Designing Spatially-Heterogeneous Strategies for Control of Virus Spread

Yan Wan ∗ Sandip Roy ∗ Ali Saberi ∗

March 13, 2007

Abstract

The spread of viruses in human populations (e.g., SARS) is closely related to the spatial topology of the network of interactions within the population. In this paper, we study the problem of allocating limited control resources (e.g., quarantine or recovery resources) in these networks in a way that exploits the topological structure, so as to maximize the speed at which the virus is eliminated. For both multi-group and automaton models for spread, this problem can be abstracted to that of designing diagonal $K$ or $D$ to minimize the dominant eigenvalue of one of the system matrices $KG$, $D + KG$, or $D + G$ under constraints on $K$ and $D$ (where $G$ is a square matrix that captures the network topology). We give explicit solutions to these problems, using eigenvalue sensitivity ideas together with constrained optimization methods employing Lagrange multipliers. Finally, we show that this decentralized control approach can provide significant advantage over a homogeneous control strategy, in the context of a model for SARS transmission in Hong Kong.

∗All authors are with the School of Electrical Engineering and Computer Science, Washington State University. The authors can be reached at \{ywan, sroy, saberi\}@eecs.wsu.edu, respectively. The authors are grateful for the support provided by the National Science Foundation (under Grant ECS0528882) and the National Aeronautics and Space Administration (under Grant NNA06CN26A).
1 Problem Formulation and Motivation

The significant impacts of epidemics in recent years highlight the need for controlling virus spread with limited resources [1, 2]. Here, we put forth the perspective that spatially-heterogeneous control strategies can allow mitigation of virus spread with sparse resources. Thus, we pose the virus-spreading control problem as a constrained decentralized design task for a dynamic network model, and give an analytical methodology for completing the design task. We apply our method to design spatially-heterogeneous controls for Hong Kong’s 2003 SARS outbreak (see [3]), which outperform the existing homogeneous controls considered in the literature.

While our primary focus is on virus-spreading control, this work also constitutes a significant contribution to our ongoing efforts in decentralized controller design. Recently, researchers in such fields as autonomous vehicle control and sensor networking have recognized the need for decentralized algorithms/controllers that exploit a network’s topological structure (see e.g. [4], see also the review article [5]). The existence of stabilizing decentralized controllers for these applications can be studied using the seminal work of Wang and Davison [6], but the design of practical but high performance controllers remains difficult. In [7], we posed an optimal decentralized design problem for a simple class of network dynamics, and developed a design tool using optimization machinery together with eigenvalue sensitivity and graph algebra notions. Our efforts here show that similar tools can be developed for a family of decentralized design problems including some constrained ones. Our results also highlight that particularly simple and structurally-insightful design tools can be obtained for certain special classes of network topologies, such as ones with non-negative weights.

The article is organized as follows. In the remainder of this section, we review two models for computer-virus and biological-virus spread, namely a multi-group model for spatially inhomogeneous populations, and an automaton model for interactions of individuals. In turn, we formulate several spread control problems as decentralized design problems. In Section 2, we develop a methodology for solving these decentralized design problems in detail. Finally, in Section 3, we apply our method to synthesize controllers for the SARS outbreak in Hong Kong.
1.1 Epidemic Control: Brief Review

Mathematical epidemiology has a history of more than two centuries. One major focus of the mathematical work on epidemics is characterization of the basic reproduction ratio $R_0$ (defined as the average number of secondary infections produced during an infected individual’s infectious period, when the individual is introduced into a population where everyone is susceptible [9]). It is well known that, for $R_0 > 1$ a disease can spread throughout the population and may eventually reach an equilibrium, while for $R_0 < 1$ the epidemic eventually terminates. The basic reproduction ratio can be computed from models and also found experimentally, see e.g. [3,9,11–13].

Control can be viewed as reducing $R_0$, and hence stopping the spread of a virus. Common control methods include: 1) vaccination; 2) reduction of local contact rates; 3) shortening of the time between symptom appearance and hospitalization; 4) restriction on long-range movement; 5) isolation of symptomatic people, and those in contact with them (quarantine) [3,12,13]. All these control schemes change one or more parameters of the epidemic model of interest. Hence, by analyzing $R_0$ or simulating dynamics over a parameter range [3, 12], we can analyze the impact of different control schemes. For example, the strategy of vaccinating newborns in a heterogeneous population (assuming contact rates are only at two levels) was studied in [9, 14], where an age-related model was used.

1.2 Spatial Control in a Multi-Group Model: Formulation

Spatial interaction structure is critical in epidemics (e.g., SARS [3,15]). However, there is little work in the literature on designing controls (e.g. isolation and quarantine), that are specialized to the network structure, to optimally fight against epidemics with limited resources. We believe such a systematic design of control parameters at different points in the network can provide guidance for effective epidemic control. Here we review spatially inhomogeneous models for epidemics, and so pose the optimal spread control problem.

Early epidemic modeling assumes homogeneous mixing, i.e. any pair of individuals are assumed equally likely to interact (or equivalently the strengths of the interactions are the same). In reality, populations are spatially heterogeneous, and in fact this spatial
structure (social interaction topology) of the population plays an important role in epidemic spread. Usually, multi-group models (models in which the population is composed of multiple interacting groups, which internally have homogeneous mixing) are used to represent the spatially inhomogeneous dynamics [3,12,16]. Of particular interest to us, references [17, 18] show how to calculate $R_0$ in a heterogeneous population, from the multi-group model. In particular, $R_0^1$ is the dominant eigenvalue of the next generation operator, the elements of which are defined as the the expected numbers of new infections within a group that are produced by one infective with another group during its infectious period. Use of the next generation matrix to calculate $R_0$ can be found in [3] and [12]. Often, stochastic models for epidemic spreads are also used because chance fluctuations can be large, especially in the early stages of an epidemic [3,13]. Here, we consider a distributed control paradigm in a multi-group model for spatial inhomogeneity.

Spatial inhomogeneous models have already been used to capture specific epidemics, and to evaluate homogeneous control of such epidemics. For instance, the outbreak of severe acute respiratory syndrome (SARS) in 2003 aroused a lot of interest in spatial modeling and control [3, 10, 13], because of the geographical patterns observed in the virus spread [3,13,15].

Of interest to us, the article [3] models SARS in Hong Kong using a stochastic multiple-group model, where each group corresponds to a (spatial) district in Hong Kong. The authors identify the basic reproduction ratio $R_0$ for the model, and show how homogeneous (identical network-wide) control can be used to reduce $R_0$ to 1. In Section 3, we show that our optimal control which exploits network structure can reduce $R_0$ further with the same amount of resource, or equivalently achieve $R_0 = 1$ with less resource.

Let us now formulate the multi-group epidemic model and associated control design problem. Since we are primarily interested in spatial spread of epidemics, we shall refer to the $n$ groups (labeled 1,..., $n$) as districts (meaning geographic regions), though the design can be used for other multi-group models also. In this sense, our model is a generalization of the model in [3] that permits inhomogeneous control, and also can be viewed as a special case of the model in [17,18] where groups correspond to spatial regions.

$^1R_0$ is usually real since an interaction network structure is most often non-negative and irreducible. We limit ourselves to the real maximum eigenvalue case in our later design.
Specifically, in our model, the nominal average rate (average rate without control) at which an individual in District \(i\) becomes infected is \(\sum_j \beta h_{ji} \lambda_j(t)\), where \(\beta\) is transmission coefficient (incorporating infectiousness and average contact rate), each \(h_{ji}\)\(^2\) is a corrective term that accounts for the relative rate of inhomogeneous mixing between District \(i\) and \(j\), and \(\lambda_j(t)\) represents the infectiousness of region \(j\) at some time and roughly scales with the infected population (see [3]). We also assume nominally that an individual requires time \(T\) to heal.

