Optimizing Epidemic Protection for Socially Essential Workers

C. Barrett R. Beckman K. Bisset J. Chen S. Eubank V. S. Anil Kumar
{cbarrett,rbeckman,kbisset,jchen,seubank,akumar}@vbi.vt.edu
Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061

B. Lewis M. V. Marathe P. Stretz
{blewis,mmarathe,pstretz}@vbi.vt.edu
Virginia Bioinformatics Institute
Virginia Tech
Blacksburg, VA 24061

T. DuBois A. Srinivasan
{tdubois,srin}@cs.umd.edu
Department of Computer Science
University of Maryland
College Park, MD 20742

ABSTRACT

Public-health policy makers have many tools to mitigate an epidemic's effects. Most related research focuses on the direct effects on those infected (in terms of health, life, or productivity). Interventions including treatment, prophylaxis, quarantine, and social distancing are well studied in this context. These interventions do not address indirect effects due to the loss of critical services and infrastructures when too many of those responsible for their day-to-day operations fall ill. We examine, both analytically and through simulation, the protection of such essential subpopulations by sequestering them, effectively isolating them into groups during an epidemic. We develop a framework for studying the benefits of sequestering and heuristics for when to sequester. We also prove a key property of sequestering placement which helps partition the subpopulations optimally. Thus we provide a first step toward determining how to allocate resources between the direct protection of a population, and protection of those responsible for critical services.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Health

Keywords

public health informatics, optimization, epidemiology

1. INTRODUCTION

Society's resilience to a severe infectious disease epidemic depends on the continued operation of essential services, e.g. water, sewer, and power systems, health care, food distribution, law enforcement, leadership, etc. Federal response plans [15] offer guidance to businesses and local agencies for continuity-of-operations planning and identifying critical infrastructures essential to the country's ability to provide necessary services. People who perform these services are essential because they are hard to replace quickly and in their absence social organization rapidly deteriorates. Reducing the overall prevalence of infections also affords protection to critical subpopulations; however, it may not be sufficient. Sequestering is a social distancing technique that further protects an essential subpopulation by placing its healthy members into small groups isolated from the general population and from each other. Protective sequestration was used effectively during the 1918 pandemic by a few socially isolated populations [13] and was used to varying effect for centuries [7], but fell out of use as societies became more tightly integrated. Nevertheless, protective sequestration continues to be an important tool to consider for ensuring a ready national defense force [8].

The logic of protective sequestering may explain its relative absence in epidemiological and public health policy literature. Sequestering a small subpopulation is costly and may have little direct effect on the progress or ultimate size of an outbreak. Instead, sequestering's benefits come from the external and indirect effects on availability of services throughout the epidemic. These effects are difficult to quantify inside the purview of contagion control and lie outside the scope of a public health system's responsibilities to contain an epidemic. Thus, assigning resources to protect a small subpopulation is difficult to rationalize using the prevailing methodology. Rather, sequestering presents a complex problem in which socio-technical interdependencies must be considered in integrated social decision-and policy-making.

The strategies needed to protect a small group of *socially essential* individuals are different from those intended to reduce morbidity and mortality in the *general public*. Consider quarantine and isolation, which are standard tools for controlling outbreaks which protect the uninfected by isolating the infected. Sequestering is the converse: protection of an *uninfected subpopulation* from the overall population during an outbreak. In this context, small groups could be housed in military barracks, a requisitioned public

IHI'12, January 28–30, 2012, Miami, Florida, USA.

school, industrial facilities, and other socially isolatable locations. The size of a group would include the number of co-located sequestered individuals and associated logistical-support personnel. Minimizing logistical connectivity to the greater population is a practical challenge, but is outside our scope. We make the simplifying assumption that such isolation is total.

The need for protective sequestration is perhaps best highlighted by considering a military context [9, 12]. In the course of its operations, the military creates circumstances conducive to the spread of infectious disease, as occurred, for example, in the 1918 influenza pandemic. Naval ships and military bases contain personnel in confined settings, wherein infectious diseases can be easily transmitted. Personnel can be called up and assigned in a way that is carefully designed to minimize their total infections. This context also provides authorities the level of control necessary to effectively sequester a population. We study how early estimates of individual infection-probabilities can help sequester individuals better in such confined settings. The results reported here arose from investigations into sequestration feasibility and optimization for DoD. Some of our findings are also described in [2, 1].

We use the SIR model (Susceptible, Infectious, Recovered) of infection [11] over a contact network to develop a formulation of the sequestration problem. Based upon environment and contact network factors before sequestering, each individual has a probability of being latently infected. Once individuals are placed into groups, those latently infected will become infectious and may spread the epidemic within the group. The transmission probability over time p for any pair depends upon the amount of time until the infectious person recovers or is removed as well as the environmental characteristics of the group and epidemic-specific factors.

Using this model we discover and prove an important property: that an optimal placement of people into groups places people with others of similar estimated probabilities of infection (EPIs). We call this the "well-ordered" property and use it as the basis for an efficient algorithm for optimal placement. We explore the computational complexity of our sequestering algorithm along with some of the associated trade-offs between running time and memory usage. In addition to providing optimal partitions of the critical population, a minor change allows our algorithm to show when better overall results are achieved by leaving some of the critical population unsequestered.

We then perform a series of simulations to study the benefits of sequestering as well as the significant trade-offs involved. The main parameters in a sequestering setup are when to start sequestering (which affects the EPIs) and what group sizes to use. The result we are most interested in is the fraction of the critical population which becomes ill. First we evaluate the number of infections using optimal group placement compared with random group placement or not sequestering at all. Then we study the trade-offs between EPIs at the beginning of sequestering, group size, and the number of infections. Finally, we find that small EPI estimation errors, or large errors in only a few estimates, changes overall infection rates only slightly.

Our results suggest that to be effective, critical workers should be sequestered very early on during an outbreak. Sequestering early enough can completely protect the critical population from illness. On the other hand, beyond a threshold level of latent infections in the critical population, sequestering them into densely connected groups may in fact produce more total infections than result from leaving them in the general population. The group sizes and the time to enforce sequestration are both important logistically as well as socially – from a logistical standpoint, one would like to keep the group sizes large; from a societal standpoint, one would like to sequester individuals as late as possible. Our work provides a method for quantifying the trade-off between these competing factors.

