Controlling Contact Network Topology to Prevent Measles Outbreaks

Ilai Bistritz and Nicholas Bambos
Department of Electrical Engineering
Stanford University, Stanford, CA, 94305, USA
Email: bistritz@stanford.edu

Dor Kahana, Irad Ben-Gal and Dan Yamin
Department of Industrial Engineering
Tel Aviv University, Ramat-Aviv, Tel Aviv, 69978, Israel
Email: yamind@post.tau.ac.il

***Abstract*—Consider an epidemic that propagates in a network of N individuals. The dynamics of the infection are governed by the N-intertwined SIR model, which is a non-linear model. Our goal is to prevent the epidemic by removing (vaccinating) nodes and removing (closing) links. Since vaccinating nodes and closing links is costly, we want to minimize this cost under the constraint that the outbreak is prevented. We first show that preventing the outbreak can be guaranteed by ensuring that the maximal eigenvalue λ_1 of a specific linear system is negative. This induces a well posed, but highly complex, combinatorial optimization problem. We propose a greedy algorithm that at each step picks the approximately best link to close or node to vaccinate, and proceeds to break the network until $\lambda_1 < 0$. We also prove that running our algorithm on a coarser and smaller graph of regions as opposed to individuals still guarantees that the epidemic is prevented in the large network of size N. We tested our algorithm on an N-intertwined SIR model that was calibrated using real data that includes measles outbreaks and contact frequencies. The contact network was generated based on raw cellular localization data of 17 billion records from Radio Network Controllers (RNCs) that cover 1.8 million users over 2 months. Our encouraging results demonstrate that algorithms which consider the topology of the network can offer great value even in practical scenarios, where the decisions and computations can only be made on the regional level.**

I. Introduction

Measles is a highly contagious disease that remains a major cause of death among young children globally, despite the availability of a safe and effective vaccine. In recent years, many cases of measles outbreaks have been observed around the globe and the rates of disease and deaths have increased [1], [2]. The dominant reason for the increase of measles cases in developed countries is vaccine hesitancy [2], [3], which is considered one of the top 10 threats to global health, and is associated with the 30% global rise in measles cases [4].

If in a certain community vaccination rates are dramatically lower, even just few cases of measles can quickly spread throughout the community. Therefore, vaccine hesitancy can lead to a serious outbreak that in turn infects vaccinated individuals that did not successfully develop antibodies or simply infants that have yet to be vaccinated. It is intuitively clear that removing nodes and closing links around hesitant communities is much more beneficial than treating random or even central nodes and links in the network. Any policy that does not look at the topology of the network as a whole will miss this type of more intricate solutions, that in general are not intuitive.

Another major change in recent years is the availability of "big data". Today data such as measles cases and vaccination rates is easily accessible. When public policies are considered, then cellular localization data is also available. Using cellular localization over time, one can even estimate the topology of the network. Knowing the topology opens the way for more sophisticated computational methods for preventing outbreaks despite vaccine hesitancy.

Our goal is to prevent the epidemic by vaccinating nodes and closing links, while minimizing the cost of these countermeasures. The knowledge of the topology of the network allows quantifying the spread rate of a process on a graph using the eigenvalues of the adjacency or the Laplacian matrices (see [5]). This way, we can estimate the spread rate reduction caused by removing a certain node or deleting a certain link. By picking more effective nodes and links, the outbreak is prevented with a lower intervention cost.

However, for epidemics control in large networks, using spectral graph theory is far from being straight forward. First, the mathematical models of infectious disease, like Susceptible-Infected-Susceptible (SIS) or Susceptible-Infected-Recovered (SIR) [6] are not linear. Hence, it is not clear how the adjacency, Laplacian or any other matrix is related to the disease spread. Second, even if such a matrix is given, it represents the individuals in a country or a city, so its dimension is huge. Repeatedly computing the eigenvalue decomposition for a large matrix is not possible in practice, since the complexity is $O\left(N^3\right)$ where N is the population size. Exhaustive search for the best link to remove has a complexity of $O\left(N^5\right)$ per decision, which is not even practical for medium-sized networks.

Our main theoretical contribution is to prove that working on a smaller graph of regions (e.g., cities, districts, neighborhoods) instead of individuals can still guarantee the prevention of the outbreak in the actual larger network. Hence the complexity of our algorithm is $O\left(R^3\right)$ per decision, where R is the number of regions which typically satisfies $R \ll N$. This reduction is important not only for complexity considerations. In practice, decisions on vaccinations and link closure are made on some regional level or another, rather than on an individual level. Furthermore, we modify the approach of [7] to quantify the benefit from closing each link using the eigenvector, therefore reducing the complexity to $O\left(N^3\right)$ (or $O\left(R^3\right)$) per decision instead of $O\left(N^5\right)$ (or $O\left(R^5\right)$).

