission in an msm population. Math Biosci **277** (2016)
23. Saha, S., Adiga, A., Prakash, B.A., Vullikanti, A.K.S.: Approximation algorithms for reducing the spectral radius to control epidemic spread. In: Proceeding of the 2015 SIAM. SIAM (2015)
24. Salathé, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M.W., Jones, J.H.: A high-resolution human contact network for infectious disease transmission. Proc. Natl. Acad. Sci. **107** (2010)
25. Sun, C., Hsieh, Y.H.: Global analysis of an seir model with varying population size and vaccination. Appl Math Model **34** (2010)
26. Tong, H., Prakash, B.A., Eliassi-Rad, T., Faloutsos, M., Faloutsos, C.: Gelling, and melting, large graphs by edge manipulation. In: Proc. of the 21st ACM CIKM. ACM (2012)
27. Vanhems, P., et al.: Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PloS one **8**, 73970 (2013)
28. Wang, N.: Modeling and analysis of massive social networks. Ph.D. thesis, UMD (2005)
29. Wang, Y., Chakrabarti, D.and Wang, C., Faloutsos, C.: Epidemic spreading in real networks: An eigenvalue viewpoint. In: 22nd SRDS. IEEE (2003)
30. WHO: Systematic screening for active tuberculosis: principles and recommendations. WHO (2013)
31. Zhang, Y., Prakash, B.A.: Data-aware vaccine allocation over large networks. TKDD **10** (2015)