Throughout our analysis, we make a number of assumptions that we state here. First, we assume that individuals who get infected are removed from the sequestered population in a predictable way. Removal is possible only after the individual is symptomatic; in case of flu-like diseases, this usually happens a couple of days after the individual is infectious. We incorporate this time into the transmission probability p. Second, we assume that EPIs can be effectively estimated. EPIs capture the uncertainty inherent in this complex process; the sensitivity studies reported in the experimental results section show that our results are fairly robust. Third, we assume completely mixed groups with uniform transmission probabilities. Again while not completely true it appears to be a reasonable assumption for flu-like illness.

2. DEFINITIONS AND MODELS

The *Optimal Sequestering Problem* (`OSP`) is defined as follows: Given a set $M' \subseteq M$ of socially essential individuals, an upper bound g on group size and an upper bound r on the number of groups, decide:

1. *Triggered Sequestering* (`TS`): the optimal time t_{opt} when the set M' of individuals should be removed from the population, isolating the subset $M' - M''$ of M' that are infectious and symptomatic, and
2. *Optimal Allocation* (`OA`): allocate the asymptomatic subset M'' of M' into groups of size at most g with the total number of groups being no more than r, so that the expected number of individuals M''_{exp} that eventually become infected is minimized.

Thus we have two subproblems: (i) finding the best time when the essential subpopulation should be isolated from the general population (`TS`), and (ii) finding the best strategy for partitioning the individuals who are not yet symptomatic into groups of size at most m(`OA`). The earlier the people are sequestered, the higher the chances of a false alarm and the cost of supporting them until the outbreak ends. The later they are sequestered, the more likely that asymptomatic infectious people will be included. This trade-off, which we begin to explore in Section 4.3, drives the `TS` subproblem.

For any sequestered subpopulation the best partitioning into groups for the `OA` problem separates all of the latently infected individuals from those who are healthy. While it is not feasible to know for certain who is latently infected, we are able to make reasonable estimates of infection probabilities. For example, those living near high areas of infectivity, those with infected friends and family, and even those living with school-age children have higher EPIs. The difference in EPIs provides us with an important, exploitable property when trying to minimize the epidemic's affects. This subproblem has certain degenerate cases, such as when individuals are diagnosed as soon as they become infectious (and

can then be removed from a healthy group before spreading the disease) or group sizes are small enough to effectively isolate individuals. We focus our study on the wide range of more interesting cases.

To put the problem in mathematical notation, we are given: a set V of n people (or nodes), and a set of groups with capacities $m_1, m_2, \ldots, m_k$. Groups are tightly constrained locations which are typically small enough that it is natural to assume that the contact graph within a group is complete [14]. Larger groups would contain their own contact networks, and assuming a complete graph within each group provides an upper bound on the total number of infections.

In such isolation, there are only two ways in which a sequestered individual may be infected: they can carry the infection from before sequestering began, or they can contract the disease from another in the same group. We quantify the chance that any person v is latently infection as the external probability of infection or EPI, denoted s_v. The symbol "s" in s_v denotes "susceptibility" and we assume that these initial exposures happen independently for all v. The EPIs can be estimated by combining computer simulations, demographic characteristics and ground measurements [10]. For any set of nodes V_i confined to a single group, a subset U_i of V_i becomes initially infected according to their EPIs.

For within-group transmission, we use a standard SIR model over a contact network. For every pair of people in contact where one is infected and the other susceptible, there is a transmission probability per minute of contact. Once someone becomes infected, they progress to infectious, and then recovered (or removed if symptoms are detected). In a uniform population we further simplify the rate of transmission to a single value p: the probability of the epidemic spreading from a person v to another u given that v becomes infected and u does not become infected some other way first.

Using this model, disease transmission within each group is equivalent to percolation in the Erdős-Rényi random graph $G(V_i, p)$, and the set of nodes from V_i that finally become infected are those reachable in the random-graph from some node in U_i. Figure 1 shows how the random-graph corresponds to disease transmissions.

Given a partition $V_1, V_2, \ldots, V_k$ of the population V, we can consider the *expected* number of finally infected nodes, where the expectation is taken both over the random choice of the initially infected set given by the EPIs, as well as the random choices made in the percolation (disease-spreading) process. The goal of the `OA` problem is to find a feasible partitioning so that the expected number of infections (also referred to as the *outbreak size*) is minimized. The inputs to the problem are the person-to-person transmission probability p as well as the values s_j, for $j = 1, 2, \ldots, n$ and m_i for $i = 1, 2, \ldots, k$; the output is the partition.

If the EPIs vary quite a bit, the partitioning can significantly affect the expected outbreak size. In particular, the natural heuristic of random assignment can perform very poorly. A simple example of this is the following: let $|V| = k^2$ and the group capacities be $m_1 = m_2 = \cdots = m_k = k$. Let $s_i = 1$ for $i = 1, \ldots, k$, and $s_j = 0$ for $j > k$. Assume the disease is very contagious, so that the presence of an infected node in a group will infect everyone else in that group: i.e., p is essentially one. The optimal solution places all of the initial k nodes with $s_i = 1$ into one group, and partition the rest into the remaining $k-1$ groups - this would have a cost of k. However, a random partitioning and assignment of V to the groups will result in $\sim k(1-1/e)$ groups having some node $i \leq k$, which results in an expected outbreak of size $\Theta(k^2)$. The optimal solution in this example groups the nodes according to similar EPI values, and our algorithm is based on this idea; see Figure 2 for an illustration.

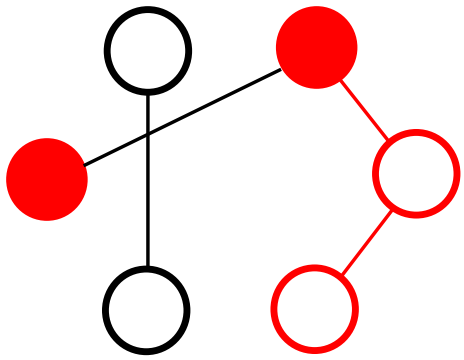

Figure 1: A example random-graph instance of a group with six people. The contact network is a complete graph (all pairs are connected). Through a random process only some people are initially infected and only some edges transmit the disease. Solid red nodes are initially infected, hollow red nodes become infected later, and hollow black nodes are uninfected. Red edges are those randomly included which result in new infections, and black edges are those randomly included which do not transmit the disease (either because they connect two uninfected nodes, or two nodes both infected from other sources).

