TRACE: Algorithmic ACTS for Preventing the Spread of Recurrent Infectious Diseases on Networks

Biswarup Bhattacharya *, Han Ching Ou *, Arunesh Sinha †,
Sze-Chuan Suen ‡, Bistra Dilkina *, and Milind Tambe *
* *Department of Computer Science, University of Southern California, Los Angeles, CA, USA*
† *Department of Computer Science and Engineering, University of Michigan, Ann Arbor, MI, USA*
‡ *Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, CA, USA*
bbhattac@usc.edu

***Abstract*—An important means of controlling recurrent infectious diseases is through active screening to detect and treat patients. Disease detection on a large network of individuals is a challenging problem, as the health states of individuals are uncertain and the scale of the problem renders traditional dynamic optimization models impractical. Moreover, efficient use of diagnostic and labor resources is a major concern, especially when the disease is prevalent in a resource-constrained region.**

In this paper, we propose a novel active screening model (ACTS) and an algorithm to facilitate active screening for recurrent (no permanent immunity) diseases. Our contributions include: (1) A new approach to modeling SEIS type diseases – diseases with a latent stage and no permanent immunity – using a novel belief-state representation, and (2) a community and eigenvalue-based algorithm (TRACE) to generate an online policy to perform multi-round active screening. We discuss in detail the advantages and disadvantages of existing eigenvalue-based, community-based and greedy approaches towards solving the problem and illustrate the need of developing an algorithmic active screening strategy to achieve better performance scalably. We demonstrate the applicability of TRACE by performing extensive experiments on several real-world publicly available datasets, most of which emulate human contact, with a range of settings to demonstrate applicability to a range of diseases. To the best of our knowledge, this is the first work on developing a multi-round active screening model and active screening algorithm for diseases with a latent stage and no permanent immunity.

***Index Terms*—Public health, SEIS Disease Model, Active Screening, Eigenvalue, Community, Belief states**

I. Introduction

Curable infectious diseases are responsible for millions of deaths every year. Tuberculosis (TB), one such disease, affected over 10 million people worldwide in 2016, and caused over 400,000 deaths in India, the country with the highest TB mortality [1]. While low-cost treatment programs are available, many rely on patients to seek medical care (*passive screening*). However, individuals mistake their symptoms for another condition and not seek care. Public health agencies therefore engage in *active screening*, where individuals in the community are asked to undergo diagnostic tests and are offered treatment if tests return positive results [2].

It is costly to seek out at-risk individuals, and active screening efforts are often limited to high risk groups such as household TB contacts [3]. This method can successfully identify patients [4], and has been extensively evaluated [5]. However, this approach can be challenging to implement widely in resource-constrained regions such as India, as there are large transmission networks of potential patients and the number of health workers is limited. Prior studies show that even when focusing on high-risk TB groups in urban slums in India, the yield can be very small — only 0.8% of screened individuals were diagnosed with TB [3]. With an estimated 1 million undiagnosed TB cases in India, efficient active screening is the need of the hour [3].

Our *first contribution* is a model of the active screening problem (ACTS) which considers the underlying disease dynamics. We focus on recurrent infectious diseases with a latent stage (SEIS model of disease [6]), such as TB. Individuals can be susceptible (S) (currently healthy, but may become exposed), exposed (E), or infected (I). We consider diseases for which there is no means to achieve permanent immunity, either through vaccination or one time infection. As for TB, we assume treatment is effective for both exposed and infected individuals, making the individuals healthy (though again susceptible). Health workers are uncertain about the health state of individuals and have a small budget relative to population size for active screening. To the best of our knowledge, models of multi-round active screening for SEIS diseases are missing in the AI literature.

Our *second contribution* is a novel algorithm—**T**argeted **R**esolution of **A**ctive diseases using **C**ommunities and **E**igenvalues (TRACE)—to guide scalable active screening. In TRACE, we use network community structure to form a community graph, and then we select nodes to screen by maximizing the reduction of the largest eigenvalue of a variant of the community graph. TRACE takes into account the underlying disease dynamics and uncertainty of individuals' health states. TRACE is easily adaptable to most SEIS or SIS type diseases.

We illustrate the benefits of TRACE via extensive testing

on several real-world human contact networks against various baselines across a wide range of disease parameters (which also demonstrates its applicability to various other diseases).

II. Disease Model and Background

We first introduce the disease model notations for our problem. An individual can be in one of the following health states: S (a healthy individual *susceptible* to disease), E (the individual has been *exposed* and has latent disease), or I (the individual is *infected*). We do not consider an explicit recovered or permanent immunity state (R) in our model, as this has been the focus of prior studies [7], [8]. Diseases like Hepatitis A and measles follow a SEIR or SIR pattern where treated individuals may achieve permanent immunity by entering R state. We focus on recurrent diseases, where permanent immunity is not possible (e.g. TB, typhoid), represented by SIS [9] or the more general SEIS [6] disease dynamics.

Disease Model: We adopt a SEIS model [6] for modeling the disease dynamics, given by:

$$\text{Susceptible } (S) \xrightarrow{\alpha} \text{Exposed } (E)$$
$$\text{Exposed } (E) \xrightarrow{\beta} \text{Infected } (I)$$
$$\text{Infected } (I) \xrightarrow{c} \text{Susceptible } (S)$$

In the context of a graph of individuals, α is the edge-wise fixed probability of a susceptible (S) individual (node) being exposed (E) to the disease from an infected (I) neighbor, β is the fixed probability of an exposed (E) individual (node) becoming infected (I), and c is the probability of an infected (I) individual (node) *voluntarily* seeking and successfully completing treatment and returning to the susceptible S stage. We assume that the treatment takes place in one time period, where a period represents the duration needed for a complete treatment regimen ($\sim$half a year for TB).

Prior Approaches for Active Screening: Most previous work on active screening deals primarily with SIR or SEIR type diseases, often referred to as the *Vaccination Problem* [7], [8], [10]–[12], where permanent immunization (entry into R state) can be viewed as removing nodes from the graph [13]–[15]. Exploiting this idea, [14], [15] 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. [11] considers 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 vaccination or screening. However, our paper focuses on multi-round active screening of SEIS diseases, where the complexity increases substantially due to lack of permanent immunity, existence of a latent stage, and uncertainty about the health states of all individuals. To the best of our knowledge, this complex setting has not been attempted previously in the AI literature. Generally, the problem of minimizing disease spread is different from the well-studied problem of influence maximization [16], [17] as well, where one optimizes the selection of seeds or starting nodes for maximizing spread, as opposed to optimizing the selection of nodes on which to intervene in order to minimize spread.

III. The Active Screening (ACTS) Problem

Setup. We define k *ACTS agents* that are to be deployed at every timestep t to diagnose and treat I and E individuals. Individuals are part of a contact network $G(V, E)$, and infection spreads via the edges in the network. There are $|V|$ individuals, and $N(i)$ denotes neighbors of individual i in the network. The network structure (graph) is known from the beginning ($t = 0$). Each individual (node) in the network is in one of the health states $\{S, E, I\}$. Let s_i^t denote the state of individual i at time t. In every round, the *ACTS agents* can either choose to screen a node i (action $a_i = 1$) or not ($a_i = 0$). Only k nodes can be screened in one round. A screened node is observed to be in state S, E, or I, and an unscreened node generates no observation. The *ACTS agents* maintain a belief about the state of every individual, starting with no information at $t = 0$. The beliefs about the health states evolve over time as the agents gain information about individuals (detailed later in this section).

