52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

Optimal Vaccine Allocation to Control Epidemic Outbreaks in Arbitrary Networks

Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George Pappas

***Abstract*— We consider the problem of controlling the propagation of an epidemic outbreak in an arbitrary contact network by distributing vaccination resources throughout the network. We analyze a networked version of the Susceptible-Infected-Susceptible (SIS) epidemic model when individuals in the network present different levels of susceptibility to the epidemic. In this context, controlling the spread of an epidemic outbreak can be written as a spectral condition involving the eigenvalues of a matrix that depends on the network structure and the parameters of the model. We study the problem of finding the optimal distribution of vaccines throughout the network to control the spread of an epidemic outbreak. We propose a convex framework to find cost-optimal distribution of vaccination resources when different levels of vaccination are allowed. We illustrate our approaches with numerical simulations in a real social network.**

I. Introduction

The problem of controlling spreading processes in networks appear in many different settings, such as epidemiology [1], [2], computer viruses [3], or viral marketing [4]. The dynamic of the spread depends on both the structure of the contact network, the epidemic model and the values of the parameters associated to each individual. The dynamic behavior of spreading processes in networks have been widely studied. In [6], Newman studied the epidemic thresholds on several random graphs models. Pastor-Satorras and Vespignani studied viral propagation in power-law networks [7]. This initial work was followed by a long list of papers aiming to study the spread in more realistic network models. Boguna and Pastor-Satorras [8] considered the spread of a virus in correlated networks, where the connectivity of a node is related to the connectivity of its neighbors. In [9], the authors analyze spreading processes in random geometric networks. The analysis of spreading processes in arbitrary contact networks was first studied by Wang et al. [10] for the case of discrete-time dynamics. In [11], Ganesh et al. proposed a continuous-time Markov process to relate the speed of spreading with the largest eigenvalue of the adjacency matrix of the contact network. The connection between the speed of spreading and the spectral radius of the network was also found for a wide range of spreading models in [12]. The relationship between the spectral radius of a contact network and its local structural properties were explored in [13], [15].

The authors are with the Department of Electrical and Systems Engineering at the University of Pennsylvania, Philadelphia PA 19104.

The development of strategies to control the dynamic of a spread process is a central problem in public health and network security. In [16], Borgs et al. proposed a probabilistic analysis, based on the theory of contact processes, to characterize the optimal distribution of a fixed amount of antidote in a given contact network. In [17], Aditya et al. proposed several heuristics to immunize individuals in a network to control virus spreading processes. In the control systems literature, Wan et al. proposed in [18] a method to design optimal strategies to control the spread of a virus using eigenvalue sensitivity analysis ideas together with constrained optimization methods. Our work is closely related to the work in [19] and [20], in which a continuous-time time Markov processes, called the N-intertwined model, is used to analyze and control the spread of a SIS epidemic model.

In this paper, we propose a convex optimization framework to efficiently find the cost-optimal distribution of vaccination resources in an arbitrary contact network. In our work, we use a heterogeneous version of the N-intertwined SIS model [5] to model a spread process in a network of individuals with different rate of being infected and recovered. We assume that we can modify the rates of infection of individuals, within a feasible range, by distributing vaccines to the individuals in the network. We assume that there is a cost associated to injecting a particular amount of vaccination resources to a each individual, where the cost function can vary from individual to individual. Our aim is to find the optimal distribution of vaccination resources throughout the network in order to control the spread of an initial infection at a minimal cost.

II. Notation & Preliminaries

In this section we introduce some graph-theoretical nomenclature and the dynamic spreading model under consideration.

A. Graph Theory

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote an undirected graph with n nodes, m edges, and no self-loops[1]. We denote by $\mathcal{V}(\mathcal{G}) = \{v_1, \ldots, v_n\}$ the set of nodes and by $\mathcal{E}(\mathcal{G}) \subseteq \mathcal{V}(\mathcal{G}) \times \mathcal{V}(\mathcal{G})$ the set of undirected edges of $\mathcal{G}$. If $\{i, j\} \in \mathcal{E}(\mathcal{G})$ we call nodes i and j *adjacent* (or neighbors), which we denote

[1] An undirected graph with no self-loops is also called a *simple* graph.

978-1-4673-5716-6/13/$31.00 ©2013 IEEE

by $i \sim j$. We define the set of neighbors of a node $i \in \mathcal{V}$ as $\mathcal{N}_i = \{j \in \mathcal{V}(\mathcal{G}) : \{i,j\} \in \mathcal{E}(\mathcal{G})\}$. The number of neighbors of i is called the *degree* of node i, denoted by d_i. The adjacency matrix of an undirected graph $\mathcal{G}$, denoted by $A_\mathcal{G} = [a_{ij}]$, is an $n \times n$ symmetric matrix defined entry-wise as $a_{ij} = 1$ if nodes i and j are adjacent, and $a_{ij} = 0$ otherwise[2]. Since $A_\mathcal{G}$ is symmetric, all its eigenvalues, denoted by $\lambda_1(A_\mathcal{G}) \geq \lambda_2(A_\mathcal{G}) \geq \ldots \geq \lambda_n(A_\mathcal{G})$, are real.