3. AN EFFICIENT ALGORITHM FOR SEQUESTERING

We now describe our main algorithm for optimal sequestering. As discussed earlier, there are two natural contrasting heuristics for grouping people: (i) load-balancing–type heuristics where we try and keep the total carrier probabilities approximately the same across the groups, which is usually well-achieved by a random partitioning, and (ii) where people with high carrier probabilities are all grouped together where possible: i.e., viewing the objective of number-of-infected-people as something like a concave function. As mentioned earlier, the former heuristic can lead to suboptimal assignment, and our algorithm is based on a refinement of the latter heuristic. We start with the following "well-ordered" property of an optimal solution, and then discuss how this leads to a natural dynamic programming algorithm. We say that a partition is well-ordered if for every pair of groups G_1 and G_2 there exists a separator x such that either $\forall v \in G_1, s_v \leq x$ and $\forall u \in G_2, s_u \geq x$ or $\forall v \in G_1, s_v \geq x$ and $\forall u \in G_2, s_u \leq x$. In any solution which is not well-ordered (i.e., there are people $u, w \in G_1$ and $v \in G_2$ such that $s_u < s_v < s_w$) then we can switch the group assignments of two people without increasing the expected number of infections. We formally capture this property in the following theorem:

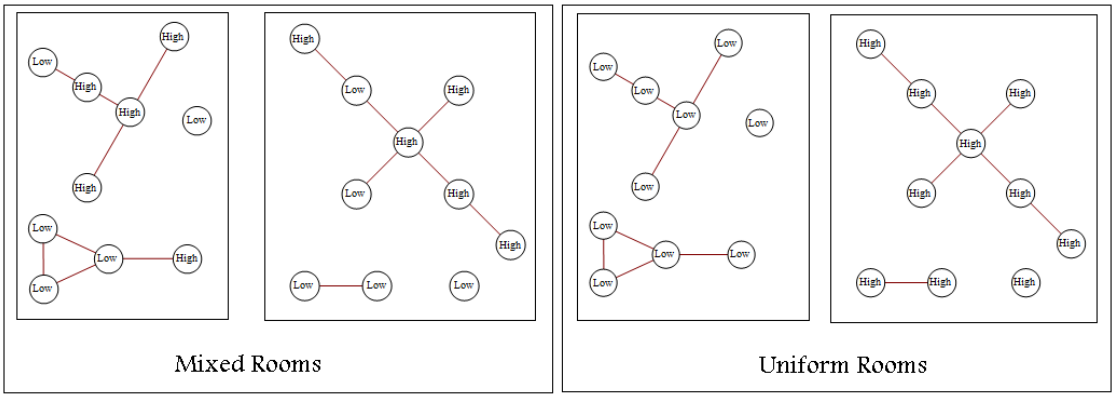

Figure 2: Here we see two partitions of a population into two groups each. The nodes represent individuals and are labeled "high" or "low" based upon the person's EPI. For this example think of "high" being close to 1 and "low" close to 0. The edges represent randomly chosen disease transmission paths. In these instances, an outbreak spreads from any externally infected person to all others in the same connected component. In the left example, the population is divided into groups randomly. All but 4 people are connected to those who are likely infected. In the right example, there is a low EPI group and a high EPI group. Here all 10 people in the low EPI group are likely to remain healthy.

THEOREM 1. *For any two groups and a subset of the people to be assigned to these two groups, there exists an optimal partition where all* EPI*s in one group are less than or equal to all* EPI*s in the other.*

Proof of Theorem 1 We use a probabilistic argument to prove the theorem, so we start by stating some terms and definitions. For any random event A, $Pr[A]$ is the probability that A occurs. For any discrete random variable X with range R, its expected value (essentially a weighted average) is $E[X] = \sum_{x \in R} x \cdot Pr[X = x]$. In our analysis, we make heavy use of the linearity of expectation, which says that for any two random variables X and Y, $E[X + Y] = E[X] + E[Y]$.

For any person a let I_a denote the probabilistic event that a is externally infected (thus $Pr[I_a] = s_a$). For any set of people S, define the random variable X_S to be the number of final infections among the people S. Similarly for any set of people S, define $Y_S = E[X_{S\cup\{a\}}|I_a] - E[X_{S\cup\{a\}}|\overline{I_a}]$ where a is any person not in S. Intuitively, Y_S is the expected marginal number of infections caused when an additional person added to the group S is externally infected versus when the additional person is not infected. Using the linearity of expectation we have that $E[X_{S\cup\{a\}}] = Y_S Pr[I_a] + E[X_{S\cup\{a\}}|\overline{I_a}]$. This means that the net cost from swapping a from a group with S for another person a' from a group with S' is

$$\begin{aligned}&E[X_{S\cup\{a'\}}] + E[X_{S'\cup\{a\}}] - E[X_{S\cup\{a\}}] - E[X_{S'\cup\{a'\}}]\\ =&(Pr[I_{a'}] - Pr[I_a])(Y_S - Y_{S'}).\end{aligned}$$

There are two parts necessary to prove the theorem. The first is that if two people a, b are in a group with additional people S, then $Pr[I_a] \leq Pr[I_b]$ implies $Y_{S\cup\{a\}} \geq Y_{S\cup\{b\}}$. We show this by conditioning on the connected components of the group. Define random variables C to be a connected component decomposition (a set of disjoint subsets whose union is the entire group) of the group $S \cup \{a, b\}$ and $C(x)$ to be the set of nodes in the component containing x.