Transition Dynamics. The probability of an individual undergoing a change in health state is given by:

$$T^0 = \begin{array}{c} \\ S \\ E \\ I \end{array} \begin{array}{c} \begin{array}{ccc} S & E & I \end{array} \\ \begin{bmatrix} q_j & 1-q_j & 0 \\ 0 & 1-\beta & \beta \\ c & 0 & 1-c \end{bmatrix} \end{array},$$

$$T^1 = \begin{array}{c} \\ S \\ E \\ I \end{array} \begin{array}{c} \begin{array}{ccc} S & E & I \end{array} \\ \begin{bmatrix} q_j & 1-q_j & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \end{array},$$

$$\text{and } q_j = (1-\alpha)^{|\{k \in N(j) \ | \ s_k^t = I\}|}$$

where, T^0 is the probability matrix for non-screened individuals and T^1 is the probability matrix for screened individuals. The rows denote the state at time t and the columns denote the state at $t + 1$. The transition probabilities follow the disease dynamics described earlier. In particular, q_j captures the probability that node j does not become exposed from his infected neighbors $\{k \in N(j) \mid s_k^t = I\}$. Both I and E individuals who are screened can be treated, but we assume E individuals do not seek treatment voluntarily since their disease is latent unlike I individuals who seek treatment voluntarily with the probability c. For model simplicity, we assume S individuals cannot directly transition directly to I state. This is not an extreme assumption for TB, where the overall duration with latent TB can be much longer than the round length (6 months).

Objective Function. We finally define a Quality Adjusted Life Years (QALY) function $\mathcal{Q}(t)$, associated with full health – a commonly used metric for such intervention studies in public health [18]–[20]. Here, $\mathcal{Q}(t) = \sum_j Q(s_j^t)$, where $Q(s_j^t)$

is defined as follows (values from [18]–[20]).

$$Q(s_j^t) = \begin{cases} +1, & s_j^t = S \\ +1, & s_j^t = E \\ +0.66, & s_j^t = I \end{cases} \quad (1)$$

We focus on maximizing health outcomes in this study and leave cost considerations to future work. We can now define the problem statement.

Problem Statement. *(ACTS Problem) Given a contact network $G(V, E)$, disease transition dynamics T^0, T^1, time horizon T, and budget k, find an active screening policy, i.e. find a set of budget-limited actions $\mathbf{a}$ at each time step, such that $QALY(T) = \sum_{t=0}^{t=T} \mathcal{Q}(t)$ is maximized.*

Belief States. We do not know the true health states of every individual at all times perfectly. We therefore model our belief of node i's health state as $\mathbf{b}_i^t = [b_{i,S}^t, b_{i,E}^t, b_{i,I}^t]$, where $b_{i,j}^t$ is the probability node i is in state j. This marginal representation of health state belief for each node i addresses scalability issues, as representations of the joint distribution of health state beliefs over all nodes can be prohibitively large. We assume marginal beliefs $\mathbf{b}_i^t$'s can be updated independently at each node. Such independence assumptions have been made in prior literature on the spread of contagion [21], [22] and experimentally found to have a minimal effect on outcomes.

Belief Update. We assume perfect observability of the health state s_i^t of any node when it is screened. We cannot observe the health state of a node at time t if we do not screen it at time t. We update the belief for each individual (node) i who voluntarily come to the clinic to an intermediate belief state $\bar{\mathbf{b}}_i^t = [0, 0, 1]$. We also update the beliefs of actively screened individuals to an intermediate belief state $\bar{\mathbf{b}}_i^t \sim s_i^t$. We update the intermediate beliefs of the remaining individuals as:

$$\bar{\mathbf{b}}_i^t = \frac{[b_{i,S}^t, b_{i,E}^t, (1-c)b_{i,I}^t]}{b_{i,S}^t + b_{i,E}^t + (1-c)b_{i,I}^t}$$

For each node i that voluntarily came to a clinic or was actively screened, the final belief update is: $\mathbf{b}_i^{t+1} = [1, 0, 0]$ because the node will be successfully treated and returned to the susceptible state if it was in E or I state. For the remaining nodes, we update to $\mathbf{b}_i^{t+1}$ as follows:

$$\mathbf{b}_i^{t+1} = \bar{\mathbf{b}}_i^t \, \mathbf{\Gamma^t}, \text{ where}$$

$$\mathbf{\Gamma^t} = \begin{bmatrix} w_i^t & 1-w_i^t & 0 \\ 0 & 1-\beta & \beta \\ c & 0 & 1-c \end{bmatrix}, \; w_i^t = \prod_{j \in N(i)} (1 - \alpha \bar{b}_{j,I}^t).$$

This belief update procedure is an important and novel aspect of our proposed ACTS model.

While the *ACTS Problem* can be interpreted as a POMDP, our model is slightly different from standard POMDP models, since in the active screening setting a screening action results in observing the current health states of the individual and not the individual's transitioned state. This difference can be handled straightforwardly, as in [23], [24] using a modified value iteration technique. However, we show in Section VI that known POMDP approaches are not scalable for our problem.

IV. Motivation for TRACE

Given the problem setup, we motivate the need for the TRACE algorithm by showing that many prior approaches or simple extensions do not achieve the desired goal.

We know from [10] that, given a network and limited resources, finding the optimal strategy for vaccinating a limited number of individuals (SIR scenario) and quarantining a limited number of individuals are NP-hard. Also, given a network and limited resources, finding the optimal strategy for placement of a limited number of sensors for monitoring the course of an epidemic is NP-hard [25]. The *ACTS Problem* as defined in Section III is a generalized (harder) case of the above problems where we try to treat infected people without removing them from the graph since there is no permanent immunity and re-infection is possible (SEIS scenario). Computing an individual's probability of infection in SIS networks (SIS is a special version of the SEIS model, where $\beta = 1$) and computing the expected number of infections are also NP-hard [25], [26].

A. Eigenvalue Based Prior Approach

We now consider the circumstances under which diseases or epidemics die out on their own. In the absence of any intervention (action), the system is a discrete non-linear dynamical system. Such systems have been studied in prior work, and the following has been shown:

Proposition 1. *[22] Let λ_A^* denote the largest eigenvalue of the adjacency matrix A of the underlying graph, otherwise known as the spectral radius. Then, the epidemic dies out if and only if*

$$\frac{\alpha}{c} < \frac{1}{\lambda_A^*} \text{ and } \beta \neq 0 \, .$$

Remark: An observation is that the bound on λ_A^* above is same as derived for SIS model (without exposed E state) in earlier work [21]. This is because in the SEIS model, the E state must eventually become I if $\beta \neq 0$; thus, in the long run, E behaves similarly to I when $\beta \neq 0$ and there is no active intervention.