We consider three sorts of control that deviate from this nominal: 1) We allow the change of the contact rate of individuals in District \(i\) by a factor of \(r_i \in [0, 1]\) which decreases spread of virus both locally and to spatial neighbors. Note \(r_i\) can be reduced by isolation of closely connected and fairly isolated groups such as a school/college. 2) we allow change of the contact rate of an individual from outside districts to a District \(i\) by a factor of \(c_i\). The external contact rate factor \(c_i \in [0, 1]\) can be reduced by prohibition of travel from another district to District \(i\), or similarly isolation of arriving travelers for some days (i.e. days longer than the incubation period). 3) We allow the average duration of infectious period of each District \(i\) to deviate from \(T\) by a factor of \(T_i\). The factor \(T_i\) can be reduced by shortening the time between symptom appearance and hospitalization in District \(i\). Control measures such as isolation of people who may have contacted an infected individual (i.e. isolation of a neighborhood with infected people) or isolation of symptomatic people (possibly some individuals have false symptoms and are not infected) reduce both \(T_i\) and \(r_i\). Upon control, the rate of infection for region \(i\) becomes \(\beta r_i (h_{ii} \lambda_i(t) + c_i \sum_{j \neq i} h_{ji} \lambda_j(t))\). Based on these rates of infection and the infectious duration \(T_i\), the next generation matrix is well known to be the following\(^3\) (see [3,17,18]):

\(^2\)In [3], \(h_{ji}\) is denoted as \(K_{ji}\). Since in the field of control, \(K\) is usually used as a scaling factor as in our paper, we use \(h_{ji}\) to indicate the corrective term simply to avoid confusion.

\(^3\)In finding the next-generation matrix, it is assumed that the susceptible population in a district is well-approximated by the entire population in that district.
\[ A = \beta T \, \text{diag}(T_i r_i N_i) \times \left( \text{diag}(h_{ii}) + \text{diag}(c_i) \right) \]

\[ \begin{bmatrix}
0 & h_{21} & \ldots & h_{n1} \\
h_{12} & 0 & \ldots & h_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
h_{1n} & h_{2n} & \ldots & 0
\end{bmatrix} \] (1)

Let us now formally pose the controller design problem. In doing so, note that one very reasonable performance measure is the dominant eigenvalue of the next generation matrix, which represents the spread rate of the epidemic. Let’s say we are interested in designing \( T_i \) and/or \( r_i \). Noting that we can write the next generation matrix as \( A = KG \), where

\[ G = \beta T \, \text{diag}(N_i h_{ii}) + \beta T \, \text{diag}(N_i c_i) \]

and \( K = \text{diag}(T_i r_i) \), we can view the design problem as that of finding a diagonal matrix \( K \) so as to minimize \( \lambda_{\text{max}}(KG) \), subject to constraints that \( 0 \leq K_i \leq 1 \) and that the \( K_i \) are not too small in total (since much resource is needed to make \( K_i \) small). Here is a formal statement:

**Problem 1.** Design diagonal matrix \( K \), such that \( \lambda_{\text{max}}(KG) \) is minimized, where \( K \) is subject to the following constraints:

1. \( \text{tr}(K) = \sum K_i \geq \Gamma \)
2. \( 0 \leq K_i \leq 1 \) for all \( i \).

We may alternately consider a constraint based on resource cost increasing inversely with \( K_i \), i.e. \( \sum \frac{1}{K_i} \leq \Gamma \).

In the case that we restrict long distance movement (i.e. movement between districts) to design \( c_i \), we can write the next generation matrix as \( A = D + KG \), where \( G = \)
\[
\begin{bmatrix}
0 & h_{21} & \ldots & h_{n1} \\
h_{12} & 0 & \ldots & h_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
h_{1n} & h_{2n} & \ldots & 0
\end{bmatrix},
\]

\[K = \text{diag}(c_i)\text{ and } D = \beta T \text{diag}(T_i r_i N_i h_{ii}).\]

In this case, we can formulate the design problem as follows:

**Problem 2.** Design diagonal matrix \(K\), such that \(\lambda_{\text{max}}(D + KG)\) is minimized, where \(K\) is subject to the constraints that

1. \(tr(K) \geq \Gamma\)
2. \(0 \leq K_i \leq 1\) for all \(i\).

### 1.3 Inhomogeneous Control in an Automaton Model

Automaton models for virus spread—those in which *individuals’* infection states (or state probabilities) are tracked—have been used to model cell-to-cell spread of influenza [19], SARS propagation [20], and computer virus spread [21], among other applications. Automaton models are motivated by the observation that homogeneity usually does not exist in real populations, perhaps not even within small groups. Automaton models are thus appealing in that they can capture the specific network interactions among individuals. It is worth noting that, within the general framework of automaton models, considerable research is focused on special classes of network topologies (e.g., scale-free, small-world, correlated, mesh) [1, 22–24].

An automaton model defined for a general network topology was proposed in [25] (which is motivated by the computer virus application). This paper approximated the epidemic threshold (a threshold on the infection rate to curing rate ratio, above which epidemic occurs, i.e. such that \(R_0 = 1\)) as the inverse of the dominant eigenvalue of the network’s adjacency matrix, assuming that (at each discrete time step) an infected node infects its adjacent node with a common probability and is cured with a different common probability. However, because the interaction probabilities are identical, this paper does not provide us with insight into topology-based network design, which can potentially lower the network’s vulnerability to virus spread. Our work here pursues
Few works have studied anti-virus network design. The article [1] proposed a targeted immunization strategy (a few nodes with the highest connectivity are immunized) for power-law networks, and evaluated its performance using simulation. The article [2] concluded that selective immunization (e.g. immunizing the upper-level nodes in a tree-like topology, or nodes with high connectivity) significantly reduces a network’s vulnerability to virus attack compared to random immunization. However, this work is also built on simulation, and hence does not provide us with an immunization strategy that meets a performance requirement.

We develop network resource allocation strategies that optimize spread-based performance requirements (e.g., epidemic diminishing rate, number of nodes affected, and the total duration of the epidemic), with the motivation that such design will aid in defending networks against virus attacks. In our effort, the network parameters (e.g., the local curing rates and infection rates) from [25] have the flexibility of design. For example, providing a selected set of individuals/nodes with faster treatment (or, in the case of computer viruses, better virus scan softwares) can increase these nodes’ local curing rates. Similarly, providing antibiotics to individuals (equivalently, providing computers with strong firewalls) can safeguard these nodes from common viruses, or at least lower the rate of infection from their neighboring nodes. Each of these control actions is associated with a cost (e.g., financial cost, productivity loss). Thus, it is not realistic to immunize or provide real-time repair to every individual in a network. Instead, we must assign limited control resources to achieve the best performance. Our study indicates how resources can be allocated in a way that appropriately uses the network topology.

To pursue control, we build on the automaton model proposed in [25], with the motivation that this model has already been of interest in studying resource allocation. Our model is a generalization of [25], in that we allow variation in local curing rates and infection rates throughout the network. As in [25], we model virus spread as a discrete-time dynamics defined on a directed graph. Each node \((i \in 1, ..., n)\) represents an individual (node) in the network, which may either be infected or susceptible. Each directed edge
represents a path along which a virus can spread from one node to another. The branch weight $\beta_{ij}$ represents the probability that virus originating from node $i$ spreads to node $j$ during a time step. Notice that $\beta_{ij}$ increases with the transmission rate from node $i$ to $j$ and the infectiousness of the virus, and decreases as protection at node $j$ is increased (e.g., through giving antibiotics to humans, or providing firewalls to computers). We set $\beta_{ij} = 0$ if node $j$ is not a neighbor to $i$ (node $i$ cannot transmit an infection to $j$) or node $j$’s protection prevents any infection. An infected node has probability $\delta_i$ to recover at a discrete time step. We consider two possible control actions in our model. 1) We allow for control that makes a node $j$ less susceptible to any virus spread. In this case, we assume that the weights $\beta_{ij}$ are scaled by a constant for all entering branches), i.e. the weights become $k_j \beta_{ij}$, $k_j \in [0,1]$. We note that decreasing $k_j$ from 1 is costly. 2) We allow control of the recovery rate $\delta_i$. We note that increasing the recovery rate is expensive, in that more medicine or quicker hospitalization is needed (better virus removal programs or quicker human intervention, respectively, for computer network applications).