Assuming $Pr[I_a] \leq Pr[I_b]$ we have:

$$\begin{aligned}Y_{S\cup\{a\}} =&E[X_{S\cup\{a,b\}}|I_b] - E[X_{S\cup\{a,b\}}|\overline{I_b}]\\ =&\sum_C Pr[C] \cdot \left(E[X_{C(b)}|I_b] - E[X_{C(b)}|\overline{I_b}]\right)\\ =&\sum_{C:a\in C(b)} Pr[C] \cdot |C(b)| \cdot \prod_{x\in(C(b)-b)} Pr[\overline{I_x}]\\ &+ \sum_{C:a\notin C(b)} Pr[C] \cdot \left(E[X_{C(b)}|I_b] - E[X_{C(b)}|\overline{I_b}]\right)\\ \geq&\sum_{C:b\in C(a)} Pr[C] \cdot |C(a)| \cdot \frac{Pr[\overline{I_b}]}{Pr[\overline{I_a}]} \cdot \prod_{x\in(C(a)-b)} Pr[\overline{I_x}]\\ &+ \sum_{C:b\notin C(a)} Pr[C] \cdot \left(E[X_{C(a)}|I_a] - E[X_{C(a)}|\overline{I_a}]\right)\\ =&\sum_C Pr[C] \cdot \left(E[X_{C(a)}|I_a] - E[X_{C(a)}|\overline{I_a}]\right)\\ =&Y_{S\cup\{b\}}.\end{aligned}$$

The second part is that in an optimal solution, if a and b are in different groups along with S_a and S_b respectively, then $Pr[I_a] < Pr[I_b]$ implies $Y_{S_a} \geq Y_{S_b}$. This can be seen easily because the cost of swapping a and b is $(Pr[I_a] - Pr[I_b])(Y_{S_b} - Y_{S_a})$. If the original partition was optimal, the change in cost incurred by the swap must be at least 0. Since by assumption $Pr[I_a] - Pr[I_b] < 0$, $Y_{S_b} - Y_{S_a}$ cannot be positive.

Taken together these give that for any a, b, c with $Pr[I_a] < Pr[I_b] < Pr[I_c]$ with a and c in the same group with S_{ac} others and b in a different group with S_b in an optimal solution, $Y_{S_{ac}\cup\{a\}} \geq Y_{S_{ac}\cup\{c\}}$. This follows because a and c are in the same group and $Y_{S_{ac}\cup\{a\}} \leq Y_{S_b} \leq Y_{S_{ac}\cup\{c\}}$ because the group assignment is optimal. Since $Y_{S_{ac}\cup\{a\}} \leq Y_{S_b} \leq Y_{S_{ac}\cup\{c\}} \leq Y_{S_{ac}\cup\{a\}}$, the three quantities must be equal and any two of the three people can be swapped while keeping the cost optimal. Given any optimal solution, as long as there exists such a, b, c we can continue making these swaps,

moving those with higher EPIs to the right, until an optimal solution of the desired form is reached. □

When there are more than two groups, applying Theorem 1 to all pairs of groups yields the following corollary:

COROLLARY 2. *Given any set of groups and a set of people with known* EPIs*, there exists an optimal partition which orders the groups, and if group i comes before group j than all of the* EPIs *in i are less than or equal to all of the* EPIs *in group j.*

Theorem 1 tells us that an optimal solution keeps people with like EPIs together, however it does not say when to stop filling one group and start on the next - this can be determined by dynamic programming. We develop a dynamic programming algorithm whose time and space complexities are exponential in the number of group-sizes r but polynomial in both the number of people and the total number of groups. Our algorithm, which heavily exploits the symmetry of infectivities is detailed in Figure 3 for the primary case in which all group capacities are uniform.

Algorithm SEQUESTER requires the function $g(S)$, which is the number of infections if the set S forms a group. In Figure 4, we describe a dynamic programming algorithm to compute $g(S)$. The proof of SEQUESTER's optimality follows from straightforward induction on the dynamic programming arrays which show that it produces an optimal well-ordered partition and Theorem 1 and Corollary 2, which show that an optimal well-ordered partition is an optimal partition.

Also, in Figure 3, Algorithm SEQUESTER is described for the setting in which all group sizes are uniform. This algorithm can be extended to handle r group types (where a type i contains u_i groups of capacity m_i and with a pairwise transmission probability p_i) by making OPT into an $r+1$ dimensional array and computing for all $a \in [1..n]$ and $b_i \in [0..u_i]$:

$$OPT(a, b_1, \ldots, b_r) = \min_{j=0}^{r} \min_{i=0}^{m_j} \left(OPT(a-i, b_1, \ldots b_j - 1, \ldots, b_r) + g(\{a-i+1, \ldots, a\})\right).$$

An interesting feature of our algorithm is that it not only determines how to optimally partition a critical subpopulation, but it can also be used to determine when it is better to leave some critical individuals unsequestered. For example, if there is one group available for sequestering, a single individual with a high EPI, and many with low EPI, then leaving the high EPI person out of the group can result in more of the critical population remaining healthy. The person who is a likely carrier is left behind to protect the rest of the critical population. We address this problem of incomplete sequestering by adding an additional group where everyone assigned to that group becomes infected. This new group represents those who are not sequestered. The optimal partition produced by our algorithm for this modified instance, minus the group where everyone becomes infected, is an optimal partial placement for the original problem.

3.1 Sequester's Complexity

To analyze SEQUESTER's efficiency we break it into two parts, one part specific to the subroutine g and concerning mostly initialization, and a second part which concerns the main function Sequester.

In initializing the function g, we compute and store for each group, and every possible number of initially infected people in the group, how many people become finally infected in expectation. If we let m_i denote the size of the i^{th} group when sorted by group sizes, then the permanent space used after initialization is $\sum_i m_i$. When computing these values, we need an m_i^3 sized temporary array for the intermediate computation. The j, x, y entry in this array records the probability that there are x people infected by a path of distance at most j, and y people infected at distance exactly j. Each entry in these arrays takes time $O(m_i^2)$ to compute, for a total run-time complexity of $O(\sum_i m_i^5)$.

Once the initialization is complete, we create an array OPT of size $n \cdot \prod_i u_i$. For each entry in OPT we create and store a matrix A once taking time and space $O(\max_i m_i^2)$ and we make $\sum_i m_i$ calls to g, each of which does $O(m_i)$ arithmetic operations for a total running time of $O(n \cdot (\prod_i u_i \cdot \sum_j m_j^2))$.