While the focus of this paper is preventing outbreaks, the theoretical ideas are readily applicable to computer viruses spreading over the internet, which share a similar model [8].

A. Previous Work

Modeling disease spread is an established and rich field of research [9]–[14]. Over the past decade, prescriptive studies have started to appear that rely on the descriptive modeling studies. Various works have considered the problem vaccinations allocation to prevent an outbreak on a graph [15]–[20]. The works that most resemble our approach are [18], [19].

In [19], a semi-definite programming (SDP) continuous optimization problem was formulated where the budget is minimized under a constraint that guarantees that the outbreak is prevented. The authors also provided a heuristic for the combinatorial optimization case. This work has considered the SIS model in continuous time, but more importantly, only considered vaccinations for nodes, leaving the links untouched. In each iteration of the algorithm in [19], an eigenvalue decomposition is performed once for each node to determine the best node to vaccinate. This results in a complexity of $O\left(N^4\right)$ per decision (vaccinating one node). Employing the same approach for link closure will result in a complexity of $O\left(N^5\right)$ per decision, which is already much worse than our $O\left(N^3\right)$, and is of a completely different scale compared to $O\left(R^3\right)$. In this work, we partially build upon the unpublished results in [18], that analyzed continuous optimization in the SIS case and only for the individuals network of size N.

B. Outline

The paper is organized as follows: in Section II we formulate our problem and describe our algorithm. In Section III we analyze the performance guarantees and properties of the proposed algorithm. In Section IV we test our algorithm on data driven networks. Section VI concludes the paper.

II. Problem Formulation

Counter-measures that break the network can prevent an outbreak even when vaccine hesitancy is significant. It has already been proposed that vaccinating infants earlier than usual might help preventing the epidemic [21]. The disadvantage is that the immune system of these infants might not be fully developed yet, making the vaccination useless for some of them. Another option is to run antibodies tests to detect what adults are not immune despite receiving a shot in the past. The disadvantage is the cost of this process. These possibilities amount to vaccinating (removing) certain nodes in our network. When vaccine hesitancy is involved, the number of non-vaccinated individuals that are willing to get vaccinated might be limited. Hence, other ways of breaking the network are necessary. Since infants and young children are at elevated risk to become infected and transmit the disease, temporarily closing daycares and kindergartens will slow down the disease spread. Closing other public places might achieve similar effects. In an emergency, temporarily closing roads might be necessary. In any case, closing only certain roads going out from a district is far more convenient for the residents than simply isolating the whole district. All these measures amount to closing certain links (contacts) in our network.

By considering the network topology, we can come up with localized policies that are much less costly than assigning vaccinations to everyone or closing all the links in a city. This, however, requires a solid mathematical model, developed next.

A. N-intertwined SIR Model

Next, we formulate our mathematical model, which uses the SIR model [6] in discrete time for each individual, and ties individuals together using a given network. A similar approach was taken in [19] for the SIS model. Let $G = (\mathcal{V}, \mathcal{L})$ be a network of N individuals, or nodes. Denote by A the adjacency matrix such that $a_{ij} = 1$ and $a_{ji} = 1$ if node i and j are connected in G and $a_{ij} = 0$ and $a_{ji} = 0$ otherwise.

The time index is discrete and denoted by $t = 0, 1, 2, ...$ Denote by $p_i(t)$ the probability that individual i is infected at time t. Denote by $s_i(t)$ the probability that individual i is susceptible (healthy) at time t. Denote by $r_i(t)$ the probability that individual i is recovered at time t. The vaccinated individuals can be either recovered from a previous infection, hence they have natural vaccination, or individuals that have received a vaccine shot. Each individual is either susceptible, infected or recovered so for all t and i

$$p_i(t) + s_i(t) + r_i(t) = 1. \tag{1}$$

Denote by $\delta_i > 0$ the recovery rate at node i. The recovery rate is uncontrollable in nature, and many studies estimated the recovery rate for Measles [22]. Denote by $\beta_i > 0$ the infection rate at node i. Define the vectors $\boldsymbol{p}(t) \triangleq (p_1(t), ..., p_N(t))^T$, $\boldsymbol{s}(t) \triangleq (s_1(t), ..., s_N(t))^T$ and $\boldsymbol{r}(t) \triangleq (r_1(t), ..., r_N(t))^T$.