B. N-Intertwined SIS Epidemic Model

Our modeling approach is based on the N-intertwined SIS model proposed by Van Mieghem et al. in [5]. Consider a network of n individuals described by the adjacency matrix $A_\mathcal{G} = [a_{ij}]$. The infection probability of an individual at node $i \in \mathcal{V}(\mathcal{G})$ at time $t \geq 0$ is denoted by $p_i(t)$.

The viral spreading is characterized by two positive parameters per node–an infection rate $\beta_i > 0$ and a curing rate $\delta_i > 0$. We use an extension of the the N-intertwined SIS model in [5] to the case of non-homogeneous infection and curing rates. The dynamics describing the infection is

$$\frac{d\boldsymbol{p}(t)}{dt} = (BA_\mathcal{G} - D)\,\boldsymbol{p}(t) - P(t)\,BA_\mathcal{G}\boldsymbol{p}(t), \quad (1)$$

where $\boldsymbol{p}(t) = (p_1(t), \ldots, p_n(t))^T$, $B = diag(\beta_i)$, $D = diag(\delta_i)$, and $P(t) = diag(p_i)$. Concerning the non-homogeneous epidemic model, we have the following result:

Proposition 1: Consider the heterogeneous N-intertwined SIS epidemic model in (1). Then, if

$$\lambda_1(BA - D) \leq -\varepsilon,$$

for some $\varepsilon > 0$, an initial infection $\boldsymbol{p}(0) \in [0,1]^n$ will converge to zero exponentially fast.

First, we have

$$\begin{aligned}\frac{dp_i(t)}{dt} &= \beta_i \sum_{j=1}^{n} a_{ij} p_j(t) - \delta_i p_i(t) - \beta_i p_i(t) \sum_{j=1}^{n} a_{ij} p_j(t)\\ &\leq \beta_i \sum_{j=1}^{n} a_{ij} p_j(t) - \delta_i p_i(t),\end{aligned}$$

since β_i, δ_i, $p_i(t)$,$a_{ij} \geq 0$. Therefore, the linear dynamic system

$$\frac{d\hat{p}_i(t)}{dt} = \beta_i \sum_{j=1}^{n} a_{ij} \hat{p}_j(t) - \delta_i \hat{p}_i(t), \quad (2)$$

upper-bounds the nonlinear dynamical system (1) when they share the same initial conditions, i.e., $\hat{\boldsymbol{p}}(t) \geq \boldsymbol{p}(t)$ for $t \geq 0$ when $\hat{\boldsymbol{p}}(0) = \boldsymbol{p}(0)$.

This linear dynamic system can be written in matrix form as

$$\frac{d\hat{\boldsymbol{p}}(t)}{dt} = (BA_\mathcal{G} - D)\,\hat{\boldsymbol{p}}(t).$$

[2]For simple graphs, $a_{ii} = 0$ for all i.

For the above linear system to be stable, we need the eigenvalues of $BA - D$ to be in the open left half-plane. The state matrix $BA_\mathcal{G} - D$ has real eigenvalues, since it can be transform via a similarity transformation to the symmetric matrix $B^{1/2}A_\mathcal{G}B^{1/2} - D$. Hence, exponential asymptotic stability, with an exponential rate ε, is equivalent to the largest eigenvalue $\lambda_1(BA_\mathcal{G} - D) < -\varepsilon$. In the above analysis, we have shown that the linear dynamics in (2) upper-bounds the mean-field approximation in (1); thus, the spectral result in Proposition 1 is a sufficient condition to control the evolution of an epidemic outbreak. In the following section, we use this result to characterize the profiles of infection rates that results in a stable linear dynamics.

III. A Convex Framework for Optimal Resource Allocation

In this section, we consider the partial vaccination problem. In the partial case, we assume that we are able to modify the infections rates β_i in the network by distributing vaccination resources throughout the individuals in the network. We assume that the infection rates of each individual can be modified within a particular feasible interval, $\underline{\beta}_i \leq \beta_i \leq \bar{\beta}_i$, where $\bar{\beta}_i > 0$ is the value of the natural infection rate for node i, which is achieved in the absence of any nodal immunization, and $\underline{\beta}_i > 0$ is the minimum possible infection rate for node i, which is achieved when we allocate a large amount of vaccines at node i. We propose an optimization framework to find the optimal distribution of resources when there is a cost function function associated to different values of β_i.

