Minimizing the Social Cost of an Epidemic

Elizabeth Bodine-Baron, Subhonmesh Bose,
Babak Hassibi, and Adam Wierman

California Institute of Technology,
Pasadena, CA 91125, USA
{eabodine,boses,hassibi,adamw}@caltech.edu

Abstract. In this paper we quantify the total cost of an epidemic spreading through a social network, accounting for both the immunization and disease costs. Previous research has typically focused on determining the optimal strategy to limit the lifetime of a disease, without considering the cost of such strategies. In the large graph limit, we calculate the exact expected disease cost for a general random graph, and we illustrate it for the specific example of an Erdös-Rényi network. We also give an upper bound on the expected disease cost for finite-size graphs, and show through simulation that the upper bound is tight for Erdös-Rényi networks and graphs with exponential degree distributions. Finally, we study how to optimally perform a one-shot immunization to minimize the social cost of a disease, including both the cost of the disease and the cost of immunization.

Keywords: Epidemic, immunization, information cascade, random matrix theory, generalized random graphs.

1 Introduction

The spread of epidemics over networks has been extensively studied since the Kermack-McKendrick SIR (Susceptible-Infectious-Recovered) model was proposed in 1927 [15]. Though initially the study of epidemics in networks was primarily motivated by the spread of disease, the results have had far reaching applications: the insights learned apply to many other settings where something is spreading over a network. For example, applications such as (i) network security, where the goal is to understand and limit the spread of computer viruses [7], [28], [10] (ii) viral advertising, where the goal is to create an epidemic to propagate interest in a product [24], [25], and (iii) information propagation, where the goal is to understand how quickly new ideas propagate through a network [12], [14], [5]. See [13] and [11] for comprehensive surveys of prior results. Note that in this paper we use the language of epidemiology, however the results and techniques can easily be seen to hold more broadly.

Typically, work in this area has focused on understanding two broad questions: First, what is the lifetime of an epidemic, and how does it depend on the structure of the network? Second, what is the optimal strategy to minimize/maximize the spread of the epidemic? The first type of research looks to determine the

"epidemic threshold" – the infection probability below which the disease will eventually die out. In [6] and [29], the authors show that this epidemic threshold depends on the largest eigenvalue of the adjacency matrix of the network graph, generalizing earlier results. In a series of papers, [20], [21], [2], [1], the authors show the absence of such an epidemic threshold in the presence of scale-free networks; the disease will not die out if it is even the smallest bit infectious. In the second branch of the research on modeling epidemics, the authors focus on ways to minimize or maximize the spread of an epidemic by targeting specific nodes. In [6], the authors use their eigenvalue model to predict which nodes are the best to immunize; in [23], the authors use similar methods to evaluate various immunization strategies and determine which is optimal. In contrast, [7] and [18] focus on real-world immunization strategies, where determining the optimal nodes to immunize might not be practical; they conclude that a good solution is to target various types of high-degree nodes. [22] comes to a similar conclusion in the case of scale-free networks. [17] proposes a static bond-percolation model for studying the classic SIR model and evaluates various immunization strategies within this model.

In this paper, we take a different, non-standard goal: instead of focusing on the spread of an epidemic, we focus on the "social cost" of an epidemic. This shift in focus is motivated by a growing trend in the medical community to quantify the cost of an epidemic by looking at the direct and indirect medical costs to both the hospitals and doctors treating and immunizing a population, as well as the cost to individuals in the population paying for medical care. See [4] and [26] for two examples of such studies. It is not just within the medical community that a focus on cost of epidemics is emerging. For example, within the computer science community [8] has recently studied the social cost of an epidemic, although in a very different model than studied in the current paper. In particular, their model assumes a single infected node and determines the maximum infection (reach) from that node. Note that looking at social cost is in general more difficult than studying the lifetime or spread of an epidemic, as it involves the transient, not just long-term, behavior of a disease.

To study the social cost of epidemics, we use a general random graph model, which can incorporate graphs with arbitrary degree distributions. The model generalizes the configuration model [19], and the details are provided in Section 2. To model the spread of an epidemic we use a standard model of, e.g., [29], [6], and [23], which is an approximation of the SIS model.

Within this context, the main contribution of the paper is the derivation of (i) the exact cost of an epidemic in the large graph limit (Theorem 1) and (ii) bounds on the cost of an epidemic for finite graphs (Theorem 3). To illustrate these results, we look at an Erdös-Rényi network and graphs with an exponential degree distribution.