If permanent immunization were possible, immunization can be viewed as removing nodes. Given the result above, one would wish to remove the set of k nodes that reduces the largest eigenvalue the most. This is a NP-complete problem. [14], [15] suggest a heuristic that greedily removes k nodes one at a time, each time selecting the node that maximizes the reduction in the largest eigenvalue. However, in this paper we deal with diseases where permanent immunity is not possible (SEIS). Based on Prop. 1, we also observe that a disease is unlikely to die out on its own in low-resource countries (c is low) with highly contagious diseases (high α), thus necessitating active screening.

B. Budgetary Threshold for Random Intervention

We can gain insight into how uncertainty in individuals' health states affects our problem by examining the fully-naive random screening strategy. We focus on the budget k, the

number of nodes that can be screened and treated in one period. Intuitively, increasing k will lead to faster reduction of disease prevalence with random screening.

Lemma 1. *Assume that we know the infected patients belong to a set I_t in every round t such that $|I_t| \leq m$, where m is an arbitrary constant corresponding to the size of the network. Then, the epidemic dies out using k random interventions every round if $k > m(\lambda_A^* \alpha - c)$.*

Proof. The k random interventions among I_t nodes increase c by at least k/m and α is unchanged. Thus, the disease will die out if $\frac{\alpha}{c+k/m} < 1/\lambda_A^*$. □

Besides providing a threshold for k for which a naive intervention can achieve disease eradication, the above result can be understood as the price of limited information. Lower values of m, meaning more information (better estimate of the true health state), requires fewer random interventions to eradicate the disease. This underscores how uncertainty in the health states is an additional challenge when the number of interventions are limited.

C. DynamicEigen *Procedure*

Prior methods to minimize the largest eigenvalue greedily chose nodes to delete in order to generate a graph with lower maximal eigenvalue. Since we do not know which nodes are infected and can transmit infection with certainty, we augment this method by incorporating uncertainty. To motivate our approach, consider a *hypothetical scenario* where the state of each node is known for sure. We only wish to intervene on infected and exposed nodes, and S nodes do not effect neighboring nodes. Using $A_{i,j} = A_{j,i} = 1$ to represent an edge from i to j in the adjacency matrix A of the input graph, we see that removing all edges from S nodes is same as multiplying the rows and columns of A corresponding to nodes in state S by zero. Then we can greedily choose among I and E nodes (lines 6-7) with the goal of reducing the largest eigenvalue of the adjacency matrix of the directed graph and return nodes that have total weights above the threshold k. Algorithm 1 describes this approach. While our intervention may be undone over time (treated nodes can be reinfected), repeated screenings may push the system towards lower disease prevalence.

Algorithm 1 Eigen($\mathbf{A}, w, k$)

Input: Adjacency matrix $\mathbf{A}$, function w for weight of each node, min total weight of nodes to remove k

1: $V \leftarrow$ Number of vertex of input graph
2: **for** all $i \in \{1, \ldots, V\}$ **do**
3: $\quad \mathbf{A}' \leftarrow \mathbf{A}$
4: $\quad \mathbf{A}'_{i,:} \leftarrow \mathbf{0}$, $\mathbf{A}'_{:,i} \leftarrow \mathbf{0}$ ▷ Remove i^{th} node
5: $\quad \lambda^i = LargestEigenvalue(\mathbf{A}')$
6: **Sort** nodes $\langle v_1, \ldots, v_V \rangle$ corresponding to increasing λ^i
7: **return** first h nodes such that $\sum_{i=1}^{h} w(v_i) \geq k$

Now let us return to our problem setup, where we do not know the exact state of each node but rather have beliefs about each node. We now propose the DynamicEigen procedure, which is shown in Algorithm 2. A natural extension of the hypothetical scenario above is to multiply the row of a node i in the adjacency matrix A by $1 - b_{i,S}^t$, the belief probability it is E or I (line 3). This is a softer version of making the row of all S nodes all zeros.

Algorithm 2 DynamicEigen($\mathbf{A}, \mathbf{b}, w, k$)

Input: Adjacency matrix $\mathbf{A}$, belief $\mathbf{b^t}$, function w for weight of each node, min total weight of nodes to remove k

1: $V \leftarrow$ Number of vertex of input graph
2: **for** all $i \in \{1, \ldots, V\}$ **do**
3: $\quad \mathbf{A}_{i,:} = \mathbf{A}_{i,:} * (1 - b_{i,S}^t)$ ▷ Multiply i^{th} row
4: **return** Eigen($\mathbf{A}, w, k$)

However, we see that Algorithm 2 is computationally expensive and may not scale up to large graphs.

Lemma 2. *The time complexity of the* DynamicEigen *algorithm is $O(|V|^3)$, where $|V|$ is the total number of nodes.*

Proof. Finding the largest eigenvalue of a $|V| \times |V|$ matrix takes $O(|V|^2)$ time. Since we are looping $|V|$ times to find the top k nodes to screen, the overall time complexity is $O(|V|^3)$. Sorting the eigenvalues to get the top k values takes $O(|V| \log |V|)$ time, but the overall time complexity remains $O(|V|^3)$. □

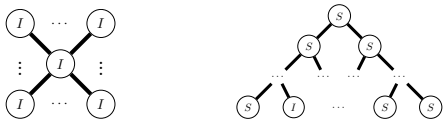
It was shown in [27] that the Eigen (or DynamicEigen) procedure can be sped up by instead finding the k nodes with the highest degree (or highest degree $\times$ $(1 - b_{i,S}^t)$) which takes $O(|V|^2)$ time. However, we do not use this approximation, since it has merit only in the case of high assortativity (ρ_D). For example, the actual human contact networks we considered in Section VI do not necessarily have high assortativity.

D. Eigenvalue and Greedy

A simple and fast alternative to the eigenvalue approach could be to select k nodes with the highest probability of being infected ($b_{i,I}$) at every time step (denoted further as *Greedy*). Unfortunately, both the eigenvalue method and Greedy method have shortcomings in our dynamic problem. We demonstrate this through some observations for different classes of networks. In all the observations, we use $(\alpha, \beta, c) = (1, 1, 0)$ as an example. Also, for the sake of comparison in these Observations, we assume that the probability of a node not being healthy is known with some degree of certainty.