A. Vaccination Cost

The cost of achieving a particular infection rate for node i is denoted by $f_i(\beta_i)$. This cost function is node-dependent and presents the following properties: (it i) The cost of achieving the natural infection rate is zero, i.e., $f_i(\bar{\beta}_i) = 0$, (it ii) the maximum cost of vaccinating node i, denoted by T_i, is achieved at the minimum infection rate, i.e., $\max_{\beta_i} f_i(\beta_i) = f_i\left(\underline{\beta}_i\right) \triangleq T_i$, and (it iii) the vaccination cost function is monotonically decreasing in the interval $\beta_i \in \left[\underline{\beta}_i, \bar{\beta}_i\right]$. Apart from the above properties, we make the following convexity assumptions on the cost function f_i to obtain a tractable convex framework:

Assumption 1: The vaccination cost function, $f_i(\beta_i)$, is twice differentiable and satisfies the following constrain:

$$f_i''(\beta_i) \geq -\frac{2}{\beta_i} f_i'(\beta_i), \quad (3)$$

for $\beta_i \in \left[\underline{\beta}_i, \bar{\beta}_i\right]$.

Notice that, since f_i is monotonically decreasing, we have that $f_i'(\beta_i) < 0$; thus, we have that Assumption 1 implies that $f_i''(\beta_i) > 0$. In other words, Assumption 1 is stronger than convexity. For example, a function that satisfies Assumption 1 with equality is:

$$f_i(\beta_i) = T_i \frac{\beta_i^{-1} - \bar{\beta}_i^{-1}}{\underline{\beta}_i^{-1} - \bar{\beta}_i^{-1}}. \quad (4)$$

In practice, for low values of $\underline{\beta}_i$ and $\bar{\beta}_i$, this function takes a shape of practical interest.

B. Problem Statements

In this subsection we propose an optimization framework to find the cost-optimal allocation of vaccines in a given contact network $\mathcal{G}$ with adjacency matrix $A_{\mathcal{G}}$. In particular, we consider the following problem:

Problem 1: Given a curing rate profile, $\{\delta_i : i \in \mathcal{V}(\mathcal{G})\}$*, and a vaccination cost function* $f_i(\beta_i)$ for $\beta_i \in \left[\underline{\beta}_i, \bar{\beta}_i\right]$*, find the optimal distribution of vaccines to control the propagation of an epidemic outbreak with an asymptotic exponential decaying rate* ε *at a total minimum cost.*

According to Proposition 1, this problem can be mathematically stated as the following optimization problem:

$$\begin{aligned} T^* = min_{\{\beta_i\}} \quad & \textstyle\sum_{i=1}^n f_i(\beta_i) \\ s.t. \quad & \lambda_1(BA_{\mathcal{G}} - D) \le -\varepsilon \\ & \underline{\beta}_i \le \beta_i \le \bar{\beta}_i, \qquad i = 1, \ldots, n, \end{aligned} \quad (5)$$

In the following subsection, we propose a convex formulation to solve this problem under Assumption 1.

C. Semidefinite Programming (SDP) Approach

Our formulation is based on writing the spectral stability condition $\lambda_1(BA_{\mathcal{G}} - D) \le -\varepsilon$ using a simple semidefinite constrain. In particular, we have the following result:

Lemma 3.1: For $A_{\mathcal{G}}$ symmetric, $B = diag(\beta_i)$ and $D = diag(\delta_i)$ with $\beta_i, \delta_i > 0$, we have that $\lambda_1(BA_{\mathcal{G}} - D) \le -\varepsilon$ if and only if $(D - \varepsilon I) B^{-1} - A_{\mathcal{G}} \succeq 0$.

Notice that $BA_{\mathcal{G}} - D$ is a matrix similar to $B^{1/2}A_{\mathcal{G}}B^{1/2} - D$, since we can pre- and post- multiply the former matrix by $B^{-1/2}$ and $B^{1/2}$, respectively, to obtain the latter. Hence, since $B^{1/2}A_{\mathcal{G}}B^{1/2} - D$ is a symmetric matrix with real eigenvalues, the eigenvalues of $BA_{\mathcal{G}} - D$, including $\lambda_1(BA_{\mathcal{G}})$, are all real. Then, we have that $\lambda_1(BA_{\mathcal{G}} - D) \le -\varepsilon$ if and only if $\lambda_i((D - \varepsilon I) - BA_{\mathcal{G}}) = \lambda_i\left((D - \varepsilon I) - B^{1/2}A_{\mathcal{G}}B^{1/2}\right) \ge 0$, which is equivalent to $(D - \varepsilon I) - B^{1/2}A_{\mathcal{G}}B^{1/2} \succeq 0$. Applying a congruence transformation to $(D - \varepsilon I) - B^{1/2}A_{\mathcal{G}}B^{1/2}$ by pre- and post-multiplying by $B^{-1/2}$, we obtain that $\lambda_1(BA_{\mathcal{G}} - D) \le -\varepsilon$ if and only if $(D - \varepsilon I) B^{-1} - A_{\mathcal{G}} \succeq 0$. Using the above Lemma, we can rewrite the optimization problem 1 as a convex optimization program, as follows. First, let us rewrite (5) using the change of variables $\gamma_i \triangleq \beta_i^{-1}$as,

