Repeated Active Screening of Networks for Diseases

Biswarup Bhattacharya, Han Ching Ou
University of Southern California
Los Angeles, California, USA
bbhattac@usc.edu

Arunesh Sinha
University of Michigan
Ann-Arbor, Michigan, USA

Sze-Chuan Suen, Bistra Dilkina, Milind Tambe
University of Southern California
Los Angeles, California, USA

ABSTRACT

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: (1) A new approach for modeling SEIS type diseases using a novel belief-state representation, (2) a novel community and eigenvalue-based algorithm (TRACE) to perform multi-round active screening. We perform extensive experiments on real-world datasets which emulate human contact, and illustrate significant benefits due to TRACE.

CCS CONCEPTS

• **Computing methodologies** → **Multi-agent planning**; **Partially-observable Markov decision processes**; • **Applied computing** → **Life and medical sciences**;

KEYWORDS

Public health

1 INTRODUCTION

Curable infectious diseases are responsible for millions of deaths every year. 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 [5]. However, 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 [2].

Our *first contribution* is a model of the active screening problem which considers the underlying disease dynamics. We focus on recurrent infectious diseases (no permanent immunity) 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). 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 real-world human contact networks against various baselines across a wide range of disease parameters (which also demonstrates its applicability to various other diseases).

2 PROBLEM SETUP

Disease Model. We adopt a SEIS model [6] for modeling the disease dynamics. TB and many other diseases follow a SEIS pattern, where treated individuals can relapse or become reinfected. The disease dynamics are therefore 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 (~half a year for TB).

Active Screening Model. We define k *active screening agents* that are to be deployed at every timestep t to diagnose and treat I and E individuals. Each individual is part of a (known) contact network and is in one of the (unknown) health states $\{S, E, I\}$, and infection spreads via the edges in the network. In every round, the agents can either choose to screen a node i (action $a_i = 1$) or not ($a_i = 0$). A screened node is observed to be in state S, E, or I, and an unscreened node generates no observation.

The objective of the model is to choose the budget-limited actions at each time step in order to maximize the number of susceptible individuals over T time-steps: $\max \sum_{t=0}^{T} R(s^t)$, interpreted as the total number of disease-free half years [4].

Belief States. The agents maintain a belief about the state of every individual — $\mathbf{b}_i^t = [b_{i,S}^t, b_{i,E}^t, b_{i,I}^t]$, where $b_{i,j}^t$ is the probability node i is in state j, starting with no information at $t = 0$. The beliefs about the health states evolve over time as the agents gain information about individuals. This belief update procedure is an important and novel aspect of our proposed active screening model, however not elaborated here due to space constraints.

Algorithm 1 TRACE Algorithm

Input: Adjacency Matrix A of graph, Belief 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: $\overline{\mathbf{A}}, \overline{\mathbf{b}}^t, size \leftarrow Coarsen(A, g_1, g_2, g_3, b^t)$
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: $\quad \underline{A}, \underline{b}^t \leftarrow$ remove all nodes in $\mathbf{U} \setminus \mathbf{u}'$ from A, b^t
10: $\quad a \leftarrow$ DynamicEigen($\underline{A}, \underline{b}^t, \mathbf{1}, \kappa$)
11: Active screen nodes $\{v \mid v \in a$ or $v \in \mathbf{u}$ for $\mathbf{u} \in \mathbf{U} \setminus \mathbf{u}'\}$

Algorithm 2 DynamicEigen(A, b^t, w, k)

Input: Adjacency matrix A, belief b, function w for weight of each node, min total weight of nodes to remove k

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

3 ALGORITHM FOR ACTIVE SCREENING

We introduce a structured algorithm to generate an online POMDP policy—Targeted Resolution of Active diseases using Communities and Eigenvalues (TRACE)—that combines elements of three approaches (Max Belief, and eigenvalue based, and community based methods) to identify the k individuals to actively screen at every time-step. The complete TRACE algorithm is shown in Algorithm 1.

As we do not know the true health state of all nodes in the network, we first assign an attractiveness score to each node ($R_i^t = \sigma b_{i,E}^t + b_{i,I}^t$) to reflect the effectiveness of intervening on the node. 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) (lines 1-3).

After labeling all nodes, locally similar nodes (nodes of the same label that share an edge) are clustered into a super-node iteratively using the known method of *graph coarsening* [3] (line 4).

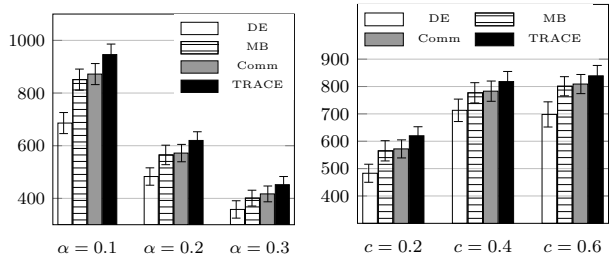
Prior methods to minimize the largest eigenvalue greedily chose nodes to delete in order to generate a graph with lower maximal eigenvalue. Since we do not know which nodes are infected with certainty, we augment this method by incorporating uncertainty (the DynamicEigen sub-procedure) and finally choosing the k nodes to screen through this process (lines 5-11).

4 EXPERIMENTS

We show results on only one network in this short abstract — **India** network [1]: A human contact network with $n = 202$ nodes, collected from a rural village in India, a setting in which TB active screening may take place ($1/\lambda_A^* \sim 0.095$). 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 — Survey stage ($t = 0 \sim 10$) and Active Screening stage ($t = 11 \sim 30$). We compare the benefit of the screening strategies over and above no intervention (**None**), where in **None** the evolution of the health states is based on disease dynamics with no active screening.

Results. Figure 1 shows 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 when compared to the three approaches: (a) **Dynamic Eigen (DE)**: Choose the nodes using just Algorithm 2 without any super-node formation; (b) **Max Belief (MB)**: Choose the nodes with the higher *belief* of being infected in that time-step, i.e. $b_{i,I}^t$; (c) **Community (Comm)**: Choose the nodes by a 0-1 knapsack algorithm (knapsack weight = budget k) after super-node formation.



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

Figure 1: Performance by the components (India network)

Our proposed novel active screening model and an algorithm (TRACE) to facilitate multi-round active screening for recurrent diseases performs significantly better than the baselines and each of its components individually in a variety of real-world inspired settings.

REFERENCES

[1] Abhijit Banerjee, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. The diffusion of microfinance. *Science* 341, 6144 (2013), 1236498.
[2] Palanivel Chinnakali, Pruthu Thekkur, Gomathi Ramaswamy, Kalaiselvi Selvaraj, et al. 2016. Active screening for tuberculosis among slum dwellers in selected urban slums of Puducherry, South India. *Annals of Tropical Medicine and Public Health* 9, 4 (2016), 295.
[3] Bruce Hendrickson and Robert W Leland. 1995. A Multi-Level Algorithm For Partitioning Graphs. *SC* 95, 28 (1995).
[4] IHME. 2010. *The Global Burden of Disease: Generating Evidence, Guiding, Policy.* World Bank.
[5] K Kranzer, H Afnan-Holmes, K Tomlin, Jonathan E Golub, AE Shapiro, A Schaap, EL Corbett, K Lönnroth, and JR Glynn. 2013. 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* 17, 4 (2013), 432–446.
[6] P Van den Driessche, M Li, and J Muldowney. 1999. Global stability of SEIRS models in epidemiology. *Canadian Applied Mathematics Quarterly* 7 (1999), 409–425.