Observation 1. *There exists a class of graphs where the Greedy method with a budget of $k = 2$ requires an expected $O(|V|!)$ rounds to completely eradicate the disease whereas an eigenvalue-based method can eradicate the disease in an expected $O(|V|^2)$ rounds with the same budget.*

Justification. Consider a star graph (Figure 1a), where all the nodes are initially in I state. With a budget of 2, the eigenvalue method will choose the star center and one arbitrary node among non-central nodes to treat in every round. The disease will thus die out in an expected $\sum_{i=1}^{|V|-1} \frac{|V|-1}{i} \sim O(|V|^2)$ rounds. On the other hand, the Max Belief method will choose k nodes randomly among the nodes in state I. If the center node is not picked in every two rounds ($S \xrightarrow{1 \text{ round}} E \xrightarrow{1 \text{ round}} I$) before the disease dies out, the center will become infected, and after two more rounds the non-central nodes will be I except $2k$ nodes which can be either in S, E or I state (we ignore this w.l.o.g.). The probability of the center node being chosen every second round (because it takes two rounds to move from S to I state) is $\frac{k}{|I|}$ where $|I|$ is the total number of infected nodes in the round with the center being in I state. The probability of the center node being chosen every second round until the disease dies out is $\prod_{i=0}^{\frac{|V|}{2k-1}-1} \frac{k}{|V|-(2k-1)i}$. This gives the desired result. □



(a) (b)

Fig. 1: Comparing Eigenvalue and Greedy

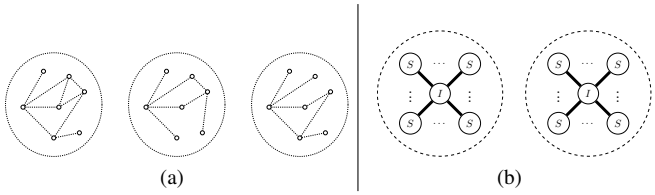
Observation 2. *There exists a class of graphs where an eigenvalue-based method can never eradicate the disease with a budget of $k < \frac{|V|}{2}$ whereas the Greedy method can eradicate the disease in one round with a budget of $k \sim O(1)$.*

Justification. Consider a binary tree (Figure 1b), with $\Theta(k)$ leaf nodes in I state and others in S state. An eigenvalue-based method chooses the nodes that equally partitions the graph, and thus in this case it will start choosing from the root and go down the tree in breadth-first order, and reach the leaf nodes only after it has chosen all the $\frac{|V|-1}{2}$ parent nodes. Greedy however can eradicate the disease in the first round by simply choosing k nodes which have the highest probability of being in I state, which are the infected leafs.

E. Utilizing Communities

Infectious diseases such as TB are transmitted via close contact with an infected person, usually within communities [28]. Curing whole communities may potentially be an efficient way to reduce infection (can be interpreted as *graph shattering* [10]), since infection propagation is stopped for large sections of the graph. Also in our case, given the lack of additional information about the network like patient attributes, it is natural to utilize this approach. We also note that forming communities might enable us to reduce the largest eigenvalue, i.e. apply Algorithm 2, in a scalable fashion.

However, we show in the following Observations that using communities alone can be both better or worse than Greedy or eigenvalue based approaches for different classes of graphs, further motivating the need for our algorithm, TRACE, which identifies communities in addition to considering beliefs and reducing the largest eigenvalue. The exact method of achieving scalability using communities is elucidated in the next section.



(a) (b)

Fig. 2: Comparing Eigenvalue, Community and Greedy

Observation 3. *There exists a class of graphs where an eigenvalue-based method can never eradicate the disease with a budget of k and the Greedy method requires an expected $O(|V|^k)$ rounds to completely eradicate the disease, but a community-based method can eradicate the disease in an expected $O(|V|^2)$ rounds.*

Justification. Let us consider a graph where there exists M disjoint clusters (Figure 2a), each of size less than or equal to the budget k with $k << M$, where M is the number of communities. All the nodes are in I state and are arranged in each cluster such that the the top k nodes, removal of which causes the most decrease in the largest eigenvalue, all lie in different clusters. In such graphs, it is evident that community-based algorithms can cure one community at a time and can achieve full eradication after an expected $M^2 \sim (|V|/k)^2 \sim O(|V|^2)$ rounds because a cured community cannot infect other communities. However, an eigenvalue-based technique may not choose communities as a whole and therefore, an eradication cannot be guaranteed unless the budget is increased to $|V|$ which is equal to the size of the graph. Similarly, the Greedy method may not choose communities as a whole and therefore takes an expected $\binom{|V|-1}{k-1}$ rounds to cure the first community, $\binom{|V|-k-1}{k-1}$ rounds to cure the second community, and so on, thus taking approximately $O(|V|^k)$ rounds to cure all the infected nodes.

Observation 4. *There exists a class of graphs where a community-based method can never eradicate the disease whereas the Greedy or eigenvalue-based method either can eradicate the disease in one round with a budget of k.*

Justification. Consider M disconnected star graphs (Figure 2b), where $M-1$ stars are of size less than k and one star is of size k, and $k \leq M$. All the center nodes of the stars are in I state, and all the other nodes are in S state. With a budget of k, community-based algorithms will keep choosing the same star with k nodes thus never eradicating the disease. However, either the Greedy or eigenvalue-based method can directly choose the k center nodes in the first round and completely eradicate the disease in one shot.

V. TRACE Algorithm for Active Screening

We introduce a structured algorithm to generate an online policy—Targeted Resolution of Active diseases using Communities and Eigenvalues (TRACE)—that combines elements of the three approaches (Greedy, and eigenvalue based, and

community based methods) to identify k individuals to actively screen at every time-step. The complete TRACE algorithm is shown in Algorithm 3. First, we discuss the two step process of community formation (lines 1–4 in Algorithm 3).

Node Type Estimation: We assign an attractiveness score to reflect the effectiveness of intervening on the node. If we knew the true health state of every node, then we would intervene only on the infected nodes as only these nodes spread infection. However, in the absence of such precise information, at every time-step the nodes are sorted according to a measure of possible benefit, defined as $R_i^t = \sigma b_{i,E}^t + b_{i,I}^t$ for each node i (line 2), where σ is an arbitrary parameter that controls the relative importance of E nodes relative to I nodes. The nodes with the highest one-third of R^t values are labeled g_1 (group 1), the next one-third to be g_2 (group 2), and the rest to be g_3 (group 3) (line 3).

Super-Node Creation: After labeling all nodes, locally similar nodes (nodes of the same label that share an edge) are clustered into a super-node iteratively. This process generates a set of super-nodes, each of which is labeled as g_1, g_2 or g_3 based on the labeling of its component nodes. There can be multiple super-nodes with the same label in the network. The $size_{\mathbf{u}}$ of a super-node $\mathbf{u}$ is the number of component nodes in the super-node. The weights of edges between nodes in different super-nodes are added to produce new inter-super-node edges. This super-nodes formation uses the known method of *graph coarsening* [29] (line 4). As an example, in Figure 3 we combine the two g_1, two g_2 and three g_3 nodes to form three super nodes with size two (and another with size one). These super-nodes emulate the communities of I, E and S in real-world networks. We refer to the resultant graph of super-nodes as the community graph, where the belief of each node $b_{\mathbf{u},S}^t$ is the average of $b_{v,S}^t$ of all component nodes v in super-node $\mathbf{u}$.

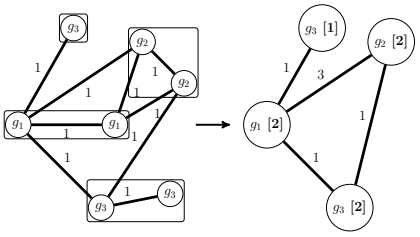


Fig. 3: Example: 4 super-nodes formed from 7 nodes

Next, we call the DYNAMICEIGEN sub-procedure to choose nodes to screen in the weighted community graph using $size$ as weights on each super-node (line 5, where $\overline{\mathbf{A}}$ is the adjacency matrix of the community graph). The procedure returns a set of super-nodes where the total size (weight) is not lower than the budget k. If the total size is higher (line 6), we remove a super-node (line 7), compute left-over budget κ (line 8), modify the original graph by removing all nodes from the left-over super-nodes (line 9), and call the sub-procedure again to select κ nodes from the modified original graph with weights 1 on each node (line 10). It must be noted that our proposed DYNAMICEIGEN procedure is also one of the novel aspects of TRACE.