$$\begin{aligned} T^* \triangleq \min_{\{\gamma_i\}} \quad & \textstyle\sum_{i=1}^n f_i\left(\gamma_i^{-1}\right) \\ s.t. \quad & (D - \varepsilon I)\Gamma - A_{\mathcal{G}} \succeq 0 \\ & \bar{\beta}_i^{-1} \le \gamma_i \le \underline{\beta}_i^{-1}, \qquad i = 1, \ldots, n, \end{aligned} \quad (6)$$

where $\Gamma = diag(\gamma_i)$. Therefore, the feasible set is convex in the space of variables γ_i, $i = 1, \ldots, n$. Furthermore, we now verify that the cost function $\sum_{i=1}^n f_i\left(\gamma_i^{-1}\right)$ is also convex under Assumption 1 by computing its second derivative,

$$\frac{d^2}{d\gamma_i^2} \sum_i f_i\left(\gamma_i^{-1}\right) = f_i''\left(\gamma_i^{-1}\right)\frac{1}{\gamma_i^4} + 2f_i'\left(\gamma_i^{-1}\right)\frac{1}{\gamma_i^3} \ge 0,$$

where the last inequality is obtained from Assumption 1, taking into account that $\gamma_i^{-1} = \beta_i$.

The convex optimization program in (6) allows us to efficiently find the cost-optimal allocation of vaccines to control the spread of an epidemic outbreak in a given contact network. In the following subsection, we illustrate our approach in a real social network.

D. Numerical Results

We illustrate our results by designing the optimal distribution of vaccines in an online social network when the cost vaccination function follows (4). We consider a social network with 247 nodes, and assume that the individuals in the network present the same recovery rate, $\delta_i = \delta = 0.1$. In this case, we can rewrite (6) as a convex program with a convenient structure, as follows. First, defining $a \triangleq \left(\underline{\beta}_i^{-1} - \bar{\beta}_i^{-1}\right)^{-1}$, we have that

$$\sum_i f_i(\beta_i) = a\sum_i \beta_i^{-1} - a\sum_i \bar{\beta}_i^{-1} = a\text{Trace}(\Gamma) - b,$$

where $b \triangleq \alpha \sum_i \bar{\beta}_i^{-1}$. Hence, minimizing $\sum_i f_i(\beta_i)$ is equivalent to minimizing Trace(Γ). Thus, the optimization problem in (6) can be written as the following semidefinite program (SDP):

$$\begin{aligned} T^* \triangleq \min_{\Gamma} \quad & \text{trace}(\Gamma) \\ s.t. \quad & (\delta - \varepsilon)\Gamma - A_{\mathcal{G}} \succeq 0 \\ & \bar{\beta}_i^{-1} \le \gamma_i \le \underline{\beta}_i^{-1}, \quad i = 1, \ldots, n, \end{aligned} \quad (7)$$

Given our network with 247 nodes, we now compute the optimal distribution of vaccinations in several cases.

The network under consideration has a maximum eigenvalue $\lambda_1(A_{\mathcal{G}}) = 13.52$. In our simulations, individuals have the same natural infection rates $\bar{\beta}_i = \bar{\beta}$, and study three cases: $\bar{\beta} \in \{1.2\beta_c, 1.8\beta_c, 2.4\beta_c\}$. We choose the value of $\underline{\beta}_i < \beta_c$ to induce a stable disease-free equilibrium in the case of full-force vaccination, i.e., we saturate all the individuals with vaccines to shift their infection rates to $\underline{\beta}_i$. In our simulations we use a minimum infection rate $\underline{\beta}_i = 0.2\bar{\beta}_i = 0.2\bar{\beta}$; hence, we obtain

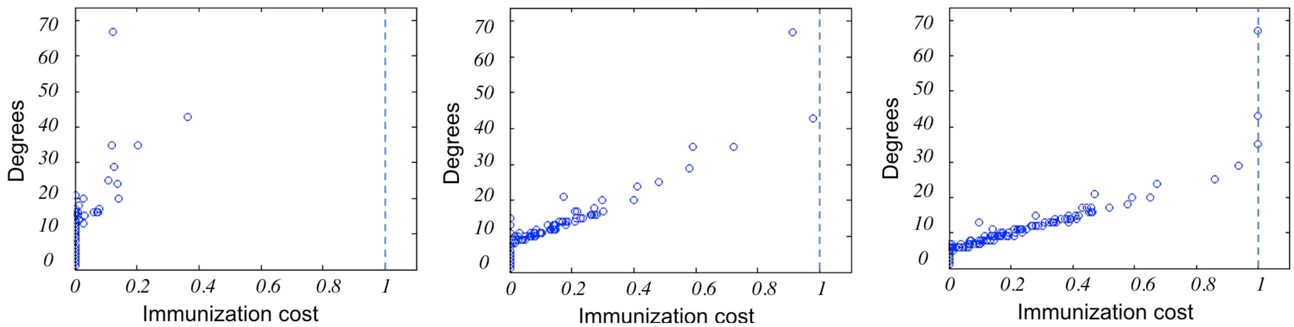

Fig. 1. Vaccination costs versus degree in a social network with 247 nodes.