To illustrate the usefulness of the cost calculations just described, we study immunization strategies for balancing the cost of disease with the cost of immunization in Section 4. In particular, we derive the optimal immunization strategy in a simple one-shot setting where immunization happens only at time zero. Our

results allow the study of both random and degree-based immunization strategies. In this context, we derive the optimal immunization strategy in the large graph limit, and illustrate using simulations that it still provides a near-optimal choice for finite graphs.

2 Model Description

In this section we introduce the network model and epidemic spread model studied in this paper.

Network model. Let there be n nodes in the population, and let A be an $n \times n$ adjacency matrix corresponding to the network. We consider only undirected graphs. We assume that the network is drawn from a general class of random graphs, $\mathcal{G}$. For example, the network represented by A could be a realization of an Erdös-Rényi random graph, $G_{n,p}$, which would correspond to allowing each edge to exist independently with probability p.

We focus on a class of random graphs, which is related to the 'configuration' model [19] and the, so-called, 'general random graph' model [9], but which is slightly more general than each.

The construction of the graph works as follows. First, define a degree distribution $p(w)$, and obtain n i.i.d. samples $w = (w_1, \ldots, w_n)$. From this vector, generate a random graph given by the adjacency matrix:

$$A_{ij} = A_{ji} = \begin{cases} 1 & \text{w.p. } w_i w_j \rho \\ 0 & \text{w.p. } 1 - w_i w_j \rho \end{cases} \quad \text{where } \rho = \frac{1}{\sum_i w_i}. \tag{1}$$

Note that the expected degree of node i is $\sum_j w_i w_j \rho = w_i$. Since this model is fully determined by one degree distribution $p(w)$, for ease of reference, we call it $G_{n,p(w)}$.

To generate the Erdös-Rényi random graph, $G_{n,p}$, we would need $w = (np, np, \ldots, np)$. With this notation, we have $G_{n,p} = G_{n,\delta(w-np)}$, where $\delta(\cdot)$ is the Dirac δ-function.

The advantage of this model is that it can generate a random graph with a specified degree distribution, allowing us to examine very general random graph models. To relate this model to the configuration model [19] and the general random graph model [9], note that in the case of the configuration model the degree sequence is enforced deterministically and that in the general random graph model the expected degree sequence is fixed rather than the distribution.

Epidemic model. To model an epidemic, we consider an initial set of seed nodes that are infected at time $t = 0$ and then we consider a discrete-time evolution as follows. At $t = 0$, αn nodes are uniformly at random chosen to be infected. Then, the infection spreads according to the following, standard SIS-inspired model. Define, $P(t)$ a vector where the ith entry is the the probability that node i is infected at time step t. Then the evolution proceeds according to

$$P(t+1) = MP(t) \tag{2}$$

where $M = (1 - \gamma)I + \beta A$ is termed the 'system matrix', γ is the probability that a node which is infected recovers in within the current time-step, and β is the infection rate, i.e., if a node has k infected neighbors then the probability it is infected in the current time-step is $\beta k + o(k)$.

This model is an approximation of the SIS (Susceptible-Infected-Susceptible) model where each node gets infected/infectious and then recovers to become susceptible to the disease again, using the approximation and independence assumptions made in [29], [6] and [23]. We give a more detailed analysis of the approximations inherent in this model in the extended version of this paper [3].

3   Cost of Disease

The primary goal of this paper is to derive an expression for the cost of an epidemic. Thus, we first need to model the cost of a disease. To do this, we assume a simple model where, c_d, is defined as the cost of an individual being infected during a time-step. Thus, c_d captures both the direct costs to the individual for medication, doctor visits, etc., as well as secondary costs such as missed work. We note that this model leaves open the question of how exactly to determine the parameter c_d. This is ongoing work in the medical community; see [4] and [26] for example studies in this area. In future work, we look to incorporate these results into determining the per-time cost of disease, but for the purposes of this paper, we leave it a general parameter of the model.

Given this model for the cost of disease to an individual, we can begin to calculate the cost of an epidemic. Define C_D, the 'disease cost', as the expected (averaged over the random spread) per node disease cost of an epidemic *during its entire course* when a fraction $\alpha < 1$ of the nodes are infected at time $t = 0$. Note that the expected per node disease cost in a *given time-step t* is simply $\frac{1^T P(t)}{n}$. Furthermore, since $P(0) = \alpha 1$ and $P(t) = M^{t-1}P(0)$, we can express the disease cost per node as

$$C_D = \frac{1}{n}\left[1^T\left(\sum_{t=0}^{\infty} M^t\right)1\alpha c_d\right]. \tag{3}$$

When the infinite sum converges, i.e., when $||M|| < 1$, we have

$$C_D = \frac{1}{n}\alpha c_d \cdot \left[1^T(I - M)^{-1}1\right]. \tag{4}$$

We emphasize that, when the underlying graph is drawn randomly from $G_{n,p(w)}$, say, the above C_D is a random variable.