Now let us analyze the network’s dynamics. Denoting the probability that each node $i$ is infected at time $k$ as $p_i[k]$, we find that the probability the node is infected at time $k+1$ is

$$p_i[k+1] = \left(1 - \prod_{j} (1 - K_i \beta_{ji} p_j[k])\right) + (1 - \delta_i) p_i[k].$$

(2)

Assuming $K_i \beta_{ji} p_j[k] << 1 \forall i, j, k$ (which is accurate for small time steps), the quantity $1 - \prod_j (1 - K_i \beta_{ji} p_j[k])$ can be well approximated by $\sum_j K_i \beta_{ji} p_j[k]$, and thus we can linearize (2) to obtain the network dynamics

$$P[k+1] = (D + KG)P[k]$$

(3)

where $G = \begin{bmatrix} 0 & \beta_{21} & \cdots & \beta_{n1} \\ \beta_{12} & 0 & \cdots & \beta_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{1n} & \beta_{2n} & \cdots & 0 \end{bmatrix}$, $D = \text{diag}(1 - \delta_i)$, $K = \text{diag}(K_i)$, and $P[k] = [p_1[k] \ p_2[k] \ \cdots \ p_n[k]]^T$. In the case where only $\delta_i$ are being designed, we find it convenient to use the notation that $P[k+1] = (D + G)P[k]$, since the $K_i$ are fixed.
The diagonal matrices $D$ and $K$ are the ones we have ability to design. Our aim is to design $D$ and $K$ so that the network structure best inhibits virus spread, and hence minimizes epidemic size. This goal leads us to consider optimization with respect to a performance measure. The one we consider here is the dominant eigenvalue of $D + KG$ (denoted as $\lambda_{\text{max}}(D + KG)$), as we know that the dominant eigenvalue governs the growth/decay rate of infection.

The matrix $D$ contains the local recovery rate of each node. Increasing $\delta_i$ (or equivalently decreasing $D_i$) can speed up the elimination of a virus, but at higher resource cost. Therefore, we aim to design $D$ so that the cost of control (i.e., sum of $\delta_i$) is under a limit, while the performance of the design is optimized. Similarly, matrix $K$ represents the virus protection strength for each node. Decreasing $K_i$ for more nodes can also speed up the elimination of a virus. We thus design $K$ so that the cost of control is less than a threshold (i.e., sum of $K_i$ is over a limit), while the performance of the design is optimized.

Let us pose these network design problems formally:

**Problem 3.** Design diagonal matrix $D$, such that $\lambda_{\text{max}}(D + G)$ is minimized, while $D$ satisfies the constraints:

1. $tr(D) \geq \Gamma$

2. $0 \leq D_i \leq 1$ for all $i$.

**Problem 4.** Design diagonal matrix $K$, such that $\lambda_{\text{max}}(D + KG)$ is minimized, while $K$ is satisfies the constraints:

1. $tr(K) \geq \Gamma$

2. $0 \leq K_i \leq 1$ for all $i$.

## 2 Network Design

In this section, we address the design problems formulated in Section 1, namely to design diagonal $D$ to minimize $\lambda_{\text{max}}(D + G)$, and to design $K$ to minimize $\lambda_{\text{max}}(KG)$ and
\( \lambda_{\text{max}}(D + KG) \), subject to the described constraints. We give methods for finding the optimal resource allocations both for general topologies \( G \), and for specific classes of topologies that are common in virus-spreading applications. We describe the solution to the \( D + G \) case first, since the full suite of results is easier to describe/interpret in this case.

### 2.1 Designing \( \lambda_{\text{max}}(D + G) \)

We address the problem of designing diagonal \( D \) such that \( D + G \) has minimum dominant eigenvalue, subject to the constraints that \( 0 \leq D_i \leq L \) and \( tr(D) \) is lower-bounded. For convenience, we refer to the optimum \( D \) as \( D^* \), the dominant eigenvalue of \( D^* + G \) as \( \lambda_{\text{max}}^* \), and the corresponding left- and right- eigenvectors as \( w_{\text{max}}^* \) and \( v_{\text{max}}^* \). Our design method is founded on the observation that the optimized topology \( D^* + G \) has a very special eigenstructure, based on which we can compute \( D^* \) and find the optimal performance. We begin with the structural result:

**Theorem 1**  
Consider a matrix \( D + G \), where \( D \) is diagonal and \( G \) is an \( n \times n \) matrix, and consider \( D = D^* \) that minimizes the dominant eigenvalue of \( D + G \) subject to the constraints 1) \( \sum D_i \geq \Gamma \) and 2) \( D_i \in [0, L] \), and assuming the dominant eigenvalue of \( D + G \) is real and non-repeated\(^4\), the optimizing \( D^* \) and corresponding eigenvalue/eigenvectors \( \lambda_{\text{max}}^*, w_{\text{max}}^* \) and \( v_{\text{max}}^* \) (appropriately normalized) satisfy one of the following conditions:

- Either: \( \sum D_i^* = \Gamma \) and for each \( i \) we either have \( 0 < D_i^* < L \) and \( w_{\text{max}}^*, v_{\text{max}}^* = 1 \), or we have \( D_i^* = L \) or \( D_i^* = 0 \).
- Or: \( \sum D_i^* < \Gamma \), and for each \( i \) we either have \( 0 < D_i^* < L \) and \( w_{\text{max}}^*, v_{\text{max}}^* = 0 \), or we have \( D_i^* = L \) or \( D_i^* = 0 \).

**Proof:** This result follows from standard theorems on eigenvalue sensitivity [26], as well as theorems on constrained optimization using Lagrange multipliers [27]. Let us denote \( D_i = d_i^2 \), since we require \( D_i \geq 0 \) for all \( i \). The procedure for finding the optimum \( D^* \) under constraint is to form the Lagrangian \( L = \lambda_{\text{max}}(D + G) + \sum a_i(d_i^2 + m_i^2 - L) - \)

\(^4\)The theorem can be easily generalized to the case that \( D + G \) has real and simple dominant eigenvalues.
\[ C(\sum d_i^2 - n^2 - \Gamma) \]

and set the derivatives of it with respect to all variables (namely \( d_i, a_i, m_i, C \) and \( n_i \)) to 0. (Here, \( m_i \) and \( n \) are slack variables to transform inequality constraints to equality constraints). This procedure leads to the equations below:

\[
\begin{align*}
  d_i^*(w_{max,i}^* v_{max,i}^* + a_i^* - C^*) &= 0 \\
  d_i^{*2} + m_i^{*2} &= L \\
  m_i^* a_i^* &= 0 \\
  n^* C^* &= 0 \\
  \Gamma + n^{*2} &= \sum d_i^{*2}
\end{align*}
\]

Note that the first equation above follows from the eigenvalue sensitivity formula. The two cases in the theorem thus follow automatically from consideration of the variables \( n^* \) and \( C^* \), one of which must be 0. Specifically, the case where \( \sum D_i^* = \Gamma \) follows from setting \( n^* \) to 0, while the case where \( \sum D_i^* > \Gamma \) follows from setting \( C^* \) to 0. □

Theorem 1 tells us that the optimum \( D^* \) can be either at or inside the constraint boundaries. When \( D^* \) is at the boundary \( \sum D_i^* = \Gamma \), each \( D_i^* \) falls into one of the three categories: 1) \( D_i^* = L \); 2) \( D_i^* = 0 \); and 3) \( w_{max,i}^* v_{max,i}^* = 1 \). When \( D^* \) is not at the boundary \( \sum D_i^* = \Gamma \), each \( D_i^* \) again is at 0 or \( L \), or \( w_{max,i}^* v_{max,i}^* = 0 \). We notice that, for any condition in one of the above forms, the number of equations and variables are equal, and so we can get a potential optimal solution from these equations. However, the number of potential optimal solutions (number of solutions which satisfy one set of equations of this type) grows exponentially with the dimension of the matrix \( G \), and so the calculation will be very complicated for even moderate-sized \( G \). In the rest of this section, we will show that when \( G \) is specially structured—e.g., non-negative or symmetric, we can develop more explicit (and hence easier-to-evaluate and interpret) expressions for the optimal solution.