that $\underline{\beta}_i \in \{0.24\beta_c, 0.36\beta_c, 0.48\beta_c\}$. In other words, our vaccine reduces the infection rate to a 20% of the natural infection rate. Using these parameter values, we run three simulations, each one with a different $\bar{\beta}$.

The results of our simulations are summarized in Fig. 1. Each one of the subplots in this figure corresponds to a different value of $\bar{\beta} \in \{1.2\beta_c, 1.8\beta_c, 2.4\beta_c\}$. For each value of $\bar{\beta}$ we present a scatter plot with 247 data points (as many as individuals in the network), where each point has an abscissa equal to $f_i(\beta_i)$ (the cost of vaccinating node i with optimal fraction β_i) and an ordinate of d_i (the degree of $i \in \mathcal{V}(\mathcal{G})$).

IV. COMBINATORIAL RESOURCE ALLOCATION

In this Section, we consider a combinatorial vaccination problem in which the optimal distribution of vaccines is allowed to be in the feasible interval $\beta_i \in [\underline{\beta}_i, \bar{\beta}_i]$. In the combinatorial vaccination problem, we restrict the resources to be in the discrete set, $\beta_i \in \{\underline{\beta}_i, \bar{\beta}_i\}$. For this case, we propose a greedy approach that provides an approximation to the optimal combinatorial solution. We also provide quality guarantees for this approximation algorithm in Subsection IV-B. The combinatorial vaccination problem can be stated as follows:

Problem 2: Given a curing rate profile, $\{\delta_i : i \in \mathcal{V}(\mathcal{G})\}$, and a vaccination cost function $f_i(\beta_i)$ for $\beta_i \in \{\underline{\beta}_i, \bar{\beta}_i\}$, find the optimal distribution of vaccines to control the propagation of an epidemic outbreak with an asymptotic exponential decaying rate ε at a total minimum cost.

The optimal distribution of vaccines in Problem 2 can be characterized by the set of individuals $I_C \subseteq \mathcal{V}(\mathcal{G})$ that are chosen to be fully immunized, i.e., the infection rates are switched from $\bar{\beta}_i$ to $\underline{\beta}_i < \bar{\beta}_i$ for $i \in I_C$. Let us assume that the vaccination cost function takes the values $f_i(\bar{\beta}_i) = 0$ and $f_i(\underline{\beta}_i) = c_i$. These extreme values are achieved using the following affine cost function

$$f_i(\beta_i) \triangleq c_i \frac{\beta_i - \bar{\beta}_i}{\underline{\beta}_i - \bar{\beta}_i}.$$

Hence, the total cost of vaccination satisfies

$$\sum_{i=1}^{n} f_i(\beta_i) = a_C \sum_i c_i \beta_i - b_C,$$

where we have defined the constants $a_C \triangleq (\underline{\beta}_i - \bar{\beta}_i)^{-1}$ and $b_C \triangleq a_C \sum_i c_i \bar{\beta}_i$. Thus, since $a_C < 0$, the optimal allocation of vaccines that minimizes $\sum_{i=1}^n f_i(\beta_i)$ is the same as the one that maximizes $\sum_i c_i \beta_i$. Therefore, defining the vectors $c \triangleq (c_1, \ldots, c_n)^T$ and $b \triangleq (\beta_1, \ldots, \beta_n)^T$, Problem 2 can be stated as the following optimization problem:

$$\begin{aligned} T_C^* = \max_{\{\beta_i\}} \quad & c^T b \\ s.t. \quad & \lambda_1(BA_{\mathcal{G}} - D) \leq -\varepsilon \\ & \beta_i \in \{\underline{\beta}_i, \bar{\beta}_i\}, \qquad i = 1, \ldots, n. \end{aligned} \tag{8}$$

The solution to this problem is combinatorial in nature. In the following subsections we provide a greedy approach that approximates the combinatorial solution, as well as a quality guarantee of our approach.