Algorithm 3 TRACE Algorithm

Input: Adjacency Matrix $\mathbf{A}$ of graph, Belief $\mathbf{b}^t$, Budget k

1: **for** all $i \in \{1, \ldots, n\}$ **do**
2: $\quad R_i^t = \sigma b_{i,E}^t + b_{i,I}^t$
3: **Sort** R^t and label each node as g_1, g_2, or g_3
4: $\overline{\mathbf{A}}, \overline{\mathbf{b}}^t, size \leftarrow Coarsen(\mathbf{A}, g_1, g_2, g_3, \mathbf{b}^t)$
5: $\mathbf{U} \leftarrow$ DYNAMICEIGEN$(\overline{\mathbf{A}}, \overline{\mathbf{b}}^t, size, k)$
6: **if** $\sum_{\mathbf{u}\in\mathbf{U}} size_{\mathbf{u}} > k$ **then**
7: $\quad \mathbf{u}' \leftarrow$ the last selected super-node from $\mathbf{U}$
8: $\quad \kappa = k - \sum_{\mathbf{u}\in\mathbf{U}\setminus\mathbf{u}'} size_{\mathbf{u}}$
9: $\quad \underline{\mathbf{A}}, \underline{\mathbf{b}}^t \leftarrow$ remove all nodes in $\mathbf{U}\setminus\mathbf{u}'$ from $\mathbf{A}, \mathbf{b}^t$
10: $\quad a \leftarrow$ DYNAMICEIGEN$(\underline{\mathbf{A}}, \underline{\mathbf{b}}^t, \mathbf{1}, \kappa)$
11: Active screen nodes $\{v \mid v \in a$ or $v \in \mathbf{u}$ for $\mathbf{u} \in \mathbf{U}\setminus\mathbf{u}'\}$

Now that we have combined community structure with belief states (denoted as *Comm* in Section VI), we compare it to the DYNAMICEIGEN procedure (without super-node formation).

Fig. 4: Comparing *Comm* and DYNAMICEIGEN

Observation 5. *There exists a class of graphs where* DYNAMICEIGEN *without super-nodes can never completely eradicate the disease with a budget of k whereas the Comm algorithm can eradicate the disease in an expected $O(|V|)$ rounds.*

Justification. The justification to this observation closely follows the reasoning in Observation 3 (Figure 4a).

Observation 6. *There exists a class of graphs where the Comm algorithm with a budget of k requires an expected $O((|V| - |V'|)!)$ rounds, to completely eradicate the disease whereas* DYNAMICEIGEN *without super-nodes can eradicate the disease in an expected $O(|V'|)$ rounds with a budget of k, where $|V'|$ is the size of the smaller star.*

Justification. Consider a graph with two stars of different sizes (Figure 4b) where the smaller star is of size $|V'| \geq k$ and the larger star has a size of $|V| - |V'|$. Initially, the center node in the larger star is in state I and the other nodes are in state S. All the nodes in the smaller star are in state I. The dynamic eigenvalue algorithm can eradicate the disease with just a budget of k in an expected $O(|V'|)$ rounds by choosing both the stars' center and then choosing one non-central node and the center, or two non-central nodes in each round based on if the center node is in I state. However, the *Comm* algorithm will cluster the smaller star and cure all of them before choosing the I node in the larger star, where by then all of the nodes in the larger star would have been infected. Based on an analysis similar to Observation 1, we can conclude that the disease will die out in an expected $O((|V| - |V'|)!)$ rounds.

We see that TRACE is considerably faster and scales to larger graphs due to the creation of super-nodes which effectively reduces the graph size.

Theorem 1. *In a scale-free graph, the number of super-nodes formed by the Coarsen algorithm is of the order* $|G| \approx O\left(\frac{3|V|}{2d}\right)$*, where* $|V|$ *is the total number of nodes and* d *is the average degree of the nodes in the original graph.*

Proof. Consider a scale-free graph. The probability that a node i shares an edge with another node j of the same type (say, g_1) is given by

$$r_{i,j} = \sum_i \sum_j \left[1 - \left(1 - \frac{2\widehat{d}_i}{|g_1|\widehat{d}}\right)^{\widehat{d}_j}\right] P(\widehat{d}_i)P(\widehat{d}_j)$$

$$\implies \mathbb{E}[r_{i,j}] = r \approx \frac{2d}{|V|} \qquad (\text{since } \sum_i \widehat{d}_i P(\widehat{d}_i) = \widehat{d})$$

where $\widehat{d} \approx \frac{|g_1|d}{|V|}$, d is average degree of the graph, and $|g_1|$ is the number of nodes of type g_1. When the first node of type g_1 arrives it forms a super-node by itself. When the second node of type g_1 arrives, it can form a super-node by itself with probability $1 - r_{2,1}$ or it can coalesce with the first node with probability $r_{2,1}$ to form one combined super-node. If there are a total of n' nodes which arrive, we can calculate the expected number of groups (super-nodes) g_1 nodes forms as:

$$\mathbb{E}[|G_1|] = 1 + \mathbb{E}[(1 - r_{2,1})] + \ldots + \mathbb{E}\Big[\prod_{i=1}^{n'-1}(1 - r_{n',i})\Big]$$

By Jensen's Inequality, $\mathbb{E}[f(x)] \leq f(\mathbb{E}[x])$ when f is convex. Here, $f(x)$ is of the form $(1-x)^n$ which is a convex function since $x < 1$ in our case. Therefore, we can rewrite the above equation as:

$$\mathbb{E}[|G_1|] \leq 1 + (1 - r) + \ldots + (1 - r)^{n'} = \frac{1 - (1 - r)^{n'}}{r}$$

The above expectation expression is also true for $\mathbb{E}[|G_2|]$ and $\mathbb{E}[|G_3|]$. Also, given the node type estimation process, the number of nodes of each type is equal, thus $n' = |V|/3$. Hence, $\mathbb{E}[|G|] \leq \mathbb{E}[|G_1|] + \mathbb{E}[|G_2|] + \mathbb{E}[|G_3|] = \frac{2d}{3|V|}$. Thus, $|G| \approx O\left(\frac{3|V|}{2d}\right)$. □

Theorem 2. *The time complexity of the TRACE algorithm (Algorithm 3) is* $O(|V|\log|V| + |E| + |G|^3)$*, where* $|V|$ *and* $|E|$ *are the total number of nodes and edges in the original graph respectively, and* $|G|$ *is the total number of super-nodes.*

Proof. The sorting procedure (line 3) takes $O(|V|\log|V|)$ time. The $Coarsen$ procedure (line 4) takes $O(|E|)$ time. By Lemma 2, we know that finding the top k nodes to screen by finding the largest eigenvalues (line 5) takes $O(|G|^3)$ time. □

We know by Theorem 1 that $|G| \approx O\left(\frac{3|V|}{2d}\right)$ for scale-free graphs, thus there is almost a $(1 - \frac{27}{8d^3})$ reduction in average runtime compared to Algorithm 2.