In the uniform case where there are u groups all of capacity m and $m \cdot u \approx n$ (meaning the total space available is approximately the number of people), this yields a space complexity of $O(n^2/m + m^2)$ and a time complexity of $O(m^5 + n^2 \cdot m)$. Note that a long as the group capacities scale slowly compared with the population, the m^5 term is not prohibitive. For example, if the maximum group capacity is at most the square root of the population size ($m = 100$ for a population of 10,000), then m^5 is no more than $n^2 \cdot m$.

3.2 Memory Efficient Adaptation

When dealing with large datasets (n of a hundred thousand or more), the space complexity of $O(n^2/m + m^2)$ can be prohibitively expensive. We can greatly improve space efficiency with a factor of 2 increase in running time using the following observation: to compute the array after person i, all we need to have stored are the $\max_t m_t$ rows from $OPT(i - \max_t m_t, *)$ to $OPT(i-1, *)$. If we only want to compute the optimal expected outbreak size, we can reuse the space from $OPT(i \mod m + 1, j)$ for every entry $OPT(i, j)$ for a space savings factor of n/m. If we want to compute the optimal partition however, we need to be able to backtrack through the OPT array, which we cannot do efficiently if we have to repeatedly recompute it from scratch for every m people. Instead we have to do something more clever.

THEOREM 3. *Suppose we are given an instance of the Sequestering problem with a single room type and total capacity $O(n)$. We can compute an optimal partition in time $O(m^5 + n^2 m)$ and space $O(m^3 + n^{1.5}/\sqrt{m})$.*

The time complexity and initialization space complexity are taken from above. Here we prove the improved space bounds for the main phase. For the single group type case, after computing OPT for c rows (for a value of c to be specified later), we store the entries for the last m of them, and reuse the space for OPT in computing the next c people. This takes space $mu_1 \cdot \frac{n}{c}$ for the saved blocks and $c \cdot u$ for the frequently overwritten memory. Combined they sum to $O(\frac{m \cdot u \cdot n}{c} + c \cdot u)$ which is minimized when the two terms are equal: $m \cdot u \cdot n/c = c \cdot u$ and therefore we choose $c = \sqrt{m \cdot n}$. Since $u = O(n/m)$ this gives a total space complexity of $O(n^{1.5}/\sqrt{m} + m^2)$. As the algorithm backtracks to find the

Algorithm SEQUESTER
Given: set $V = \{1, \ldots, n\}$ of people, EPIs s_i for each $i \in V$, and k groups of size at most m
Output: partition of V into groups of size at most m, so that the final expected outbreak size is minimized.

1. Define $OPT(a, b)$ to be the expected number of finally-infected people, in an optimal solution for the problem restricted to the people indexed $\{1, 2, \ldots, a\}$, and using groups $\{1, 2, \ldots, b\}$ (for any a, b that satisfy $a \leq mb$). Let $g(S)$ denote the expected number of infections if the group of individuals S is put in one group.
2. Sort the people in V, such that $i < j \rightarrow s_i < s_j$.
3. For all b set $OPT(0, b) = 0$.
4. For $b = 1, \ldots, k$ and for $a = 1, \ldots, n$, compute
$$OPT(a, b) = \min_{i=0}^{m}[OPT(a - i, b - 1) + g(\{a - i + 1, a - i + 2, \ldots, a\})].$$
5. $OPT(n, k)$ gives the expected infection size of an optimal sequestering, and tracing back through OPT reveals the partitioning which achieves that value.

Figure 3: Algorithm SEQUESTER for the simplest setting in which all allowed group capacities are uniform, though the final group sizes need not be uniform. The algorithm is a dynamic program, based on the recursive expression for the optimum.

optimal assignment, each section of c people must be recomputed exactly once from the saved m people before it. This method results in a factor of 2 increase in running time, which is a reasonable trade-off.

4. EXPERIMENTAL ANALYSIS

In this section we examine the behavior of our algorithm when applied to a few artificial datasets. We look at three specific aspects:

- How much better is an optimal solution than a random solution?
- How sensitive are our results to errors in the EPI estimates?
- What are the trade-offs between when sequestering is triggered, group size, and final infection rates.

4.1 Optimal Versus Random Placement

First we study how the EPI estimates can be used by our algorithm. We evaluate the effectiveness of our algorithm, relative to a random assignment which does not use the EPI estimates, and find that the random assignment could lead to outbreaks which are twice as large as our algorithm or larger. Next, we study the sensitivity of our algorithm to the accuracy of the EPI estimates, and find that it is fairly robust. Thus, the EPI estimates provide valuable information to policy planners.

In our study, we assume that a critical population (e.g., the military) is sequestered at a base a few days after the onset of an outbreak of a simulated disease in the general population. We assume that the EPI estimates of such individuals are known; and for our simulations we assume the EPIs are exponentially distributed. For a given maximum group size (we use sizes of $m = 20$, 30, or 50), we find the optimal assignment using our algorithm, and compare it with a random assignment. Figure 5 shows a histogram of the ratio of an optimal solution to a random solution for a large number of simulations using a variety of values for p and quantized exponential like distributions for the EPIs. The optimal sequestering by our algorithm is up to 50% better than a random sequestering, and often at least 25% better. We see the best improvements at moderate disease transmission probabilities, and the worst at the extremes of low or high transmission probabilities. This is because moderate transmission probabilities permit the most room for improvement. With low enough probabilities, connected components are very small, effectively isolating sick individuals from others in the room. When probabilities are large enough, rooms become almost fully connected, and most of the population falls ill despite our best efforts.

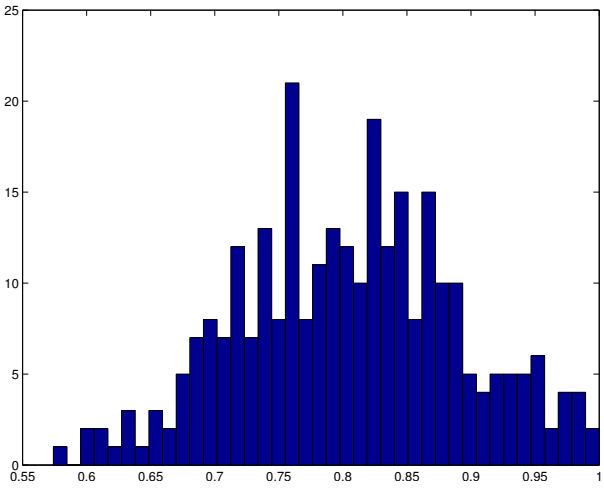

Figure 5: Histogram of optimal sequestering's ratio of expected infection size over that of random sequestering.