An individual is at recovered state at time t if he was at recovered state at time $t-1$ or if he was infected at time $t-1$ and got cured at time t:

$$r_i(t) = r_i(t-1) + \delta_i p_i(t-1) \tag{2}$$

Next we derive the dynamics of $p_i(t)$. Assume that individuals get infected independently at random in time and between individuals. Hence, the probability that an individual i does not get infected from all of his neighbors is given by

$$\zeta_i(t) = \prod_{j=1}^{N} \left(p_j(t)\left(1 - \beta_i A_{ji}\right) + 1 - p_j(t)\right) = \prod_{j=1}^{N} \left(1 - p_j(t)\beta_i A_{ji}\right). \tag{3}$$

Individual i is infected at time t if he was infected at $t-1$ and did not get cured at t, or if he was susceptible at $t-1$ and got infected at t from at least one of his neighbors. Hence

$$p_i(t) = (1 - \delta_i)\, p_i(t-1) + \left(1 - \zeta_i^t\right) s_i(t-1) \tag{4}$$

B. Our Algorithm

Corollary 3 in Section III allows us to pose the constraint of "preventing the outbreak" in mathematical terms. Our suggested policy, or algorithm, is then based on a heuristic that aims to minimize the costs under this constraint. We now formalize this combinatorial optimization problem.

Denote by $\mathcal{V}$ the set of nodes we choose to vaccinate, and by $\mathcal{L}$ the set of links we choose to close. Let c_i be the cost of vaccinating node i and $c_{i,j}$ be the cost of closing link (i,j). Note that we can model a node i that refuses to get vaccinated by setting $c_i = \infty$. Define the diagonal matrices $B = \text{diag}\left(\{\beta_i\}_{i=1}^N\right)$ and $D = \text{diag}\left(\{\delta_i\}_{i=1}^N\right)$. Also define $P(t) = \text{diag}\left(\{p_i(t)\}_{i=1}^N\right)$ and $S(t) = \text{diag}\left(\{s_i(t)\}_{i=1}^N\right)$. Our goal is to solve the following combinatorial optimization problem:

$$\begin{array}{c} \min\limits_{\mathcal{V},\mathcal{L}} \sum_{i\in\mathcal{V}} c_i + \sum_{(i,j)\in L} c_{i,j} \\ s.t.\, \lambda_1(AB-D) < 0, \end{array} \tag{5}$$

where $\lambda_1(AB-D)$ is the maximal eigenvalue of the matrix $AB-D$.

In order to solve (5), we propose Algorithm 1. The idea behind the algorithm is that at each step, we want to pick a link (or a node, which is equivalent to the set of its links) such that the decrease of $\lambda_1(AB-D)$ is maximal. This idea is formalized in Subsection III-A.

The number of individuals N in a neighborhood, let alone a city or a state, can be huge. The eigenvalue decomposition in Step 1 of Algorithm 1 typically requires a time complexity of $O\left(N^3\right)$. This is impossible in practice already for $N = 10^4$. Furthermore, in a realistic scenario decision makers will be limited to policies that target regions rather than individuals. This leads us to the notion of the regions network:

Definition 1. In the regions network G_R of G, there are R regions which are the smallest geographical units for which different decisions about vaccinations and link closures can be made (e.g., neighborhoods, districts or cities). In G_R:

- There are two nodes for each region, one for all the individuals that can get vaccinated and the other one all the individuals that cannot or unwilling to get vaccinated.
- The weight of link (i_R, j_R) is the number of links in G that connect an individual in i_R to an individual in j_R.

III. Theoretical Performance Guarantees

In this section we prove that our algorithm is guaranteed to prevent the outbreak. Therefore, while being a heuristic, the heuristic part only affects the minimization of the cost of the countermeasures. The main goal of preventing the outbreak is fully analytical. However, our heuristic of minimizing the cost is not a guess but rather is based on analytical grounds, as explained in Subsection III-A. Subsection III-B proves that our algorithm prevents the outbreak even if it runs on a coarse regions network with size R and not the actual individual network with size N.

Algorithm 1 Greedy Link Closure Algorithm

Initialization: Let $c_{i.j}$ be the cost of closing link (i,j) and c_i the cost of vaccinating node i. Let $A = \{a_{ij}\}$ be the adjacency matrix of G or G_R, that the algorithm modifies at each iteration (see Definition 1).