In the following Theorem 2, we show that for a very broad class of topology matrices \( G \), the optimizing \( D \) is one that uses maximum total resource, i.e. one for which \( \sum_i D_i = \Gamma \).

**Theorem 2** Consider \( D + G \), where \( D \) is diagonal, \( G \) is an \( n \times n \) matrix, and the
The largest eigenvalue of $D + G$ is real and non-repeated for all $D$ such that 1) $\sum D_i \geq \Gamma$ and 2) $D_i \in [0, L]$. The matrix $D^*$ that minimizes the dominant eigenvalue of $D + G$ subject to these constraints satisfies $\sum D_i^* = \Gamma$, if the left and right eigenvectors of $D + G$ corresponding to the dominant eigenvalue have the same sign patterns for all $D$. Classes of matrices satisfying this condition include 1) irreducible non-negative matrices and 2) diagonally symmetrizable matrices (matrices for which there exists diagonal $Q$ such that $Q^{-1}GQ$ is symmetric).

**Proof:** The condition that the left and right eigenvectors of the dominant eigenvalue have the same sign pattern implies the relationship that, for all $i$, $w_{max,i}v_{max,i} > 0$. Thus, according to the eigenvalue sensitivity theorem, the dominant eigenvalue decreases monotonically with the decrease of $D_i$ for all $i$, since $\frac{\partial \lambda_{max}(D+G)}{\partial D_i} = w_{max,i}v_{max,i}$ is positive. Therefore, the optimum $D^*$ is on the boundary $\sum D_i^* = \Gamma$. From the Perron Frobenius Theorem, the dominant eigenvalue of any non-negative and irreducible matrix is real and non-repeated, and the left and right eigenvectors associated with the dominant eigenvalue are positive [28], and hence have the same sign pattern. For a diagonally symmetrizable $G$, it is easy to check through a similarity transform that the eigenvalues are real and the left and right eigenvectors associated with any eigenvalue are identical related by a positive diagonal scaling and hence have the same sign pattern. Hence, the theorem is proved. □

Theorem 2 guarantees the optimum $D^*$ is located on the boundary $\sum D_i^* = \Gamma$, whenever $G$ is an irreducible and non-negative square matrix, and hence simplifies the search for $D^*$ when $G$ has this special structure. This simplification is relevant to our applications, because for both the multi-group and automaton models, $G$ is non-negative and (for meaningful interaction topologies) irreducible. In the illustrating example, $G$ is irreducible and non-negative, so we expect that the optimum $D^*$ satisfies $\sum D^* = \Gamma$. In this case, $D^*$ can be found by searching only through the first set of possibilities in Theorem 1. This search is formulated in Theorem 3 for some matrices of this sort. Before that, in Lemma 1, we characterize the pattern of the eigenvector associated with the dominant eigenvector under constraints $D_i \in [0, L]$ and $\sum D_i \geq \Gamma$. 
Lemma 1  Consider the matrix $D + G$, where $D$ is diagonal, and $G$ is an $n \times n$ irreducible non-negative symmetric matrix. A matrix $D^* = D^\star$ minimizes the dominant eigenvalue of $D + G$ subject to the constraints $\sum D_i \geq \Gamma$ and $D_i \in [0, L]$ if and only if the eigenvector $v$ associated with the dominant eigenvalue of $D^* + G$ has the pattern $v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=L) < v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L) < v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=0)$, and $v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L)$ are identical.

Proof: First let us show necessity. Suppose $v_{\text{max}}^*$ is the eigenvector associated with the dominant eigenvalue of $D^* + G$. The conclusion that $v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L)$ are identical directly follows from Theorem 1, and the symmetry and non-negativity of $G$. More specifically, when $G$ is a symmetric matrix, so is $D^* + G$, and thus $v_{\text{max},i} = w_{\text{max},i}$ for all $i$. Also, from Theorem 1, we know that the eigenvectors $v_{\text{max}}^*$ and $w_{\text{max}}^*$ of $D^* + G$ satisfy $v_{\text{max},i}^* = w_{\text{max},i}^*$ and $v_{\text{max},i}^*w_{\text{max},i}^* = 1$ for each $i$ such that $0 < D_i^* < L$. Finally, from positivity of $G$, we see that $v_{\text{max}}^*$ has only positive entries. Combining, we recover that $v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L)$ are identical. Now we need to show that $v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=L) < v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L) < v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=0)$. Since $D^*$ achieves the minimum dominant eigenvalue, decreasing $D_i^*$ for $i$ such that $D_i^* = L$ (making them less than $L$) or increasing $D_i^*$ for $i$ such that $D_i^* = 0$ (making them larger than 0) while maintaining $\sum D_i = \Gamma$ should increase $\lambda_{\text{max}}$. Eigenvalue sensitivity naturally leads to the inequality of eigenvector components, since the derivative of $\lambda_{\text{max}}$ with respect to $D_i$ equals $v_{\text{max},i}^2$.

For the sufficient condition, we know that if $v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=L) < v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L) < v_{\text{max},i}(\forall \ i, \ s.t. \ D_i=0)$ and $v_{\text{max},i}(\forall \ i, \ s.t. \ 0<D_i<L)$ are the same, the corresponding $D$ achieves a local minimum (from above). It follows from convexity (which can be proved easily using e.g. the Courant-Fisher theorem) that the local minimum is in fact global. □

Lemma 1 presents the pattern of eigenvectors associated with the minimized dominant eigenvalue of $D + G$. This allows us to check whether a solution $D$ is optimum by simply evaluating the dominant eigenvectors of $D + G$. 
Theorem 3  Consider a topology matrix $G$ that is non-negative, irreducible, and diagonally symmetrizable. The matrix $D^*$ that minimizes the dominant eigenvalue of $D + G$ can be found using the following algorithm:

1. Find diagonal matrix $Q$ such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as $\hat{G}$.

2. Choose some $D_i$ as 0 and some $D_i$ as 1, and rearrange the rows and columns of $D + \hat{G}$ in such a way that those $D_i$ fixed at 0 and 1 are in the lower right corner. The permuted matrix can be decomposed as $\begin{bmatrix} \hat{G}_{11} + D_A & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} + D_B \end{bmatrix}$, where the matrix $D_A$ is unknown, and the matrix $D_B$ has entries fixed at 0 or 1.

3. Solve the equation $-1^T\hat{G}_{11}1 - 1^T\hat{G}_{12}(\lambda I - \hat{G}_{22} - D_B)^{-1}\hat{G}_{21}1 + 1^T\lambda 1 = \Gamma - tr(D_B)$ for $\lambda$. This can be done through a simple numerical procedure.

4. Calculate $D_A$ using $D_A = \text{diagonalize}(-\hat{G}_{11}1 - \hat{G}_{12}(\lambda I - \hat{G}_{22} - D_B)^{-1}\hat{G}_{21}1 + \lambda 1)$. If $0 \leq D_A \leq LI$, and the pattern of eigenvector associated with the dominant eigenvalue of $D + \hat{G}$ follows Lemma 1, the optimum $D$ is $\text{diag}(D_A, D_B)$. Otherwise, go to step 2) until a solution is achieved.