Algorithm for computing $g(S)$.
Given: set $S \subseteq V$ forming a single group
Output: the expected number of infections, $E[X_S]$, in S.

1. Initialization the array $P(j, x, y)$ to contain the probability that there are x infected nodes at distance at most j from one of i initially infected nodes, and y nodes at exactly distance j in a $G(|S|, p)$ random graph.
2. Initialize $P(0, i, i) = 1$ and $P(0, *, *) = 0$ for all other entries.
3. For each j from 1 to $|S|$,
$$P(j, x, y) = \sum_{0 \le z \le x-y} P(j-1, x-y, z) \binom{\ell - x + y}{y} ((1-p)^z)^{\ell - x} (1 - (1-p)^z)^y.$$
Save the array $B(i) = \sum_{x=0}^{|S|} x \cdot \sum_{y=0}^{|S|} P(|S|, i, y)$.
4. Upon each invocation of $g(S)$, compute $A(i, j)$ – the probabilities that there are i initial infections among the first j people.
5. $A(0, 0) = 1$, $A(*, 0) = 0$ otherwise, and $A(i, j) = A(i, j-1) \cdot (1 - s_j) + A(i-1, j-1) \cdot s_j$. Computing A in this way computes values for several subsets at once, and they can be stored between calls to g.
6. $g(S) = \sum_{i=1}^{|S|} A(i, |S|) * B(i)$.

Figure 4: Algorithm for computing $g(S)$ exactly, though in practice a Monte-Carlo estimate can be used.

4.2 EPI Error Sensitivity

This sequestering scenario has an obvious susceptibility to errors in the EPI estimates. For example, if the EPI estimates are no better than random, an optimal sequestering with respect to those estimates would be essentially random with respect to the true values. To examine the effect of estimation error we performed a number of simulations where we found the optimal sequestering assignment for a set of EPIs, and then compared that assignment's cost to the cost of the same assignment but with the EPIs perturbed. The first way we perturbed the EPIs was to change each one by up to 30% while keeping the sum of the EPIs constant. This set was designed to measure random, directionally symmetric errors. In all of our simulations the relative change (absolute value of perturbed minus optimal divided by optimal) was at most .0012. This tells us that our assignments are very insensitive to random, symmetric errors.

We also examine the effect of one sided estimation error. Using the same sets of parameters from the symmetric case, we separately examine the effect of increasing and decreasing the EPIs after an optimal assignment is made. When we increase all of the EPIs by 30%, we see a maximum increase over the estimated values in the expected infected set of at most 23.4% with an average increase of 8.7%. When we decrease the EPIs by 30% we see a decrease in the expected infected set of at most 26% with an average decrease of 11%. While the 30% increase is an upper bound (in general a factor of x bound on EPI errors yields at most of factor of x bound on epidemic size error because in any single connected component two nodes with increased EPIs will partially cancel each others increase out), it is interesting to note that in both cases the effect of one sided error is significantly less than the factor of perturbation of the EPIs. A full plot of the scaling factors in our simulations is presented in Figure 6. In summary, neither under nor over-estimating the EPI values has a compounding effect on our results, and symmetrical incorrect estimates cancel out in practice. Therefore our sequestering scheme is fairly insensitive to estimation errors.

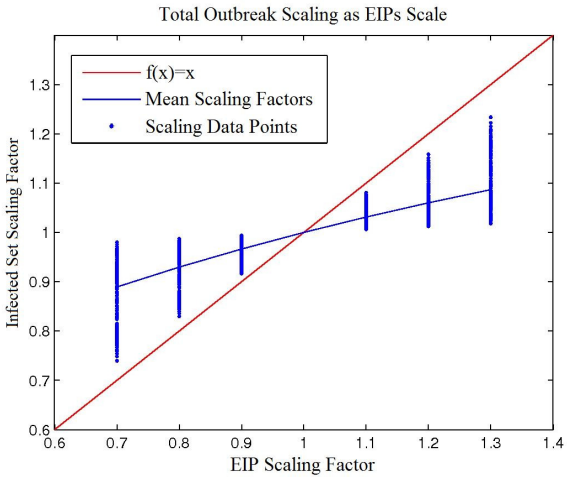

Figure 6: Outbreak scaling as a function of EPI scaling.

4.3 Trade-offs in Sequestering

We conclude this section by empirically showing some of the trade-offs involved in an implementation of protective sequestering.

In this section we use *Simfrastructure*, *Simdemics* and *EpiFast*, high performance computing (HPC)-based modeling environments [3, 6] that provide very detailed and disaggregate individual interaction models of large populations, contagious diseases, and interventions. They represent the current state of a long development process for HPC-based complex system modeling and have been used in a variety of previous studies [4, 6, 5, 10]. They allow one to combine very detailed and realistic population-level data with models of within-host disease progression and contagious disease transmission. Simdemics uses highly resolved population-activity models that yield detailed and representative urban-scale social contact networks. The networks are produced using synthetic information methods in Simfrastructure. In the studies described here using Simdemics, we implement and solve `OSP` for a socially essential population of approximately 188,000 individuals embedded in a representative and realistic urban social-contact network of approximately 3.7 million in the Washington D.C. metro area. These individuals have demographic characteristics that are representative of critical sub-population (e.g. defense force, critical social workers, etc).

To establish the statistical properties of the system's range of behavior, an epidemic outbreak over a large population is simulated for each experimental case for 50 iterations with identical conditions and different random seeds. Infections within the essential subpopulation are tallied for each day of a 254-day epidemic. This gives us, for every day t, the distribution of the number of infections within the socially essential subpopulation on that day. This is used to estimate the EPI for a given individual on any day t. We then simulate a triggered decision to sequester the protected population on each successive day t and compute the total number infected in that subpopulation during the entire epidemic as a consequence of this decision. This amounts to a controlled representation of a decision to trigger sequestering at that day. It allows us to establish the effects of what different threshold triggers would be, had they been given a priori and used by the authorities. Therefore we can experimentally compare *in silico*, the effects of different triggers.