While $\lambda_1 > 0$ **Do**

1) Compute the maximal eigenvalue λ_1 of $AB-D$ and its eigenvector v_1.
2) Assign each link (i,j) a grade $Q(i,j) = \frac{1}{c_{i.j}} \sum v_1(i)\, v_1(j)$ and each node i a grade $Q(i) = \frac{1}{c_i} \sum_j Q(i,j)$.
3) Compute $(l_1^*, l_2^*) = \arg\max\limits_{(i,j)} Q(i,j)$ and $i^* = \arg\max\limits_{i} Q(i)$.
   a) If $Q(i^*) > Q(l_1^*, l_2^*)$ then vaccinate node i^* by setting $a_{i^*j} = 0$ for all j.
   b) If $Q(i^*) \leq Q(l_1^*, l_2^*)$ then close link (l_1^*, l_2^*) by setting $a_{l_1^* l_2^*} = a_{l_2^* l_1^*} = 0$.

End

The SIR N-intertwined model in (5) is non-linear for two reasons - the multiplicative term $(1-\zeta_i^t)$ and the dependence of $p_i(t)$ on $s_i(t-1)$. The following lemma suggests an alternative linear system. This system is not an approximation for the N-intertwined SIR Model, but instead is an upper bound for the probability that node i is infected for all i and all times t. Hence, if we can guarantee that the outbreak is prevented in this new system, we know it is prevented also in the actual N-intertwined SIR model.

Lemma 2. *The system*

$$\tilde{\boldsymbol{p}}(t) = (I + BA - D)\,\tilde{\boldsymbol{p}}(t-1) \tag{6}$$

dominates (4) *in the sense that* $p_i(t) \leq \tilde{p}_i(t)$ *for all* i *and* t.

Proof: We have, for any t, that

$$\begin{aligned} p_i(t) &= (1-\delta_i)\, p_i(t-1) + \\ &\quad s_i(t-1)\left(1 - \prod_{j=1}^N \left(1 - p_j(t-1)\,\beta_i a_{ij}\right)\right) \underset{(a)}{\leq} \\ &\quad (1-\delta_i)\, p_i(t-1) + s_i(t-1) \sum_{j=1}^N p_j(t-1)\,\beta_i a_{ij} \underset{(b)}{\leq} \\ &\quad (1-\delta_i)\, p_i(t-1) + \sum_{j=1}^N p_j(t-1)\,\beta_i a_{ij} \end{aligned} \tag{7}$$

where uses $\prod_{n=1}^N (1-x_n) \geq 1 - \sum_{n=1}^N x_n$ that holds since $0 \leq x_n \leq 1$ for all n, and (b) follows since $s_i(t-1) \leq 1$ for all i. ∎

A linear system is stable if its maximal eigenvalue λ_1 is negative. This leads to the following corollary, which is the main theoretical guarantee of our algorithm.

Corollary 3. *Denote by $\lambda_1(M)$ the maximal eigenvalue of a matrix M. If $\lambda_1(BA-D)<0$ then $\boldsymbol{p}(t)$ vanishes faster than $e^{-t\lambda_1(BA-D)}$.*

A. Greedy link closure algorithm

In this subsection we explain the idea behind Algorithm 1. Since closing a link is a binary option, finding the best link to close among $O\left(N^2\right)$ many links is highly complicated. However, if instead we were to infinitesimally reduce the weight of a link instead of closing it, then it is easy to compute the best link we should pick to infinitesimally decrease λ_1. This idea is formulated in the following lemma.

Lemma 4. *Let A_w be a weighted adjacency matrix. Let λ_1 be the maximal eigenvector of A_w, and v_1 be its corresponding eigenvector such that $v_1^T v_1 = 1$. Let $l=(i,j)$ have a weight w_l. Then*

$$\frac{d\lambda_1(w_l)}{dw_l} = 2v_1(i)\,v_1(j) \tag{8}$$

Proof: Using the Rayleigh quotient to express the maximal eigenvalue

$$\lambda_1(w_l) = \frac{v_1^T A_w v_1}{v_1^T v_1} = v_1^T\left(\sum_{l\in\mathcal{L}} A_l\right) v_1 \tag{9}$$

where for $l=(i,j)$, A_l is a matrix that has $a_{ij}=w_l$ and $a_{ji}=w_l$, and zero anywhere else. The result now follows by differentiating $\lambda_1(w_l)$ with respect to w_l. ∎

B. Large Scale Graphs and Regional Constraints

At first glance it looks like since decisions are made on a regional level, there is no need to do computations on the individual network. However, even if decisions are made on a regional level, the disease spreads on the network of individuals. Hence, in order to estimate the spread rate in each step after making a regional decision, we still must work with matrices of the size of the whole population N.