A. Greedy approach

In this subsection, we provide a greedy algorithm that iteratively updates the set of nodes that will be (fully) vaccinated in order to control the spreading of an epidemic outbreak. In each step of our algorithm, we denote the set of nodes that are chosen to be part of the vaccination group as S_t. We iteratively add to this group the node that provides the most benefit per unit cost, where the benefit of vaccinating a is the increment it induces in $\lambda_1(BA_{\mathcal{G}} - D)$. More formally, given a vaccination group S_t, we define the diagonal matrix of associated infection rates as $B_{S_t} \triangleq diag(\bar{\beta}_i) - (\bar{\beta}_i - \underline{\beta}_i) diag(\mathbf{1}_{S_t})$, where $\mathbf{1}_{S_t}$ is the n-dimensional indicator vector for the set S_t. Thus, the benefit per unit cost of adding node i to S_t is measured by the function

$$\Delta(i, S_t) \triangleq \frac{\lambda_1(B_{S_t} A_{\mathcal{G}} - D) - \lambda_1(B_{S_t + \{i\}} A_{\mathcal{G}} - D)}{c_i}.$$

Parameters	Metric	Greedy	Reverse Greedy	Degree Threshold	Centrality Threshold	D^*
$\bar{\beta} = 2.4\beta_c$	$c'b$	3.6298	3.6440	3.2892	2.4518	3.9425
$\underline{\beta} = 0.2\bar{\beta}$	$\lambda_1(\delta B^{-1} - A)$	0.0054	0.0355	0.0422	0.1982	n/a
$\bar{\beta} = 1.8\beta_c$	$c'b$	3.0098	3.0098	2.9246	2.0092	3.1406
$\underline{\beta} = 0.2\bar{\beta}$	$\lambda_1(\delta B^{-1} - A)$	0.0850	0.1383	0.0774	0.2575	n/a
$\bar{\beta} = 1.2\beta_c$	$c'b$	2.1484	2.1484	2.1201	1.7369	2.1787
$\underline{\beta} = 0.2\bar{\beta}$	$\lambda_1(\delta B^{-1} - A)$	0.4383	0.4383	0.6278	1.0101	n/a

Fig. 2. Table with values of the objective function $c'b$ and the residual value of $\lambda_1\left(\delta B^{-1} - A_{\mathcal{G}}\right)$ for each possible value of $\bar{\beta}_i$.

A conventional greedy approach could be defined by the iteration $S_{t+1} = S_t + \{i_t\}$ with $S_1 = \{\}$ and $i_t \triangleq \arg\max_i \Delta(i, S_t)$, where this iteration is repeated until $\lambda_1(B_{S_t} A_{\mathcal{G}} - D) \leq -\varepsilon$ is satisfied. Notice that the resulting vaccination group is feasible and satisfies the spectral condition needed to control the spreading of an epidemic outbreak.

In practice, we observe that a modification of this greedy approach provides better results. In this modified version, we start with a vaccination set $S_1 = \mathcal{V}(\mathcal{G})$ (i.e., all the individuals are vaccinated) and iteratively remove individuals according to the iteration $S_{t+1} = S_t - \{j_t\}$ with $j_t = \arg\min_j \Delta(j, S_t \setminus \{j\})$, where this iteration is repeated until $\lambda_1(B_{S_t} A_{\mathcal{G}} - D) \geq -\varepsilon$ is satisfied. The final vaccination group is chosen to be S_{t-1}. Notice that, the resulting vaccination group is feasible and $\lambda_1\left(B_{S_{t-1}} A_{\mathcal{G}} - D\right) \leq -\varepsilon$. We denote this approach the *reverse greedy algorithm*.

Since our approach is heuristic for a combinatorial problem, we provide a quality guarantee via Lagrange duality theory in the following subsection.

B. Quality Guarantee

Using Lagrange duality theory, we provide quality guarantees for the performance of our greedy approach by computing the dual optimal D_C^*.

Theorem 4.1: Given the optimization problem

$$\begin{aligned} T_C^* = \quad & \max_b \quad c^T b \\ & \text{s.t.} \quad (D - \varepsilon I) B^{-1} - A_{\mathcal{G}} \succeq 0 \\ & \qquad \beta_i \in \{\underline{\beta}_i, \bar{\beta}_i\}, \ \forall i, \end{aligned} \tag{9}$$

the primal optimal T_C^* can be upper bounded by D_C^* computed according to the Lagrange dual

$$\begin{aligned} D_C^* = \quad & \min_{Z,u} \quad \mathbf{1}^T u - trace(A_{\mathcal{G}} Z) \\ & \text{s.t.} \quad u_i \geq c_i \bar{\beta}_i + \frac{\delta_i}{\bar{\beta}_i} Z_{ii} \, \forall i \\ & \qquad u_i \geq c_i \underline{\beta}_i + \frac{\delta_i}{\underline{\beta}_i} Z_{ii} \, \forall i \\ & \qquad Z \succeq 0, \end{aligned} \tag{10}$$

which is a convex Semidefinite Program.