Proof: We know from similarity that the eigenvalues of $D + G$ are same as that of $D + \hat{G}$, where $\hat{G} = Q^{-1}GQ$ and $Q$ is the positive diagonal matrix such that $\hat{G}$ is symmetric. Hence, we can without loss of generality consider designing $D^*$ to minimize $D^* + \hat{G}$. Since all $v_{\text{max}}$ (of $D^* + G$) whose corresponding $D_i^*$ are not fixed at 0 or 1 are equal (let us normalize them to 1), we know the optimum $D$ satisfies $\begin{bmatrix} \hat{G}_{11} + D_A & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} + D_B \end{bmatrix} \begin{bmatrix} 1 \\ v \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ v \end{bmatrix}$, where $D_B$ contains the entries in $D$ that are 0 and 1. $D_A$ must be found, and $\hat{G}_{11}$, $\hat{G}_{12}$, $\hat{G}_{21}$, and $\hat{G}_{22}$ are submatrices of $\hat{G}$ (upon appropriate permutation). A little bit of algebra leads to the solution of $\lambda$, $D$ and $v$. If the pattern of the eigenvector associated with the dominant eigenvalue of $D + \hat{G}$ follows Lemma 1, and the $D$ matrix is within the constraints $0 \leq D \leq LI$, we have found a global optimum solution according to Lemma 1. □
Although we have presented an algorithm for diagonally-symmetrizable $G$, a slightly more complicated algorithm exists for all $G$ that are non-negative matrices; we omit this algorithm here in the interest of space. We also note that the above algorithm may be computationally intensive, in that the steps may have to be repeated up to $3^n$ times to find the optimum. For some topology matrices $G$, the calculation of $D^*$ can further be simplified. We will describe the procedure to calculate $D^*$ under these circumstances in Theorem 4. As a preliminary step, let us first explicitly compute the optimal $D$ when the constraints on individual $D_i$ are relaxed:

**Lemma 2** First, we notice all entries of $\bar{v}_{max}$ are identical, based on Lemma 1 and the fact that here only the constraint $\sum D_i \geq \Gamma$ is enforced. Thus, the optimizing $\bar{D}$ and eigenvalue satisfy $(\bar{D} + G)1 = \bar{\lambda}_{max}1$. Therefore, the $\bar{D}_i$’s make the row sums of $D + G$ equal. $\bar{D}_i$ also satisfies $\sum \bar{D}_i = \Gamma$, since $G$ and thus $\bar{D} + G$ is symmetric. A little bit of algebra leads to the expressions for $\bar{\lambda}_{max}$ and $\bar{D}_i$. □

Lemma 2 states that, without the constraints that $D_i \in [0, L]$, the optimum $D$ (denoted $\bar{D}$) is the one that equalizes the row sum of $D + G$, i.e., resource is allocated to each part of the network so as to make all their impacts identical. However, when the individual $D_i$ are constrained, sometimes the optimum $\bar{D}$ can not be reached. Building on Lemma 2, Theorem 4 illustrates an easy way to find the optimum $D$ under several circumstances.

**Theorem 4** Consider a matrix $D + G$, where $D$ is diagonal, and $G$ is an $n \times n$ non-negative, irreducible, and diagonally-symmetrizable matrix. We can find $D = D^*$ that minimizes the dominant eigenvalue of $D + G$ subject to the constraints 1) $\sum D_i \geq \Gamma$ and 2) $D_i \in [0, L]$ using the following algorithm:

1. Find diagonal $Q$ such that $Q^{-1}GQ$ is symmetric. We name $Q^{-1}GQ$ as $\hat{G}$.
2. Calculate $\bar{\lambda}$ and $\bar{D}_i$ following Lemma 2, and check if $0 \leq \bar{D}_i \leq L$ for all $i$. Denote the set of indices $i$ such that $\bar{D}_i > L$ as $\mathcal{L}^+$ and the set such that $\bar{D}_i < 0$ as $\mathcal{L}^-$. If $\mathcal{L}^+ = \phi$ and $\mathcal{L}^- = \phi$ (i.e., both sets are empty), we have $\lambda^*_{max} = \bar{\lambda}$ and $D^*_i = \bar{D}_i$. 
3. If $L^+ \neq \phi$, and $L^- = \phi$, we set $D_i = L$ for $i \in L^+$. That is, $D_i$ can be calculated by first applying: $D_i = L$ for $i \in L^+$, $v_{i_1} = 1$ for $i \notin L^+$, and $\sum_i D_i = \Gamma$, and solving the above equations for $D_i$ as in Theorem 3. If $D_i \leq L \forall i$, we have $D^* = D$. Otherwise, we must recursively reduce those $D_i$ such that $D_i > L$ to $L$, and recompute $D$ until all $D_i \leq L$.

4. If $L^- \neq \phi$, and $L^+ = \phi$, we set $D_i = 0$ for $i \in L^-$. That is, $D_i$ can be calculated by first applying: $D_i = 0$ for $i \in L^-$, $v_{i_1} = 1$ for $i \notin L^-$, and $\sum_i D_i = \Gamma$, and solving the above equations for $D_i$ as in Theorem 3. If $D_i \geq 0 \forall i$, we have $D^* = D$. Otherwise, we must recursively increase those $D_i$ such that $D_i < 0$ to 0, and recompute $D_i$ until all $D_i \geq 0$.

**Proof:** When $L^+ = \phi$, and $L^- = \phi$, obviously $D^* = \bar{D}$ is the optimum solution, as shown in Lemma 2.

When $L^- = \phi$, and $L^+ \neq \phi$, the optimal solution has the special feature that $D^*_i = L$, for $i \in L^+$. We show this in the following.

Let’s say the unconstrained optimum ($\bar{D}$) yields $m$ elements in $L^+$ ($m$ of the $\bar{D}_i$ are greater than 1), and let’s consider the optimal solution when these $\bar{D}_i$ are constrained to $L$ (while the other entries are left unconstrained for now). We need to prove that the eigenvector components associated with these $m$ elements are smaller than the eigenvector components associated with all the other elements (e.g. $\hat{v}_{\text{max}, \in L^+} < \hat{v}_{\text{max}, \notin L^+}$), and hence prove that this optimum is in fact the global one according to Lemma 1. We show this using induction.

**Basis:** Let’s use $Z = \bar{D} + \hat{G}$ for the matrix that produces the unconstrained global optimal, call its dominant eigenvalue $\bar{\lambda}$, and assume (wlog) that the first $m$ entries of $D$ are too big (e.g., larger then $L$). Let’s us first consider decreasing the gain in entry 1 to unity while optimizing the other gains (right now, we don’t apply constraints to these gains). Here we denote $D_i$ as gain in entry $i$. Then the matrix of interest changes from $Z$
to 

\[
\begin{bmatrix}
-\alpha \\
\Delta D_2 \\
\vdots \\
\Delta D_n
\end{bmatrix}
\]

, where \( \alpha > 0 \) and \( \sum_{i=2}^{n} \Delta D_i = \alpha \). We would like to prove two statements: first, the eigenvector component corresponding to entry 1 is less than the components associated with all the other entries (e.g. \( v_{\text{max},1} < v_{\text{max},i \neq 1} \)); and second, the gain corresponding to entries from 2 to \( m \) are still at least \( L \).

Recalling Theorem 1, we know that all \( v_{\text{max},i \neq 1} \) are the same, say \( c_1 \). For convenience, let us call \( v_{\text{max},1} \) as \( v_1 \). Hence, the eigenvector associated with the dominant eigenvalue (denoted by \( \hat{\lambda} \)) for the new matrix is \( \hat{v} = [v_1 \ c_1 \ldots \ c_1] \). The Courant-Fischer theorem [29] tells us that \( \bar{\lambda} = \max_v v^T Z v \) and \( \hat{\lambda} = \max_v v^T (Z + \Delta) v \). \( \hat{\lambda} \) can be written as \( \hat{\lambda} = \max_{c_1} (\hat{v}^T Z \hat{v} - v_1^2 \alpha + c_1^2 \alpha) \). When we decrease the gain in entry 1, we know that maximum eigenvalue of the new matrix, say \( \hat{\lambda} \), is larger than \( \bar{\lambda} \). And we also know \( \hat{v}^T Z \hat{v} < \bar{\lambda} \), since \( \hat{v} \) is not the eigenvector for \( \bar{\lambda} \). Thus, from the expression for \( \hat{\lambda} \), the first entry in the eigenvector for the new matrix is smaller than the rest of the entries (i.e. \( v_1 < c_1 \)). Thus the first statement is proved.