Fortunately, one can construct a contracted graph G_R that only represents the regions (see Definition 1) but the maximal eigenvalue of its adjacency matrix satisfies $\lambda_1(A_R)>\lambda_1(A)$. Running Algorithm 1 on the regions graph will find a policy that guarantees $\lambda_1(A_RB_R-D_R)<0$, so also $\lambda_1(AB-D)<0$. In other words, by only doing computations on the regional level we can still guarantee that the epidemic is prevented on the individual level.

To formalize this we define the node contraction operation, that combines two nodes into one.

Definition 5. Let $G=(\mathcal{V},\mathcal{L})$. A contraction of vertex v_1 and vertex v_2 into a new node v_1v_2 is the operation that results in the set of nodes $\tilde{\mathcal{V}}=(V\cup v_1v_2)\setminus\{v_1,v_2\}$, and for any $u\in\tilde{V}$, the weight $a_{v_1v_2,u}$ of the edge between v_1v_2 and u is:

- $a_{v_1v_2,u}=a_{v_1u}+a_{v_2u}$ for any $u\neq v_1,v_2$.
- $a_{v_1v_2,u}=2a_{v_1v_2}+a_{v_2v_2}+a_{v_1v_1}$ if $u=v_1$ or $u=v_2$.

The following Lemma guarantees that $\lambda_1(A')>\lambda_1(A)$ after each contraction of G into G'. Since the regions network results from the individual network after enough contractions, then also $\lambda_1(A_R)>\lambda_1(A)$. Moreover, removing nodes and closing links in a contracted graph is equivalent to doing the same operations for the corresponding nodes and links in the original graph. The gap $\lambda_1(A_R)-\lambda_1(A)$ might be very large in practice, so that a policy that achieves $\lambda_1(AB-D)<0$ at a much lower cost is very likely to exist. However, computing and even enforcing such an individualized policy is impossible in practice. Moreover, the main concern is preventing the outbreak. As a bonus, according to Corollary 3, a large gap $\lambda_1(A_R)-\lambda_1(A)$ means that the outbreak vanishes very fast.

Lemma 6. *Let $\tilde{A}$ be an adjacency matrix that results from A by contracting l vertices. Let $\tilde{\lambda}_1$ be the maximal value of $\tilde{A}$. Then $\tilde{\lambda}_1\geq\lambda_1$*

Proof: Let x be any non-zero N dimensional vector. Let $\tilde{x}$ be an $N-1$ dimensional vector and assume without the loss of generality that we contract vertex 1 and vertex 2. The Rayleigh quotient of $\tilde{x}$ is

$$\begin{aligned}\frac{\tilde{x}^T\tilde{A}\tilde{x}}{\tilde{x}^T\tilde{x}} = \frac{\sum_i\sum_j \tilde{a}_{ij}\tilde{x}_i\tilde{x}_j}{\tilde{x}^T\tilde{x}} = \frac{\sum_{i>1}\sum_{j>1}\tilde{a}_{ij}\tilde{x}_i\tilde{x}_j}{\tilde{x}^T\tilde{x}} + \\ \frac{\sum_{j>1}\tilde{a}_{1j}\tilde{x}_1\tilde{x}_j}{\tilde{x}^T\tilde{x}} + \frac{\sum_{i>1}\tilde{a}_{i1}\tilde{x}_1\tilde{x}_i}{\tilde{x}^T\tilde{x}} + \frac{\tilde{a}_{11}\tilde{x}_1\tilde{x}_1}{\tilde{x}^T\tilde{x}} \underset{(a)}{=} \\ \frac{\sum_{i>1}\sum_{j>1}a_{i+1,j+1}\tilde{x}_i\tilde{x}_j}{\tilde{x}^T\tilde{x}} + 2\frac{\sum_{j>1}(a_{1,j+1}+a_{2,j+1})\tilde{x}_1\tilde{x}_j}{\tilde{x}^T\tilde{x}} + \\ \frac{(a_{11}+a_{22}+2a_{12})\tilde{x}_1\tilde{x}_1}{\tilde{x}^T\tilde{x}}.\end{aligned} \tag{10}$$

where (a) follows from Definition 5. Now choose $\tilde{x}$ such that $\tilde{x}_i=x_{i+1}$ for all $i>1$ and $\tilde{x}_1=\sqrt{x_1^2+x_2^2}$. Note that for this choice $x^Tx=\tilde{x}^T\tilde{x}$. Hence