Proof: Notice that the matrix in the above semidefinite constraint can be written as $(D - \varepsilon I) B^{-1} - A_{\mathcal{G}} = \sum_i e_i e_i' \frac{\delta_i - \varepsilon}{\beta_i} - A_{\mathcal{G}}$, where e_i is the unit vector in the standard basis. From (9), we construct the Lagrangian

$$\mathcal{L}(b, Z) = c^T b + trace\left(Z\left(\sum_i e_i e_i' \frac{\delta_i}{\beta_i} - A_{\mathcal{G}}\right)\right), \tag{11}$$

where $\beta_i \in \{\underline{\beta}_i, \bar{\beta}_i\}$ is kept as a domain constraint and $Z \succeq 0$. See Section 5.9 of [21] for further details on the Lagrange dual of semidefinite constraints. Using the properties of trace to simplify and decouple we get

$$\mathcal{L}(b, Z) = \sum_i \left(c_i \beta_i + \frac{\delta_i}{\beta_i} Z_{ii}\right) - trace(Z A_{\mathcal{G}}). \tag{12}$$

The dual objective is derived by maximizing the Lagrangian with respect to the primal variables

$$q(Z) = \sum_i \left(\max_{\beta_i} c_i \beta_i + \frac{\delta_i}{\beta_i} Z_{ii}\right) - trace(Z A_{\mathcal{G}}). \tag{13}$$

Due to the decoupling in (12) the primal optimization in (13) can be done for each node, independently. Since each node has only 2 options we can consider each case explicitly by defining

$$u_i = \max\left\{c_i \bar{\beta}_i + \frac{\delta_i}{\bar{\beta}_i} Z_{ii}, c_i \underline{\beta}_i + \frac{\delta_i}{\underline{\beta}_i} Z_{ii}\right\}. \tag{14}$$

It is possible to compute u_i as a threshold function of Z_{ii}, but for the purpose of constructing the dual it is better to use an epigraph formulation to rewrite (13) as

$$q(Z, u) = \sum_i u_i - trace(Z A_{\mathcal{G}}) \tag{15}$$

with the addition constraints that

$$u_i \geq c_i \bar{\beta}_i + \frac{\delta_i}{\bar{\beta}_i} Z_{ii} \tag{16}$$

$$u_i \geq c_i \underline{\beta}_i + \frac{\delta_i}{\underline{\beta}_i} Z_{ii}. \tag{17}$$

Since the dual is a minimization and $q(Z, u)$ is strictly increasing in u, either (16) or (17) must be achieved with equality, ensuring that the definition (14) is satisfied at the optimal point. To conclude, our dual (10) is given by minimizing (15) subject to the domain constraint $Z \succeq 0$ and the epigraph constraints (16) and (17). This is a standard form SDP as defined in section 4.6 of [21]. The solution D_C^* is guaranteed to satisfy $D_C^* \geq T_C^*$ by weak duality, [21] Section 5.2. ■

Theorem 4.1 tells us that for any optimization problem of the form (9) we can get an accuracy certificate

$$T_C^* - c^T b \leq D_C^* - c^T b \tag{18}$$

by solving the dual (10). Since we do not have a strong duality, we do not expect $c^T b = D_C^*$ to be attainable (i.e, $P_C^* < D_C^*$).

Remark 4.1: The solution to the dual gives us some insight into the primal optimizers via the threshold solution to (14),

$$u_i(Z_{ii}) = \begin{cases} c_i \bar{\beta}_i + \frac{\delta_i}{\bar{\beta}_i} Z_{ii} & \text{if } Z_{ii} \leq \frac{c_i}{\delta_i} \bar{\beta}_i \underline{\beta}_i \\ c_i \underline{\beta}_i + \frac{\delta_i}{\underline{\beta}_i} Z_{ii} & \text{if } Z_{ii} \geq \frac{c_i}{\delta_i} \bar{\beta}_i \underline{\beta}_i \end{cases}. \tag{19}$$

It appears we can deduce the primal optimizers b^* from Z^*, but in practice for most nodes i, $Z_{ii}^* = \bar{\beta}_i \underline{\beta}_i c_i / \delta_i$ making it impossible to determine β_i^*. In some cases there are nodes that have Z_{ii}^* not equal to the threshold. These nodes have their optimal action specified by Z_{ii}^* and (19). This at least allows for a reduction of the dimension of the primal problem which due to its combinatorial form could be a very large improvement.