Observation 7. *Suppose the belief states equal the actual health states and* $(\alpha, \beta, c) = (1, 1, 0)$*. Then, TRACE is guaranteed to perform better than or at least as well as its individual components, in terms of both budget and time, in all the classes of graphs discussed in the Observations.*

Justification. For example, in Figure 1a, in case of exact beliefs, it is guaranteed that TRACE will choose the central node since that is the best choice by eigenvalue (all I nodes have equal belief of [0,0,1]) and thereby eradicate the disease in $O(|V|^2)$ rounds with a budget of $k = 2$. Thus, following Algorithm 3, we can similarly show that TRACE will in fact perform at least as well as its individual components in all the discussed classes of graphs (variants of trees, stars, and clusters). We omit the details for brevity. □

Thus, TRACE is able to leverage the advantages of each approach. Although these special graphs do not by themselves represent real-world human contact graphs, real graphs are formed from a combination of these special graphs. Estimating that the belief space representation is a reasonably accurate embedding of the information we do have (there is no mis-information in observations while screening), We hypothesize that TRACE's superior performance in these skeleton graphs can be extended to interpret good performance in realistic graphs as well. This hypothesis is validated via experiments.

VI. Experiments

We consider a variety of real-world, publicly available datasets on which we perform experiments. Table I lists all the networks and their properties. Most of the networks were collected in actual human contact settings. The networks are chosen to have a varied range of size ($|V|$), degree (d), assortativity (ρ_D), and epidemic threshold ($1/\lambda_A^*$).

As previously mentioned in Section III, we first attempt to solve the *ACTS Problem* as a POMDP, by using the state-of-the-art modified POMCP algorithm [23]. We show in Figure 5 that POMCP takes exponential time with increasing $|V|$ and fails to scale up beyond 10 nodes (**India** network) for fixed values of k and T while TRACE is able to generate an online POMDP policy without an exponential increase in runtime. Factored POMDPs [40] and newer algorithms like DESPOT [41] also fail to scale up beyond a few nodes due to memory overflow. All results are averages over 20 simulation runs.

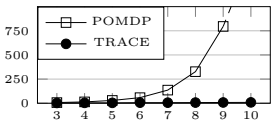


Fig. 5: Runtime (s) v/s Number of nodes ($|V|$); $k = 3, T = 10$

Settings. Next, we analyze TRACE's performance under various α, β, c settings. α, β, c may depend on social contact patterns and biological factors which may vary across populations [42]. We explore a range of these parameters to show disease behavior under a variety of scenarios. Since eradication does not depend on β (by Proposition 1), we vary only α, c and fix $\beta = 0.25$ for the experiments. The passive treatment rate c may vary widely, as it depends on resource availability (clinic

Network	$\|V\|$	$\frac{1}{\lambda_A^*}$	d	ρ_D	$\alpha = 0.1, c = 0.2$ (Milder disease)						$\alpha = 0.3, c = 0.2$ (Strong disease)					
					Random	SE	DE	MB	Comm	TRACE	Random	SE	DE	MB	Comm	TRACE
Hospital [30]	75	0.027	15.19	-0.18	61	29	73	87	108	**131**	50	17	26	67	73	**100**
Office [31]	92	0.050	8.20	-0.05	76	32	98	120	132	**151**	71	27	39	77	87	**108**
Friendship [32]	134	0.102	3.03	0.29	123	41	151	193	201	**228**	114	41	52	129	139	**152**
India [33]	202	0.095	3.43	0.02	145	60	169	216	245	**260**	127	54	107	160	173	**198**
Exhibition [34]	410	0.042	6.74	0.23	212	68	257	326	430	**529**	199	66	184	267	292	**356**
Flu [35]	788	0.003	150.12	0.05	574	82	662	810	1164	**1371**	545	70	492	719	943	**1191**
Irvine [36]	1899	0.021	7.29	-0.18	849	106	1294	1808	2147	**2474**	952	83	869	1464	1783	**2120**
Adolescent [37]	2539	0.076	4.12	0.25	1607	125	2013	2436	2847	**3250**	1331	89	1224	1597	1994	**2641**
Chess [38]	7301	0.015	7.65	0.37	5824	TLE	TLE	7692	9624	**11697**	4403	TLE	TLE	6640	8032	**9904**
Escorts [39]	16730	0.032	2.33	-0.03	13083	TLE	TLE	16565	18630	**21441**	9464	TLE	TLE	12901	14347	**16990**

TABLE I: Increase in $QALY(T = 30)$ over **None** (rounded to the nearest integer). TLE implies time limit of 10000s for all rounds exceeded. All TRACE values are statistically significant ($p < 0.05$).

accessibility, outreach campaigns, etc.). In all simulations, the budget is $k = 5\%$ of the total population, and $\sigma = 0.5$.

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 equal number of S, E, I individuals and ends at $t = 10$. No active screening is done and the disease evolves naturally. The initial belief b^0 for all nodes is assumed to be $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$ since we have no prior information. Beliefs are updated when individuals come to the clinic voluntarily (with probability c).
- Stage 2 (**ACTS Stage**): Here, we consider various screening algorithms. We perform active screening from $t = 11$ to $t = T = 30$ to represent 10 years of time (each round is 6 months [43]). Beliefs are updated according to the belief update scheme presented in Section III.

Metric. We compare the benefit of these screening strategies over and above no intervention (**None**) based on the QALY objective ($QALY(T)$, Table I). 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 first measure the performance of TRACE against simple baselines:

(1a) **Random**: Randomly select nodes for active screening.

(1b) **Static Eigen (SE)**: Choose the nodes using Algorithm 1 (no belief information) on the network (no super-node formation). This baseline uses only the graph structure information.

TRACE provides significant improvement over **None** compared to **SE** and **Random** ($p < 0.05$). The improvement is also practically significant (Cohen's $d > 1$: large effect).

Comparison with individual components. We then show the performance of the three approaches that were combined to form TRACE, illustrating that no single approach is solely responsible for TRACE's performance. TRACE's performance is both statistically and practically significant ($p < 0.05$ and Cohen's $d \sim 0.6$: medium effect) when compared to the three approaches:

(2a) **Dynamic Eigen (DE)**: Choose the nodes using just Algorithm 2 without any super-node formation.

(2b) **Greedy or Max Belief (MB)**: Choose the nodes with the higher *belief* of being infected in that time-step, i.e. $b^t_{i,I}$.

(2c) **Community (Comm)**: Choose the nodes by a 0-1 knapsack algorithm (knapsack weight = budget k) after super-node formation. This does not utilize eigenvalues.

Supernodes. By Theorems 1 and 2, we indicated that the number of supernodes formed is a fraction of the total number of nodes $|V|$. We illustrate this experimentally by showing the average number of supernodes formed per round (Figure 6) for all the networks.