$$\begin{aligned}\frac{\sum_{i>1}\sum_{j>1}a_{i+1,j+1}x_{i+1}x_{j+1}}{x^Tx} + \\ 2\frac{\sum_{j>1}(a_{1,j+1}+a_{2,j+1})\left(\sqrt{x_1^2+x_2^2}\right)x_{j+1}}{x^Tx} + \\ \frac{(a_{11}+a_{22}+2a_{12})\left(x_1^2+x_2^2\right)}{x^Tx} \underset{(a)}{\geq} \\ \frac{\sum_{i>1}\sum_{j>1}a_{i+1,j+1}x_{i+1}x_{j+1}}{x^Tx} + 2\frac{\sum_{j>1}a_{1,j+1}x_1x_{j+1}}{x^Tx} + \\ 2\frac{\sum_{j>1}a_{2,j+1}x_2x_{j+1}}{x^Tx} + \frac{a_{11}x_1^2+a_{22}x_2^2+a_{12}x_1x_2+a_{21}x_2x_1}{x^Tx} \\ = \frac{\sum_i\sum_j a_{ij}x_ix_j}{x^Tx} = \frac{x^TAx}{x^Tx}\end{aligned} \tag{11}$$

where (a) follows since $a_{12}=a_{21}$ and because

$$\begin{aligned}\left(a_{1,j+1}+a_{2,j+1}\right)^2\left(x_1^2+x_2^2\right) = \\ \left(a_{1,j+1}^2+a_{2,j+1}^2+2a_{1,j+1}a_{2,j+1}\right)\left(x_1^2+x_2^2\right) \underset{(a)}{\geq} \\ a_{1,j+1}^2x_1^2+2a_{1,j+1}a_{2,j+1}x_1x_2+a_{2,j+1}^2x_2^2 = \\ \left(a_{1,j+1}x_1+a_{2,j+1}x_2\right)^2\end{aligned} \tag{12}$$

where in (a) we used that $x_1^2 \geq x_1 x_2$ or $x_2^2 \geq x_1 x_2$. Since (11) holds for any x, it also holds for the eigenvector v_1 of λ_1, which shows that

$$\tilde{\lambda}_1 = \max_{y \neq 0} \frac{y^T \tilde{A} y}{y^T y} \geq \frac{\tilde{x}^T \tilde{A} \tilde{x}}{\tilde{x}^T \tilde{x}} \geq \frac{v_1^T A v_1}{v_1^T v_1} = \lambda_1. \qquad (13)$$

■

IV. Simulations on a Data-driven Contact Network

A. N-intertwined SIR Model Calibration from Data

We have used measles incidence data from the Israeli Ministry of Health [23] to calibrate the N-intertwined SIR model parameters and the individual network G. The calibration metric was to minimize the squared error between the model predictions and incidence data of the overall number of infected by subdistrict.

The infection rate β of all nodes was calibrated to fit measles cases data by subdistrict of the last outbreak in Israel between March 2018 and March 2019. To better match the actual outbreak in the calibration process, we have added an exposed state to the SIR dynamics (known as SEIR dynamics). Susceptible individuals can get exposed to measles from their infected contacts with probability β as usual. After the incubation period, exposed individuals become infected with probability $\frac{1}{\sigma}$, where they can now infect their contacts. The average incubation period was taken as 8 days, based on [22] .The recovery rate δ of all nodes was chosen to be $\delta = \frac{1}{8}$ based on [22]. Recovered (vaccinated) individuals are generated based on vaccination coverage data by city [24] and according to vaccine efficacy based on the individual's age group [25], [26]. Individuals for which the vaccination was efficient are removed from the network (not counted in "N").

The contact network was generated based on raw cellular data from Radio Network Controllers (RNCs) covering central Israel. The data includes 17 billion records describing the location of 1.8 million Israeli users over 2 months. The home area of the users was inferred by their location during the night, and their age group was inferred based on proximity to schools. A contact probability matrix was developed, in which each element is the probability that an individual from area i will contact an individual from area j. This probability is a summation over all the areas of the probability that both individuals will visit the same area, multiplied by the proportion of individuals from area j attending the relevant area. In the contact network, each node represents an individual with a home area and an age group, based on the Israel Central Bureau of Statistics demographic data. The number of contacts for each node was generated independently from a geometric distribution, and the contacts between nodes were generated according to the contact probability matrix. This contact network captures spatial and sociodemographic dynamics of the population of Israel. For details, see [27].

In order to adjust the contact network to the measles model age groups, a further age stratification was needed. We adjusted the contact matrix that was used to create the network and stratified it to the required age groups using data from a few worldwide surveys collected in 8 European countries and were projected to other 144 countries, including Israel, using a Bayesian hierarchical model [28]. These data include contact rates between 5-year age groups. After the adjustment of the contact matrix, a 100,000-node contact network was generated as described in [27].