For the second statement, we prove it as follows. Since \( v_1 < c_1 \), the eigenvector associated with the optimal gain changes from \( 1_n \) to \( 1_n + \begin{bmatrix} v \\ c_1 \end{bmatrix} \), where we know \( v < 0 < c \) (since only one gain has been moved so far). Further, we know that maximum eigenvalue of the new matrix, say \( \hat{\lambda} \), is larger than \( \lambda \). Plugging into the eigenvector equation and doing some algebra, we finally get the following: \( (\hat{\lambda} - \lambda) 1_{n-1} + c(\hat{\lambda} 1_{n-1} - Z_{2;n,2:n} 1_{n-1} - Z_{2;n,1} v) = (1 + c) \begin{bmatrix} \Delta D_2 \\ \vdots \\ \Delta D_n \end{bmatrix} \). However, since \( \hat{\lambda} > \lambda \), \( v < 0 < c \) and \( Z \) is a non-negative matrix, we recover that all entries in the expression on the left are positive, and hence the change in gains \( \begin{bmatrix} \Delta D_2 \\ \vdots \\ \Delta D_n \end{bmatrix} \) must be all positive and so each other gain strictly increases. Thus, we see that the gain corresponding to entries from 2 to \( m \) are still at least \( L \).

**Induction:** Suppose that we bring any \( l \) of the \( D_i > L \) to \( L \) (\( l < m \)). Let us assume
first that the eigenvector components corresponding to the \( l \) entries are less than the components associated with all the other entries (e.g. \( v_{\text{max},1..l} < v_{\text{max},i,t} \)); and second, that the gain corresponding to entries other than these \( l \) ones are still at least \( L \). Let us show that after we bring another (the \( l+1 \)st) \( D_i \) to \( L \), we still have the appropriate eigenvector component majorization and condition on the gains.

Let us consider bringing \( l+1 \) of the \( \tilde{D}_i \) from the unconstrained optimum s.t. \( i \in \mathcal{L}^+ \) to \( L \). We can do this using two steps. The first step is to move all but one of the offending gains to \( L \) (\( l \) gains), and the second step is to bring the last gain to \( L \). First note that this is possible, since after the first step, all other gains remain greater than \( L \) by assumption. Without loss of generality, for notational convenience, let us assume that we first bring the first \( l \)\( \tilde{D}_i \) s.t. \( i \in \mathcal{L}^+ \) to \( L \), and then move \( \tilde{D}_{l+1} \). Denote the dominant eigenvalue after the first step as \( \lambda_1 \) and the one that after bringing the last gain to \( L \) as \( \lambda_2 \). Again applying the Courant-Fischer theorem, \( \lambda_1 = \max_v v^T (D + \hat{G})v \) and \( \lambda_2 = \max_v v^T (D + \hat{G} + \Delta)v \), where diagonal matrix \( \Delta \) corresponding to changing \( \tilde{D}_{l+1} \) to \( L \) have entries as the following: \( \Delta_{l+1,l+1} = -b \ (b > 0), \Delta_{i,i} = 0 \) for \( i \in [1..l] \) and \( \Delta_i = \Delta D_i \) for \( i \in [l+2..n] \). We also have \( \sum_{i=l+2}^{n} \Delta D_i = b \). With a similar argument to that given in the basis argument, we can show that the \((l+1)\)st eigenvector component is less than the (identical) components after position \( l+1 \). Repeating this argument with each \( \tilde{D}_i \) out of the \( l+1 \) possibilities set to \( L \) last, we can reach the conclusion that the eigenvector components corresponding to all the \( l+1 \) entries are less than the rest common entries.

The proof that the remaining gains \( D_i \) increase (and hence that they remain larger than \( L \) if they were originally larger than \( L \) without the constraints) after bringing these \( l+1 \) gains \( D_i \) to \( L \) is based on the knowledge that the eigenvector components corresponding to all the \( l+1 \) entries are less than the remaining (identical) entries. This can be proved formally in a very similar fashion to the case where a single gain is moved, which we have addressed in the basis step of the induction. Thus the details are omitted.

In case some other gains exceed \( L \) in the process, these can be reduced in the same fashion.

For the case that \( \mathcal{L}^- \neq \phi \), and \( \mathcal{L}^+ = \phi \), the proof is analogous to the case \( \mathcal{L}^- = \phi \), and \( \mathcal{L}^+ \neq \phi \) that we have proved here, and hence it is omitted. \( \square \)
Theorem 4 provides an easy way to calculate the diagonal matrix $D$ that minimizes the dominant eigenvalue of $D + G$ for a diagonally-symmetrizable and non-negative $G$. We can first calculate the optimum $\hat{D}$ without the individual constraints on $D_i$, i.e., if every $\hat{D}_i$ satisfies its constraint, we have found the optimum. Otherwise, if $\hat{D}_i$ that violate their constraints are either all larger than $L$ or all less than 0, the actual optimal $D_i$ for positions where constraints are violated are equal to boundary values. This allows us to quickly locate the $D_i$’s at the boundary rather than to try all combinations of $k$ at boundary to find the optimal solution. In fact, at most $n$ cases (rather than $3^n$) need to be considered. If $\hat{D}$ has entries less than 0 and greater than $L$ at the same time, we must fall back on Theorem 3, i.e. search through the possible combinations of $D_i$ at the boundaries.

In the illustrating example, when $\Gamma = 1.5$, the $D_i$ obtained following step 2) in Theorem 4 are feasible. Hence this single step finds the optimum $D$. When $\Gamma = 2.9$, $D_3$ obtained following step 2) is larger than 1, while $D_1$ and $D_2$ are smaller than 1. This satisfies the condition of step 3). Hence by fixing $D_3$ at 1, and following the calculation in Theorem 3, we find the optimum $D^*$. 

In this section, we have discussed designing a matrix $D$ to minimize the dominant eigenvalue of $D + G$ subject to constraints. This design specifically allows us to allocate limited repair resources to an automaton network to best fight against the spread of a virus. From this viewpoint, it is instructive to study the structure of the optimizing $D = D^*$, and hence the structure of $D^* + G$.

The structure of the optimizing $D$ is highly dependent on the structure of the matrix $G$, which describes the connection topology of the network. The theorems give us the insight that, for a symmetric $G$, the matrix $D$ should be chosen to best equalize the row sums of $D + G$, within the permitted constraints. In terms of resource allocation, this means that placing the most resource at the nodes that have strong connections best prevents virus spread. This makes sense since these nodes have the strongest potential to spread the virus throughout the network if they are infected, and similarly to heal the network when they are healthy. Eliminating viruses at these nodes as soon as possible...
can quickly quench the spread. In case the individual constraints prevent placing enough resource at a node, nearby nodes are provided with extra resources to prevent spread.

It is worth noting the this design is suitable for repair resource allocation before the break-out of a virus or real-time during a virus. In other words, this design is robust to the initial location of the virus. This is useful even when real-time allocation of resources after the start of an epidemic is not possible, or when it is hard to locate and respond to infected nodes network-wide in an epidemic. When the initially affected nodes are known, the design can be improved further using this additional information. We leave this improvement to later work.

![Figure 1: Illustration of a three-node network topology.](image)

**Example** We illustrate calculation of $D = D^*$ to minimize the dominant eigenvalue of $D + G$ for the nonnegative and symmetric topology matrix $G = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{4} & 0 & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{16} & 0 \end{bmatrix}$, subject to the constraints $D_i \in [0, 1]$ and $\sum_i D_i \geq \Gamma$ (see Fig. 1). We consider the following two cases.

1. $\Gamma = 1.5$. In this case, we have $D_1^* = 0.4167$, $D_2^* = 0.4792$, $D_3^* = 0.6042$, and $\lambda^* = 0.7917$ and $v_{max} = [1 \ 1 \ 1]^T$.