Next, we consider group size. Because the number of people in a sequestered group strongly affects the impact of sequestering an infectious person in that group, we include a sweep of this factor in our experimental design. In the simulation, we sequester the protected population into group sizes of 30, 50, or 70. Transmissibility of diseases vary, so the experimental design also factors person-to-person transmission rates (defined as the probability per unit of contact time that an uninfected person will catch the disease from a nearby infectious person) of 0.05 and 0.1. These values correspond roughly to infection incidences of $4{*}10^{-5}$ per minute over 20 and 40 hours respectively (a rate calibrated to infect roughly quarter of the general population). To compute the number of infections if we start sequestering on day t, we run our optimal partition algorithm on the subpopulation given the EPIs for that day. We then add the expected number of people infected within the sequestered groups to the number infected before sequestering began. The results are shown in Figure 7.

These results show that sequestering is most effective when triggered before the disease has spread very much (when EPIs are low) and when the outbreaks within groups are likely to be small (when the transmissibility times group size is small). In fact, unless these two factors are kept small enough, sequestering may lead to more infections within the critical subpopulation. These plots also suggest a trade-off between group size, latent infection rate, and final infection rate.

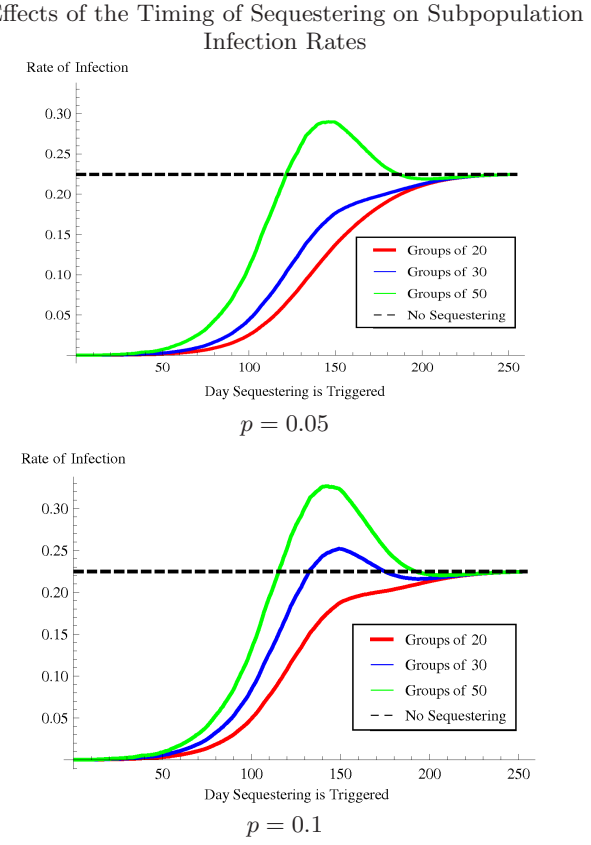

Figure 7: For every day of a simulated epidemic, these plots show the fraction of the socially essential population that get sick during the epidemic, if we start sequestering on that day. Group sizes of 20, 30, and 50 are shown, along with the baseline case without sequestering. Transmissibilities on the left are $p = 0.05$ and $p = 0.1$ on the right.

In any sequestering scenario, total infections can be reduced by using smaller group sizes (which may be more expensive logistically) or triggering when there are lower EPIs (leading to a higher rate of unnecessary sequesterings). It is important to understand how each of these three factors influences the others. We conduct a series of simulations varying both EPIs and group sizes, which for each experiment are uniform throughout the population. For each simulation, we record the fraction of the critical population which becomes ill at any time.

We examine this trade-off in greater detail in Figure 8. With EPIs of s and groups of size g, each group has $s \cdot g$ initial infections in expectation. For any group size g and transmission probability p, we find that the final infection rate varies roughly linearly with s up until most of each group becomes infected. The multiplicative factor between s and final infection rates depends heavily upon g and p. When $g \cdot p \ll 1$, initial infections tend to infect only a small

number of others. Conversely when $g \cdot p \gg 1$, even a single initial infection will likely lead to much of the group becoming ill, and thus the multiplier approaches g. From this we derive the heuristic to keep either $s \cdot g \ll 1$ so that many groups contain no infectious people or $g \cdot p \ll 1$ so that each initial infection leads to only a small number of final infections.

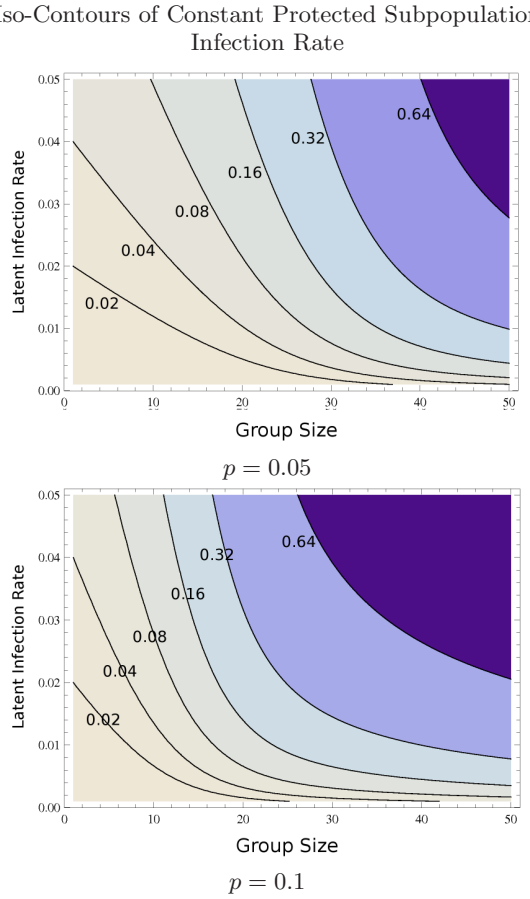

Figure 8: The contour lines indicate equal infection rates as group size and latent infection rate vary. The key observation here is that the higher the latent infection rate, the more important group sizing becomes. If we trigger sequestering late, we can make up for it to a point, but only with significantly smaller group sizes or settling for much higher infection rates.