Several papers in the literature advocate for vaccination strategies based on popular centrality measures, such as the degree or eigenvector centrality [22]. In this subsection, we compare our greedy heuristic to vaccination strategies based on centrality measures. In our simulation, we use the adjacency matrix with 247 nodes previously used in Subsection III-D and the same values for the parameters $\delta_i = \delta = 0.1$, $\beta_c = \delta / \lambda_{\max}(A_{\mathcal{G}}) = 7.4e-3$, $\bar{\beta}_i \in \{1.2\beta_c, 1.8\beta_c, 2.4\beta_c\}$ and$\underline{\beta}_i = 0.2\bar{\beta}_i$ for all i. In Table 2, we include the values of the objective function $c'b$ and the residual value of $\lambda_1 \left(\delta B^{-1} - A_{\mathcal{G}}\right)$ for each possible value of $\bar{\beta}_i$. In each case, we run the greedy algorithm and the reverse greedy algorithm (both proposed in Section IV-A), as well as two previously proposed algorithms based on the degree and the eigenvalue centrality metrics. In the last column of Table 2, we also include the upper bound provided by Theorem 4.1. Observe that our greedy algorithms are always within 10% of the upper bound D_C^*. Furthermore, the reverse greedy algorithm outperforms the others, specially those based on centrality measures.

V. Conclusions

We have studied the problem of controlling the dynamic of the SIS epidemic model in an arbitrary contact network by distributing vaccination resources throughout the network. Since the spread of an epidemic outbreak is closely related to the eigenvalues of a matrix that depends on the network structure and the parameters of the model, we can formulate our control problem as a spectral optimization problem in terms of semidefinite constraints. In the partial vaccination case, where intermediate level of vaccination are allowed, we have proposed a convex optimization framework to efficiently find the optimal allocation of vaccines when the function representing the vaccination cost satisfies certain convexity assumptions. In the combinatorial vaccination problem, where individuals are not allowed to be partially vaccinated, we propose a greedy approach with quality guarantees based on Lagrangian duality. We illustrate our results with numerical simulations in a real online social network.

References

[1] N. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed. Charlin Griffin, 1975.

[2] R.M. Anderson and R.M. May, *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 1991.

[3] M. Garetto, W. Gong, D. Towsley, "Modeling Malware Spreading Dynamics," in *Proc. IEEE INFOCOM*, 2003.

[4] J. Leskovec, L.A. Adamic, and B.A. Huberman, "The Dynamics of Viral Marketing," *ACM Transactions on the Web* , vol. 1, 2007.

[5] P. Van Mieghem, J. Omic, and R. Kooij, "Virus Spread in Networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1–14, 2009.

[6] M.E.J. Newman, "Spread of Epidemic Disease on Networks," *Physical Review E*, vol. 66, no. 1, 016128, 2002.

[7] R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Review Letters*, vol. 86, no. 14, 2001.

[8] M. Boguna and R. Pastor-Satorras, "Epidemic Spreading in Correlated Complex Networks," *Physical Review E*, vol. 66, no. 4, 047104, 2002.

[9] V.M. Preciado and A. Jadbabaie, "Spectral Analysis of Virus Spreading in Random Geometric Networks," *Proc. IEEE Conference on Decision and Control*, 2009.

[10] Y. Wang, D. Chakrabarti, C. Wang, and C. Falutsos, "Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint," *Proc. IEEE Reliable Distributed Systems*, 2003.

[11] A.J. Ganesh, L. Massoulie, D.F. Towsley, "The Effect of Network Topology on the Spread of Epidemics," *Proc. IEEE INFOCOM*, pp. 1455–1466, 2005.

[12] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic Thresholds in Real Networks," *ACM Trans. on Information and System Security*, vol. 10, no. 4, 2008.

[13] V.M. Preciado and A. Jadbabaie, "Moment-Based Spectral Analysis of Large-Scale Networks Using Local Structural Information," to appear in *ACM/IEEE Transactions on Networking*, 2013.

[14] V.M. Preciado, A. Jadbabaie, and M. Draief, "Structural Analysis of Viral Spreading Processes in Social and Communication Networks Using Egonets," ArXiv:1209.0341v1, 2013.

[15] V.M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. Pappas, "Optimal Resource Allocation to Control Spreading Processes in Networks: A Convex Framework," *Submitted for publication*.

[16] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, "How to Distribute Antidote to Control Epidemics," *Random Structures and Algorithms*, vol. 37, pp. 204–222, 2010.

[17] B. Aditya Prakash, L. Adamic, T. Iwashnya, H. Tong, and C. Faloutsos, "partial Immunization on Networks," in *Proc. SIAM Data Mining*, 2013.

[18] Y. Wan, S. Roy, and Ali Saberi, "Designing Spatially Heterogeneous Strategies for Control of Virus Spread," *IET Systems Biology*, vol. 2, pp. 184–201, 2008.

[19] E. Gourdin, J. Omic, and P. Van Mieghem, "Optimization of Network Protection Against Virus Spread," in *Proc. Design of Reliable Communication Networks*, 2011.

[20] F. Darabi Sahneh and C. Scoglio, "Optimal Information Dissemination in Epidemic Networks," *IEEE Conference on Decision and Control*, 2012.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[22] F. Chung, P. Horn, and A. Tsiatas, "Distributing Antidote Using PageRank Vectors," *Internet Mathematics*, vol. 6, pp. 237–254, 2009.