2. $\Gamma = 2.9$. In this case, we find that $D_1^* = 0.928$, $D_2^* = 0.972$, $D_3^* = 1$, which yields $\lambda^* = 1.2661$, and $v_{max} = [1 \ 1 \ 0.7047]^T$. We notice that the sum of each row of $D^* + G$ is not identical in this case, and that $D_1$ is increased by a larger fraction.
2.2 Designing $\lambda_{\text{max}}(KG)$ and $\lambda_{\text{max}}(D + KG)$

In this section, we address the other two design problems needed for virus-spreading control, design of diagonal $K$ to minimize the dominant eigenvalue of $KG$ and $D + KG$ ($D$ diagonal) respectively.

The results are similar to those for the $D + G$ case, so the proofs are omitted. Here, we denote the optimum gain $K$ as $K^*$, the optimum dominant eigenvalue of $KG$ ($D + KG$) as $\lambda_{\text{max}}^*$, and the corresponding left and right eigenvector of $KG$ ($D + KG$) as $w_{\text{max}}^*$ and $v_{\text{max}}^*$. Because the results for $KG$ and $D + KG$ problems are so similar, we present them together. We begin with a general structural necessary condition on the optimum:

**Theorem 5** Consider the matrix $KG$ (or $D + KG$), where $K$ and $D$ are diagonal and $G$ is a $n \times n$ matrix. Consider the matrix $K = K^*$ that minimizes the dominant eigenvalue of $KG$ (or $D + KG$) subject to the constraints 1) $\sum K_i \geq \Gamma$ and 2) $K_i \in [0, L]$, assuming the dominant eigenvalue of $KG$ is real and non-repeated. The optimizing $K^*$ and the corresponding left and right eigenvectors $v_{\text{max}}^*$ and $w_{\text{max}}^*$ satisfy one of the following conditions:

Either: $\sum K^*_i = \Gamma$ and for each $i$ we either have $0 < K^*_i < L$ and $\lambda_{\text{max}}^* K^*_i w_{\text{max}}^* v_{\text{max}}^* = 1$, or we have $K^*_i = L$ or $K^*_i = 0$.

Or: $\sum K^*_i < \Gamma$, and for each $i$ we either have $0 < K^*_i < L$ and $w_{\text{max}}^* v_{\text{max}}^* = 0$, or we have $K^*_i = L$ or $K^*_i = 0$.

The optimum $K$ can be found through a search algorithm, as stated in Theorem 6. Theorem 6A is for the $KG$ case, and Theorem 6B is for the $D + KG$ case.

**Theorem 6A** Consider a topology matrix $G$ that is non-negative, irreducible, and diagonally symmetrizable, the $K = K^*$ that minimizes the dominant eigenvalue of $KG$ subject to the constraints 1) $\sum K_i \geq \Gamma$ and 2) $K_i \in [0, L]$ can be found using the following
**algorithm:**

1. Find diagonal $Q$ such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as $\hat{G}$.

2. Set some $K_i$ to 0 and some $K_i$ to $L$, remove the rows and columns of $\hat{G}$ corresponding to the $K_i$ chosen as 0, and rearrange the remaining rows and columns of $K\hat{G}$ in such a way that those $K_i$ fixed at $L$ are at the lower right corner. The resulting matrix can then be written as

$$\begin{bmatrix} K_A & 0 \\ 0 & LI \end{bmatrix} \begin{bmatrix} \hat{G}_{11} & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} \end{bmatrix}.$$

3. The eigenvalue $\lambda$ satisfies

$$\lambda 1^T (\hat{G}_{11} - \hat{G}_{12} (L\hat{G}_{22} - \lambda I)^{-1}L\hat{G}_{21})^{-1} 1 = \Gamma - \text{trace}(LI),$$

and can be found through a simple numerical procedure. $K_A$ can be found as

$$K_A = \text{diagonalize}(\lambda 1^T (\hat{G}_{11} - \hat{G}_{12} (L\hat{G}_{22} - \lambda I)^{-1}L\hat{G}_{21})^{-1}).$$

Check whether $K_i$ is feasible.

4. Repeat with different $K_i$ set to their boundary values until the global minimum $\lambda$ is found.

**Theorem 6B** Consider a topology matrix $G$ that is non-negative, irreducible, and diagonally symmetrizable, the $K = K^*$ that minimizes the dominant eigenvalue of $D + KG$ subject to the constraints 1) $\sum K_i \geq \Gamma$ and 2) $K_i \in [0, L]$ can be found using the following algorithm:

1. Find diagonal $Q$ such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as $\hat{G}$.

2. Set some $K_i$ to 0 and some $K_i$ to $L$, and rearrange the rows and columns of $D + K\hat{G}$ in such a way that those $K_i$ fixed at 0 or $L$ are at the lower right corner. The resulting matrix can then be written as

$$\begin{bmatrix} D_A & 0 \\ 0 & D_B \end{bmatrix} + \begin{bmatrix} K_A & 0 \\ 0 & K_B \end{bmatrix} \begin{bmatrix} \hat{G}_{11} & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} \end{bmatrix}.$$

3. The eigenvalue $\lambda$ satisfies

$$(\lambda 1^T - D_A 1^T) (\hat{G}_{11} - \hat{G}_{12} (D_B + K_B \hat{G}_{22} - \lambda I)^{-1}K_B \hat{G}_{21})^{-1} 1 = \Gamma - \text{trace}(K_B I),$$

and can be found through a simple numerical procedure. $K_A$ can be found as

$$K_A = \text{diagonalize}(\lambda 1^T - D_A 1^T) (\hat{G}_{11} - \hat{G}_{12} (D_B + K_B \hat{G}_{22} - \lambda I)^{-1}K_B \hat{G}_{21})^{-1}.$$

Check whether $K_i$ is feasible.

4. Repeat with different $K_i$ set to their boundary values until the global minimum $\lambda$ is found.
When $G$ is positive definite in addition to diagonally symmetrizable, we can show that $\lambda_{\text{max}}(KG)$ is a convex function. In this case, the search algorithms given in Theorem 6 can be simplified: We can check whether a solution is a local optimum and stop or continue the search accordingly, since a local optimum is guaranteed to be a global optimum. Theorem 7A is for the $KG$ case, and Theorem 7B is for the $D + KG$ case.

**Theorem 7A**  Consider a topology matrix $G$ that is non-negative, irreducible, positive definite, and diagonally symmetrizable. The $K = K^*$ that minimizes the dominant eigenvalue of $KG$ subject to the constraints 1) $\sum K_i \geq \Gamma$ and 2) $K_i \in [0, L]$ can be found using the following algorithm:

1. Find diagonal $Q$ such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as $\hat{G}$.

2. Find $K = \text{diagonalize}(\hat{G}^{-1}1\Gamma/(1^T\hat{G}^{-1}1))$. If $0 \leq K_i \leq L$, this solution is optimal. Otherwise, go to step 3.

3. Set some $K_i$ to 0 and some $K_i$ to $L$, remove the rows and columns of $G$ corresponding to the $K_i$ chosen as 0, and rearrange the remaining rows and columns of $KG$ in such a way that those $K_i$ fixed at $L$ are at the lower right corner. The resulting matrix can then be written as

$$
\begin{bmatrix}
K_A & 0 \\
0 & LI
\end{bmatrix}
\begin{bmatrix}
\hat{G}_{11} & \hat{G}_{12} \\
\hat{G}_{21} & \hat{G}_{22}
\end{bmatrix}.
$$

4. The eigenvalue $\lambda$ satisfies

$$
\lambda 1^T(\hat{G}_{11} - \hat{G}_{12}(L\hat{G}_{22} - \lambda I)^{-1}L\hat{G}_{21})^{-1}1 = \Gamma - \text{trace}(LI),
$$

and can be found through a simple numerical procedure.

5. $K_A$ can be found as $K_A = \text{diagonalize}(\lambda 1^T(\hat{G}_{11} - \hat{G}_{12}(L\hat{G}_{22} - \lambda I)^{-1}L\hat{G}_{21})^{-1})$. If $0 \leq K_i \leq L$, and the left eigenvectors associated with the dominant eigenvalue of $KG$ has the pattern that the entries corresponding to $K_i = L$ are less then those corresponding to $0 < K_i < L$, and the latter are less than those corresponding to $K_i = 0$, this solution is optimal. Otherwise, go back to step 3.