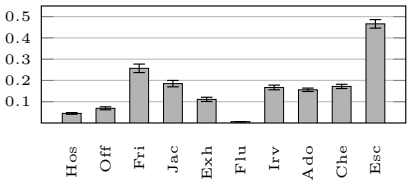
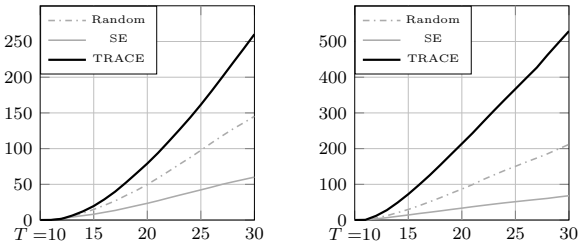


Fig. 6: Average $|G|/|V|$ per round for all the networks, $\alpha = 0.1, \beta = 0.25, c = 0.2$

Discussion. We see from Table I that TRACE not only performs better than other algorithms in all the datasets, it has an even superior performance especially in cases of higher average degree (d), because of better community formation. **Comm**'s performance also follows this pattern due to the same reason. It is also interesting to note that although we expect the QALYs to increase with increasing closeness of the α/c value to the epidemic threshold value ($1/\lambda_A^*$), we see that the values strongly follow the average degree trend. This is because the α/c values are probably too high compared to the $1/\lambda_A^*$ values.

We note that **MB** is a competitive algorithm, thus illustrating that the belief space representation is indeed a possible working embedding of the underlying health states. The performance of **DE** and other screening algorithms worsens with increasing α values due to higher contagiousness compared to available budget while non-belief based baseline **SE** performs poorly as expected since it essentially chooses the same nodes repeatedly.

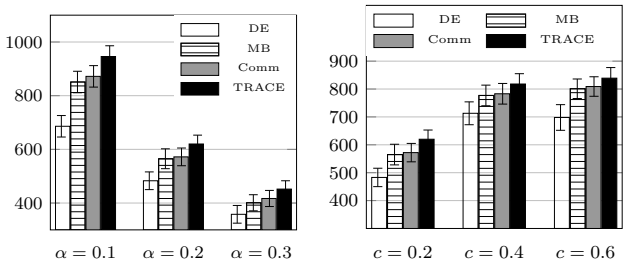
We also see an increasing divergence of TRACE's performance over time (Figure 7 shows this for two representative networks) compared to **Random** and **SE**. This clearly illustrates that the synergy of belief states, eigenvalues and community gives TRACE an advantage from early timesteps.



(a) India network (b) Exhibition network

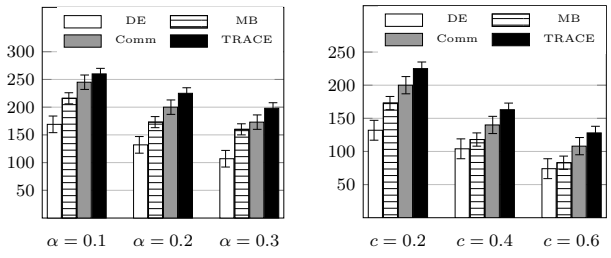
Fig. 7: Increase in $QALY(T)$ over None for varying T ($\alpha = 0.1, \beta = 0.25, c = 0.2$)

Even though our objective function measures QALYs – a widely accepted metric in public health literature – we know that E individuals cannot be classified as "healthy" for the purpose of reducing infection in a network even if they generally do not face a decreased "quality of life". Hence, we experiment by also measuring the increase in the number of healthy (S) individuals over time ($\sum_t^T |S|_t$) over **None**, equivalent to the total number of healthy half life years gained if no active screening was conducted, for varying α and c values (Figure 8). Concurrently, we also note the increase in QALYs for the same α and c values for a representative network – **India** network – a human contact network with 202 nodes, collected from a rural village in India, a setting in which TB active screening may take place (Figure 9).



(a) Varying α ($c = 0.2$) (b) Varying c ($\alpha = 0.2$)

Fig. 8: Increase in $\sum_{t=0}^{t=30} |S|_t$ for TRACE components over None (India network)



(a) Varying α ($c = 0.2$) (b) Varying c ($\alpha = 0.2$)

Fig. 9: Increase in $QALY(T = 30)$ for TRACE components over None (India network)

It is clear that for high α cases, there is not enough budget to estimate and cure the rapidly forming I nodes quickly. However, the most interesting observation is the seemingly opposite performance of the screening algorithms depending on the metric of comparison for the varying c case. This is occurring due to the fact that in high c situations, the network is automatically able to cure a lot of the I nodes by itself, thus the active screening algorithms end up curing more E nodes. While curing E nodes increases the number of S nodes, it does not lead to an increase in the QALY as much (by Eqn. 1). Also, the QALY of **None** itself becomes high due to the higher proportion of I individuals seeking voluntary cure. Similar trends are seen in the other networks – not displayed here due to space constraints.

Further, we analyze the minimum additional budget required to achieve performance comparable to TRACE in Figure 10, revealing the budgetary savings from using TRACE. TRACE with all its components produces significant savings over attempting to use each component alone.

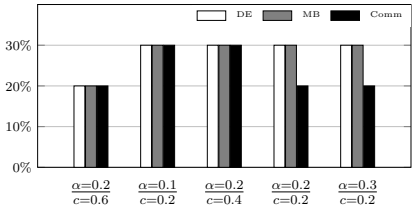


Fig. 10: Minimum extra budget (in %) required for another algorithm to match the performance of TRACE (India network)

Finally, we note that the average village population even in highly populated countries like India is around 2000 [44], much less than the size of our largest tested dataset. This indicates potential for field testing.

VII. Conclusion

We proposed a novel active screening model (ACTS) and an algorithm (TRACE) to facilitate multi-round active screening for recurrent diseases. Unlike existing works in AI literature, the ACTS model incorporates uncertainty of health states as well as the SEIS disease complexities of no permanent cure and a latent stage. TRACE performs significantly better, in a scalable fashion, than the baselines and each of its components individually in a variety of real-world inspired settings.

Future directions include incorporating more complex disease models (e.g. including maternal immunity, carrier states etc.), including birth and death processes, and introducing patient heterogeneity (age, gender, medical history and other features) and costs of treatment and screening into the model.

References

[1] WHO, "Global tuberculosis report 2017," 2017.

[2] K. Kranzer, H. Afnan-Holmes, K. Tomlin, J. E. Golub, A. Shapiro, A. Schaap, E. Corbett, K. Lönnroth, and J. Glynn, "The benefits to communities and individuals of screening for active tuberculosis disease: a systematic review [state of the art series. case finding/screening. number 2 in the series]," *The international journal of tuberculosis and lung disease*, vol. 17, no. 4, pp. 432–446, 2013.

[3] P. Chinnakali, P. Thekkur, G. Ramaswamy, K. Selvaraj *et al.*, "Active screening for tuberculosis among slum dwellers in selected urban slums of Puducherry, South India," *Annals of Tropical Medicine and Public Health*, vol. 9, no. 4, p. 295, 2016.

[4] N. T. C. Association *et al.*, "Guidelines for the investigation of contacts of persons with infectious tuberculosis. recommendations from the national tuberculosis controllers association and cdc." *MMWR. Recommendations and reports: Morbidity and mortality weekly report. Recommendations and reports*, vol. 54, no. RR-15, p. 1, 2005.