B. Numerical Results

We tested Algorithm 1 on four different contact networks labeled A-D in Table 1. The networks differ in the geographical area they cover (i.e., the subdistricts they include) and in the vaccination coverage. All network were contracted to the statistical area level (e.g. regions, see Definition 1) and the number of such regions is detailed in Table 1. To simplify the interpretability of our results, we assumed that the costs of all vaccinations and all link closures are uniform across the network. The factor ρ is the ratio between the cost of vaccinating a region and that of closing a link. It is proportional to the average degree in the weighted regions graph, since this is the average number of links that are effectively being closed when a node is vaccinated. The proportions are 1, 2 and 20. In all cases, the outbreak was prevented in the individual network G, as guaranteed by our analysis.

Network A covers the subdistricts of Tel Aviv, Jerusalem, Petah Tikva, Ramla, Rehovot, Ashkelon and Judea & Sameria. The average vaccination rate is 96%, according to [24]. Network B is identical to Network A, but with a different randomization of the vaccinated individuals. The similarity of the results on Network A and Network B demonstrates the robustness of our algorithm.

Network C includes only Tel-Aviv district, with a vaccination rate of 96%. We can see that the performance of our algorithm scales roughly linearly with N, since in Networks A, Network B and Network C closing ~30% of the links (when vaccinations are too costly and not used) was enough to prevent the outbreak.

Network D is another instance of Network A, where the vaccination rate was artificially lowered by 10% in every city, modeling a significant vaccine hesitancy. As anticipated, more aggressive intervention was needed to prevent the outbreak.

For benchmark purposes, we compare the performance of Algorithm 1 to two variations where the grade $Q(i,j)$ is modified. In the "Centrality" variation the grade of link i,j is given by $Q(i,j) = d_i + d_j$ where d_i is the degree of node i (in the weighted graph). In the "Random" variation the grade of link i,j is a random variable, e.g., uniform on $[0,1]$, which leads to a random link (or node) selection. However, all these variations use the eigenvalue decomposition of Algorithm 1 and hence rely on both Corollary 3 and Lemma 6. It can be seen that Algorithm 1 is generally better than its centrality variation, especially when only link closure is considered. Both variations are significantly better than random link and nodes selection, which shows that exploiting the topology of the network is important. However, even the random selection uses the topology of the network in its stopping condition.

Network	N	R	#Links	d	ρ	#closed links Algorithm 1	#vaccinated regions Algorithm 1	#closed links Random	#vaccinated regions Random	#closed links Centrality	#vaccinated regions Centrality
A	10138	1696	18250	8.86	177.25	7618	0	9559	0	7887	0
A	10138	1696	18250	8.86	17.70	6751	23	-	-	5768	68
B	9765	1667	17239	8.35	167.10	7129	0	-	-	7391	0
B	9765	1667	17239	8.35	16.71	6418	21	-	-	5615	63
C	862	214	1145	4.68	4.68	188	23	289	34	162	32
C	862	214	1145	4.68	9.36	360	3	535	6	315	10
C	862	214	1145	4.68	93.65	409	0	628	0	442	0
D	14548	1768	30184	6.945	13.89	7869	189	-	-	6932	249
D	14548	1768	30184	6.945	277.88	13578	0	-	-	13860	0

TABLE I
PERFORMANCE ON DATA-DRIVEN NETWORKS

V. CONCLUSIONS

In this paper, we suggested an algorithm that vaccinates some nodes and closes some links in a network in order to prevent a measles outbreak. The algorithm is heuristically trying to minimize the cost of these countermeasures, while provably guaranteeing that the epidemic is prevented. Moreover, we proved that the algorithm can run on a coarse regions network (e.g., cities, districts, neighborhoods) instead of the actual network of all population, and still guarantee that the epidemic is prevented. This is important since decisions are made in practice on a regional level and not per individual, for both fairness and complexity reasons. Our algorithm exhibited good performance on data-driven networks extracted from cellular localization data and measles cases, which are available for decision makers in practice. Therefore, our results are a step forward in closing the gap between theoretical spectral graph results and practical decision making.

ACKNOWLEDGMENTS

This research was supported by the Koret Foundation grant for Smart Cities and Digital Living.

REFERENCES

[1] J. C. Bester, "Measles and measles vaccination: a review," *JAMA pediatrics*, vol. 170, no. 12, pp. 1209–1215, 2016.

[2] V. K. Phadke, R. A. Bednarczyk, D. A. Salmon, and S. B. Omer, "Association between vaccine refusal and vaccine-preventable diseases in the united states: a review of measles and pertussis," *Jama*, vol. 315, no. 11, pp. 1149–1158, 2016.