5. CONCLUSION

Planning for a severe epidemic involves many public health considerations, the understanding of which is continually improved by epidemiological research. Additionally, and outside the scope of questions concerning control of a contagious disease, societal resilience requires the protection of certain identifiable individuals who have difficult-to-replace skills; their protection during catastrophic events will require planning and commitment. The problem we introduce herein has practical roots and can present the issue to decision-makers in a way that can support policy decisions, yet requires a sophisticated analytical foundation to support them scientifically. We have shown that preventing a loss of essential skills through protective sequestering is a very different analytical problem than those normally addressed by epidemiological research. Moreover, the problem is technically difficult and subtle. Our studies highlight strategies based on theoretical and analytical research but also demonstrate the practical importance of detailed, computationally-enabled informatic methods to inform decision makers responsible for plans and actual implementation.

As a demonstration, we analyzed an influenza-like disease in a single large American city. Nevertheless the broader implications go beyond this scenario and we believe the concepts, methods, theoretical setting and qualitative aspects of the results clearly are applicable to a wide range of similar critical national and international problems.

We conclude that sequestering essential workers early can effectively protect them from infection. It is also important to start sequestering early in an epidemic – before too many in the protected subpopulation contract the disease. Additionally, using small group sizes and removing individuals who become symptomatic reduces the incidence of large outbreaks. If the latent infection probabilities of individuals can be estimated or if an approximate ordering of individuals by these probabilities can be obtained then outbreak sizes within the essential subpopulation can be further reduced by as much as 25% or more.

Protection of the essential subpopulation will not usually affect the overall population's epidemic outcomes. The use of resources for protecting socially essential subpopulations must be evaluated with respect to the value of the protected functions, e.g., health care delivery, emergency response, etc. and outside the context of traditional disease control alone. Protective sequestering will compete for resources with conventional public health interests. The trade-offs involved in this allocation of resources pose an important open problem going forward.

Our work provides a solid scaffolding for several areas of future research into the practical application of sequestering in an emergency situation. Perhaps the most important considerations are the trade-offs involved in diverting resources from the protection of the general population to the protection of the critical population. There are other considerations as well. Within a sequestered population, it is likely that additional countermeasures will be available to further contain the epidemic. How should these countermeasures be allocated? Complete isolation of sequestered groups is impractical for any length of time, can we accurately account for the affect of the critical populations contact with the outside world either through the course of their jobs or their logistical support? Can we account for different roles within the critical population - where we wish to minimize the overall infection rate, but the infection rates of several different categories of critical workers?

6. ACKNOWLEDGMENTS

This work is supported in part by NSF ITR Award CNS-0426683, NSF Award CNS-0626636, NSF Award CNS-1010789, U.S. Army Research Office grant W911NF1010350, NIH MIDAS project 2U01GM070694-7, NSF PetaApps Grant OCI-0904844, DTRA R&D Grant HDTRA1-0901-0017, DTRA CNIMS Grant HDTRA1-07-C-0113, NSF Netse CNS-1011769, NSF SDCI OCI-1032677, NSF CNS-0845700, and DOE DE-SC0003957.

7. REFERENCES

[1] K. Atkins, C. L. Barrett, R. J. Beckman, et al. An analysis of public health interventions at military bases during a pandemic influenza event. Technical Report 07-019, NDSSL, 2007.

[2] K. Atkins et al. National force readiness during an infleunza pandemic. Technical Report 06-060, NDSSL Technical Report, Prepared at the request of the Defense Threat Reduction Agency,, 2006.

[3] C. L. Barrett, K. R. Bisset, S. Eubank, X. Feng, and M. V. Marathe. Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In *Proceedings of the ACM/IEEE Conference on High Performance Computing (SC)*, page 37, 2008.

[4] C. L. Barrett et al. TRANSIMS: Transportation Analysis Simulation System. Technical Report LA-UR-00-1725, Los Alamos National Laboratory Unclassified Report, 2001. Volume 3, available at http://ndssl.vbi.vt.edu/transims.php.

[5] K. Bisset, X. Feng, M. Marathe, and S. Yardi. Modeling interaction between individuals, social networks and public policy to support public health epidemiology. In *Proceedings of the 2009 Winter Simulation Conference*, 2009.

[6] K. R. Bisset, J. Chen, X. Feng, V. A. Kumar, and M. V. Marathe. Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In *ICS '09: Proceedings of the 23rd international conference on Supercomputing*, pages 430–439, New York, NY, USA, 2009. ACM.

[7] D. Defoe and E. Brayley. *A journal of the plague-year; or, Memorials of the great pestilence in London, in 1665*. Routledge, 1882. available at http://books.google.com/books?id=kpENAAAAYAAJ.

[8] C. DiGiovanni, A. Monto, and J. Malone. Mitigation of threats to the continuation of marine recruit training posed by category 4/5 influenza pandemic. Technical Report 07-006, Defense Threat Reduction Agency, DTRA, 2007.

[9] K. Earhart, C. Beadle, L. Miller, M. Pruss, G. Gray, E. Ledbetter, and M. Wallace. Outbreak of influenza in highly vaccinated crew of U.S. Navy ship. *Emerging Infectious Diseases*, 7:463–465, 2001.

[10] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.

[11] W. O. Kermack and A. G. McKendrick. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772):700–721, 1927.

[12] S. Kilic and G. Gray. Nonpharmaceutical interventions for military populations during pandemic influenza. *TAF Prev Med Bulletin*, 6(4):285–290, 2007.

[13] H. Markel, A. Stern, J. Navarro, J. Michalsen, A. Monto, and C. DiGiovanni Jr. Nonpharmaceutical influenza mitigation strategies, US communities, 1918-1920 pandemic. *Emerg Infect Dis*, 12(12):1961–1964, 2006.

[14] M. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[15] U.S. Department of Health and Human Services. HHS Pandemic Influenza Plan. `http://www.hhs.gov/pandemicflu/plan/`.