[5] E. Mitchell, S. den Boon, and K. Lonnroth, "Acceptability of household and community-based tb screening in high burden communities: a systematic literature review. who," 2013.

[6] P. Van den Driessche, M. Li, and J. Muldowney, "Global stability of seirs models in epidemiology," *Canadian Applied Mathematics Quarterly*, vol. 7, pp. 409–425, 1999.

[7] F. G. Ball, E. S. Knock, and P. D. O'Neill, "Stochastic epidemic models featuring contact tracing with delays," *Mathematical biosciences*, vol. 266, pp. 23–35, 2015.

[8] C. Sun and Y.-H. Hsieh, "Global analysis of an seir model with varying population size and vaccination," *Applied Mathematical Modelling*, vol. 34, no. 10, pp. 2685–2697, 2010.

[9] B. Armbruster and M. L. Brandeau, "Optimal mix of screening and contact tracing for endemic diseases," *Mathematical biosciences*, vol. 209, no. 2, pp. 386–402, 2007.

[10] N. Wang, "Modeling and analysis of massive social networks," Ph.D. dissertation, 2005.

[11] Y. Zhang and B. A. Prakash, "Data-aware vaccine allocation over large networks," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 2, p. 20, 2015.

[12] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 2. IEEE, 2005, pp. 1455–1466.

[13] J. Aspnes, K. Chang, and A. Yampolskiy, "Inoculation strategies for victims of viruses and the sum-of-squares partition problem," *Journal of Computer and System Sciences*, vol. 72, no. 6, pp. 1077–1093, 2006.

[14] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos, "Gelling, and melting, large graphs by edge manipulation," in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 245–254.

[15] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti, "Approximation algorithms for reducing the spectral radius to control epidemic spread," in *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 2015, pp. 568–576.

[16] D. Kempe, J. Kleinberg, and É. Tardos, "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, pp. 137–146.

[17] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

[18] S. Suen, E. Bendavid, and J. Goldhaber-Fiebert, "Cost-effectiveness of improvements in diagnosis and treatment accessibility for tuberculosis control in India," *The International Journal of Tuberculosis and Lung Disease*, vol. 19, no. 9, pp. 1115–1124, 2015.

[19] D. E. Winetsky, D. M. Negoescu, E. H. DeMarchis, O. Almukhamedova, A. Dooronbekova, D. Pulatov, N. Vezhnina, D. K. Owens, and J. D. Goldhaber-Fiebert, "Screening and rapid molecular diagnosis of tuberculosis in prisons in russia and eastern europe: a cost-effectiveness analysis," *PLoS medicine*, vol. 9, no. 11, p. e1001348, 2012.

[20] J. A. Salomon, T. Vos, D. R. Hogan, M. Gagnon, M. Naghavi, A. Mokdad, N. Begum, R. Shah, M. Karyana, S. Kosen *et al.*, "Common values in assessing health outcomes from disease and injury: disability weights measurement study for the global burden of disease study 2010," *The Lancet*, vol. 380, no. 9859, pp. 2129–2143, 2012.

[21] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security (TISSEC)*, vol. 10, no. 4, p. 1, 2008.

[22] B. A. Prakash, D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," *Knowledge and information systems*, vol. 33, no. 3, pp. 549–575, 2012.

[23] Y. Qian, C. Zhang, B. Krishnamachari, and M. Tambe, "Restless poachers: Handling exploration-exploitation tradeoffs in security domains," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2016, pp. 123–131.

[24] T. Ayer, O. Alagoz, and N. K. Stout, "Or forum—a pomdp approach to personalize mammography screening decisions," *Operations Research*, vol. 60, no. 5, pp. 1019–1034, 2012.

[25] M. Shapiro and E. Delgado-Eckert, "Finding the probability of infection in an sir network is np-hard," *Mathematical biosciences*, vol. 240, no. 2, pp. 77–84, 2012.

[26] J. Kjeldsen, "The probability for infection in sir and sis networks," Master's thesis, 2013.

[27] C. Li, H. Wang, and P. Van Mieghem, "Degree and principal eigenvectors in complex networks," in *International Conference on Research in Networking*. Springer, 2012, pp. 149–160.

[28] C. N. Classen, R. Warren, M. Richardson, J. H. Hauman, R. P. Gie, J. H. Ellis, P. D. van Helden, and N. Beyers, "Impact of social interactions in the community on the transmission of tuberculosis in a high incidence area," *Thorax*, vol. 54, no. 2, pp. 136–140, 1999.

[29] B. Hendrickson and R. W. Leland, "A multi-level algorithm for partitioning graphs." *SC*, vol. 95, no. 28, 1995.

[30] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, and N. Voirin, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PloS one*, vol. 8, no. 9, p. e73970, 2013.

[31] M. Génois, C. L. Vestergaard, J. Fournet, A. Panisson, I. Bonmarin, and A. Barrat, "Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers," *Network Science*, vol. 3, no. 3, pp. 326–347, 2015.

[32] R. Mastrandrea, J. Fournet, and A. Barrat, "Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys," *PloS one*, vol. 10, no. 9, p. e0136497, 2015.

[33] A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson, "The diffusion of microfinance," *Science*, vol. 341, no. 6144, p. 1236498, 2013.

[34] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, "What's in a crowd? analysis of face-to-face behavioral networks," *Journal of theoretical biology*, vol. 271, no. 1, pp. 166–180, 2011.

[35] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proceedings of the National Academy of Sciences*, vol. 107, no. 51, pp. 22 020–22 025, 2010.

[36] T. Opsahl and P. Panzarasa, "Clustering in weighted networks," *Social networks*, vol. 31, no. 2, pp. 155–163, 2009.

[37] J. Moody, "Peer influence groups: identifying dense clusters in large networks," *Social Networks*, vol. 23, no. 4, pp. 261–283, 2001.

[38] "Chess network dataset – KONECT," Apr. 2017. [Online]. Available: http://konect.uni-koblenz.de/networks/chess

[39] L. E. Rocha, F. Liljeros, and P. Holme, "Information dynamics shape the sexual networks of internet-mediated prostitution," *Proceedings of the National Academy of Sciences*, vol. 107, no. 13, pp. 5706–5711, 2010.

[40] J. D. Williams, P. Poupart, and S. Young, "Factored partially observable markov decision processes for dialogue management," in *Proc. IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2005, pp. 76–82.

[41] A. Somani, N. Ye, D. Hsu, and W. S. Lee, "Despot: Online pomdp planning with regularization," in *Advances in neural information processing systems*, 2013, pp. 1772–1780.

[42] S.-c. Suen, E. Bendavid, and J. D. Goldhaber-Fiebert, "Disease control implications of India's changing multi-drug resistant tuberculosis epidemic," *PloS one*, vol. 9, no. 3, p. e89822, 2014.

[43] CDC, "Tuberculosis: General information," 2011. [Online]. Available: https://www.cdc.gov/tb/publications/factsheets/general/tb.pdf

[44] B. Bhattacharya, "Restless bandits visiting villages: A preliminary study on distributing public health services," in *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM, 2018, p. 7.