[3] D. A. Salmon, M. Z. Dudley, J. M. Glanz, and S. B. Omer, "Vaccine hesitancy: causes, consequences, and a call to action," *Vaccine*, vol. 33, pp. D66–D71, 2015.

[4] WHO. (2019) Ten threats to global health in 2019. [Online]. Available: https://www.who.int/emergencies/ten-threats-to-global-health-in-2019

[5] D. A. Spielman, "Spectral graph theory and its applications," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. IEEE, 2007, pp. 29–38.

[6] L. J. Allen, "Some discrete-time si, sir, and sis epidemic models," *Mathematical biosciences*, vol. 124, no. 1, pp. 83–105, 1994.

[7] A. Ghosh and S. Boyd, "Growing well-connected graphs," in *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE, 2006, pp. 6605–6611.

[8] Q. Hu, X. Zhang, and D. Saha, "Modeling virus and anti-virus dynamics in topology-aware networks," in *IEEE Global Telecommunications Conference, 2004. GLOBECOM'04.*, vol. 4. IEEE, 2004, pp. 2077–2081.

[9] B. Buonomo, A. d'Onofrio, and D. Lacitignola, "Global stability of an sir epidemic model with information dependent vaccination," *Mathematical biosciences*, vol. 216, no. 1, pp. 9–16, 2008.

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

[11] F. D. Sahneh and C. M. Scoglio, "Optimal information dissemination in epidemic networks," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 1657–1662.

[12] M. D. Shirley and S. P. Rushton, "The impacts of network topology on disease spread," *Ecological Complexity*, vol. 2, no. 3, pp. 287–299, 2005.

[13] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 17, no. 1, pp. 1–14, 2009.

[14] X. Zhang, J. Wu, P. Zhao, X. Su, and D. Choi, "Epidemic spreading on a complex network with partial immunization," *Soft Computing*, pp. 1–9, 2018.

[15] E. A. Bakare, A. Nwagwo, and E. Danso-Addo, "Optimal control analysis of an sir epidemic model with constant recruitment," *International Journal of Applied Mathematics Research*, vol. 3, no. 3, p. 273, 2014.

[16] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, "How to distribute antidote to control epidemics," *Random Structures & Algorithms*, vol. 37, no. 2, pp. 204–222, 2010.

[17] K. Drakopoulos, A. Ozdaglar, and J. N. Tsitsiklis, "An efficient curing policy for epidemics on graphs," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 2, pp. 67–75, 2014.

[18] L. Mastroleon, *Scalable resource control in large-scale computing/networking infrastructures*. Stanford University, 2009.

[19] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. Pappas, "Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*. IEEE, 2013, pp. 7486–7491.

[20] Y. Wan, S. Roy, and A. Saberi, "Designing spatially heterogeneous strategies for control of virus spread," *IET Systems Biology*, vol. 2, no. 4, pp. 184–201, 2008.

[21] H. Gans, L. Yasukawa, M. Rinki, R. DeHovitz, B. Forghani, J. Beeler, S. Audet, Y. Maldonado, and A. M. Arvin, "Immune responses to measles and mumps vaccination of infants at 6, 9, and 12 months," *The Journal of infectious diseases*, vol. 184, no. 7, pp. 817–826, 2001.

[22] G. H. Dayan and H. Q. McLean, "Immunology and vaccine preventable diseases, pink book , measles," *International Encyclopedia of Public Health*, pp. 565–569, 2016.

[23] M. of Health Israel. (2019) Weekly epidemiological reports,. [Online]. Available: https://www.health.gov.il/UnitsOffice/HD/PH/epidemiology/Pages/epidemiology_report.aspx

[24] ——. (2018) Measles vaccination coverage. [Online]. Available: https://gis.health.gov.il/vaccine/

[25] ——. (2019) Measles - vaccination tips, ministry of health. [Online]. Available: https://www.health.gov.il/English/Topics/Vaccination/Pages/Measles.aspx/

[26] CDC. (2018) Measles vaccination. [Online]. Available: https://www.cdc.gov/measles/vaccination.html

[27] D. Kahana and D. Yamin, "Accounting for the spread of vaccination behavior to optimize influenza vaccination programs," 2019. [Online]. Available: SSRN:https://www.ssrn.com/abstract=3386213

[28] K. Prem, A. R. Cook, and M. Jit, "Projecting social contact matrices in 152 countries using contact surveys and demographic data," *PLoS computational biology*, vol. 13, no. 9, p. e1005697, 2017.