**Theorem 7B**  Consider a topology matrix $G$ that is non-negative, irreducible, positive definite, and diagonally symmetrizable. The $K = K^*$ that minimizes the dominant
eigenvalue of $D + KG$ subject to the constraints 1) $\sum K_i \geq \Gamma$ and 2) $K_i \in [0, L]$ can be found using the following algorithm:

1. Find diagonal $Q$ such that $Q^{-1}GQ$ is symmetric. We denote $Q^{-1}GQ$ as $\hat{G}$.

2. Find $\lambda$ that satisfies $\hat{G}^{-1}(\lambda 1^T - 1^TD)1 = \Gamma$ through a simple iteration. And then find $K$ from $K = \text{diagonalize}(\hat{G}^{-1}(\lambda 1^T - 1^TD))$. If $0 \leq K_i \leq L$, this solution is optimal. Otherwise, go to step 3.

3. Set some $K_i$ to 0 and some $K_i$ to $L$, and rearrange the rows and columns of $D + KG$ in such a way that those $K_i$ fixed at 0 or $L$ are at the lower right corner. The resulting matrix can then be written as $\begin{bmatrix} D_A & 0 \\ 0 & D_B \end{bmatrix} + \begin{bmatrix} K_A & 0 \\ 0 & K_B \end{bmatrix} \begin{bmatrix} \hat{G}_{11} & \hat{G}_{12} \\ \hat{G}_{21} & \hat{G}_{22} \end{bmatrix}$.

4. The eigenvalue $\lambda$ satisfies $(\lambda 1^T - D_A 1^T)(\hat{G}_{11} - \hat{G}_{12}(D_B + K_B \hat{G}_{22} - \lambda I)^{-1}K_B \hat{G}_{21})^{-1}1 = \Gamma - \text{trace}(K_B)$, and can be found through a simple numerical procedure.

5. $K_A$ can be found as $K_A = \text{diagonalize}((\lambda 1^T - D_A 1^T)(\hat{G}_{11} - \hat{G}_{12}(D_B + K_B \hat{G}_{22} - \lambda I)^{-1}K_B \hat{G}_{21})^{-1})$. If $0 \leq K_i \leq L$, and the left eigenvectors associated with the dominant eigenvalue of $D + KG$ has the pattern that the entries corresponding to $K_i = L$ are less than those corresponding to $0 < K_i < L$, and the latter are less than those corresponding to $K_i = 0$, this solution is optimal. Otherwise, go back to Step 3.

## 3 Control of the Hong Kong SARS Epidemics

Recall that the article [3] developed a spatial model for the spread of SARS in Hong Kong’s 18 districts (See Fig. 2), and proposed a homogeneous control for reducing the basic reproduction ratio $R_0$ to 1. Here, we find the optimal heterogeneous control that uses the same total resource amount as the controller in [3]. This controller reduces the basic reproduction ratio to 0.64. Thus, we see that an epidemic can be stopped more quickly with the same control resources, by allocating more resources to some districts
Figure 2: The Map of Hong Kong’s 18 districts (obtained from http://en.wikipedia.org/wiki/Hong_Kong#Administrative_divisions). We use the model parameters in [3]: $h_{ii} = 1$, $h_{i,j} = 0.57$ when District $i, j$ are neighbors, $h_{ij} = 0.02$ when $i, j$ are not adjacent, $\beta = 0.062$, and $T = 10.6$.

than others. Equivalently, it is easily shown that $R_0 = 1$ can be achieved even when the total control resource is reduced to 79% of the one with equal allocation (See Table 1).

This intelligent allocation takes advantage of the spatial structure of the population, by placing more control resources in the districts that are important for the spread of an epidemic (see Fig. 2). In this way, the limited control resources are best able to reduce the rate at which the epidemic diminishes. Such a control would reduce the impact on people’s daily lives in some districts (which have less control resources allocated) and overall, while still stopping the virus spread quickly.

For illustration, we also consider how the heterogeneous resource allocation changes when the local mixing rate $h_{ii}$ is increased (compared to the mixing rate of neighboring districts $h_{ij}$). As expected, increasing the local mixing rate makes use of spatial information less important, and also makes the allocation more homogeneous (See Table 1).

In all of these experiments, we see that the most resources are placed in Districts 5 and 7, and the smallest resources are placed in District 1. This is expected since Districts 5 and 7 are the ones with pivotal locations (e.g. with many neighbors) and hence the successful control of these districts is important in the control of disease spread. In
contrast, District 1 is almost isolated and hence has the least contribution to the spread of diseases in Hong Kong.

Table 1: Resource Allocation Needed to Reduce $R_0$ to 1.

<table>
<thead>
<tr>
<th>District</th>
<th>$h_{ii} = 1$</th>
<th>$h_{ii} = 1.5$</th>
<th>$h_{ii} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>District 1</td>
<td>0.9096</td>
<td>0.6222</td>
<td>0.3031</td>
</tr>
<tr>
<td>District 2</td>
<td>0.9096</td>
<td>0.5368</td>
<td>0.2525</td>
</tr>
<tr>
<td>District 3</td>
<td>0.5816</td>
<td>0.4338</td>
<td>0.2546</td>
</tr>
<tr>
<td>District 4</td>
<td>0</td>
<td>0</td>
<td>0.1919</td>
</tr>
<tr>
<td>District 5</td>
<td>0</td>
<td>0.0160</td>
<td>0.0915</td>
</tr>
<tr>
<td>District 6</td>
<td>0.5816</td>
<td>0.4242</td>
<td>0.1705</td>
</tr>
<tr>
<td>District 7</td>
<td>0</td>
<td>0</td>
<td>0.0834</td>
</tr>
<tr>
<td>District 8</td>
<td>0.9096</td>
<td>0.5926</td>
<td>0.2708</td>
</tr>
<tr>
<td>District 9</td>
<td>0</td>
<td>0.0789</td>
<td>0.1734</td>
</tr>
<tr>
<td>District 10</td>
<td>0</td>
<td>0</td>
<td>0.1419</td>
</tr>
<tr>
<td>District 11</td>
<td>0.5816</td>
<td>0.4553</td>
<td>0.2194</td>
</tr>
<tr>
<td>District 12</td>
<td>0</td>
<td>0.2118</td>
<td>0.1798</td>
</tr>
<tr>
<td>District 13</td>
<td>0.5816</td>
<td>0.4457</td>
<td>0.1985</td>
</tr>
<tr>
<td>District 14</td>
<td>0.9096</td>
<td>0.5428</td>
<td>0.2587</td>
</tr>
<tr>
<td>District 15</td>
<td>0.9096</td>
<td>0.4784</td>
<td>0.2475</td>
</tr>
<tr>
<td>District 16</td>
<td>0.9096</td>
<td>0.4784</td>
<td>0.2475</td>
</tr>
<tr>
<td>District 17</td>
<td>0</td>
<td>0.1919</td>
<td>0.1908</td>
</tr>
<tr>
<td>District 18</td>
<td>0</td>
<td>0.1919</td>
<td>0.1908</td>
</tr>
<tr>
<td>$\sum K_i$ using heterogeneous control</td>
<td>7.7842</td>
<td>5.7007</td>
<td>3.6667</td>
</tr>
<tr>
<td>$\sum K_i$ using homogeneous control</td>
<td>5.0138</td>
<td>4.4009</td>
<td>3.22</td>
</tr>
</tbody>
</table>

The $K_i$ for each district are shown in the table. Notice that smaller $K_i$ corresponds to more resources. $\sum K_i$ equals the subtraction of total utilized resources from $N$ (the number of districts, e.g., 18 in this case). Note that the reduction in resource used for $h_{ii} = 1$ is $\frac{18 - 7.7842}{18 - 5.0138} \times 100\% = 79\%$. 
References