The remainder of this section focuses on understanding (4). Though we cannot explicitly calculate it in all settings, it turns out that when $n \to \infty$ it converges almost surely to a deterministic constant which we can explicitly compute. Throughout, we use the Erdös-Rényi network as an illustrative example.

3.1   Exact Disease Cost for Large n

We first state a result for general random graphs.

Theorem 1. *For $G_{n,p(w)}$ where $p(w)$ has finite variance and M is almost surely stable*

$$\lim_{n\to\infty} C_D(n) = \alpha c_d \left(\frac{1+F^2}{\gamma} - \frac{F^2}{\gamma - \beta - \beta\sqrt{\langle w\rangle}/F} \right) \quad a.s.$$

$$F = \int_1^\infty \frac{wp(w)dw}{\gamma\sqrt{\langle w\rangle}/\beta - Fw} \quad and \quad \langle w\rangle = \int_1^\infty wp(w)dw$$

We omit the proof for brevity; see the extended version of this paper, [3], for details. However, in order to demonstrate the idea of the above proof, we show the calculation of the expected per node disease cost for the case of an Erdős-Rényi network, where $p(w) = \delta(w - np)$. The proof for the general model is parallel but requires more detailed technical arguments.

Theorem 2. *For $G_{n,\delta(w-np)}$ with $\gamma > \beta np$, M is a.s. stable and*

$$\lim_{n\to\infty} C_D(n) = \frac{\alpha c_d}{\beta\sqrt{\langle w\rangle}} \left(\frac{F}{1 - \sqrt{\langle w\rangle}F} \right) \quad a.s.$$

where $F = \dfrac{\gamma - \sqrt{\gamma^2 - 4\beta^2\langle w\rangle}}{2\beta\sqrt{\langle w\rangle}}$ *and* $\langle w\rangle = np$.

Proof. We can write the system matrix M as

$$M = (1-\gamma)I + \beta A = (1 - \gamma - \beta p)I + \beta \underbrace{(A - p11^T + pI)}_{=G} + \beta p 11^T$$

$$= (1-\gamma)I + \beta G + \beta p 11^T.$$

Note that G is a zero-mean matrix with i.i.d. lower triangular entries (taking the values $-p$ with probability $1-p$ and the value $1-p$ with probability p). The variance of the entries of G are

$$\mathrm{e}G_{ij}^2 = p^2(1-p) + (1-p)^2 p = p - p^2 \tag{5}$$

Using the Matrix Inversion Lemma, we have

$$1^T(I-M)^{-1}1 = 1^T \left(\underbrace{(\gamma + \beta p)I - \beta G}_{=B} - \beta p 11^T \right)^{-1} 1 = 1^T B^{-1}1 - \frac{1^T B^{-1}11^T B^{-1}1}{-\frac{1}{\beta p} + 1^T B^{-1}1}.$$

Let

$$s = \frac{1}{n}1^T B^{-1}1 = \frac{1}{n}1^T \left((\gamma + \beta p)I - \beta G \right)^{-1} 1,$$

then

$$C_D(n) = \frac{\alpha c_d}{n} \left[1^T (I - M)^{-1} 1 \right] = \left(\frac{\alpha c_d s}{1 - \beta p n s} \right)$$

and

$$\lim_{n \to \infty} C_D(n) = \lim_{n \to \infty} \left(\frac{\alpha c_d s}{1 - \beta p n s} \right). \tag{6}$$

Let $H = \frac{1}{\sqrt{(p - p^2)n}} G$. Then H is a symmetric matrix whose lower triangular entries are i.i.d. random variables with zero-mean and $\frac{1}{n}$ variance. Such a random matrix is called a *Wigner matrix*. We use the following Lemma to finish the calculation. □

Lemma 1. *Let H be a symmetric $n \times n$ random matrix whose lower triangular entries are i.i.d. random variable with zero-mean and $\frac{1}{n}$ variance. Then*

$$\lim_{n \to \infty} \frac{1}{n} 1^T (xI - H)^{-1} 1 = \lim_{n \to \infty} e \frac{1}{n} \operatorname{tr}(xI - H)^{-1} \, a.s.$$

$$= \begin{cases} \frac{x - \sqrt{x^2 - 4}}{2}, & |x| \geq 2 \\ \infty, & \text{otherwise} \end{cases}$$

We remark that the complex function $\lim_{n \to \infty} e \frac{1}{n} \operatorname{tr}(zI - H)^{-1} = e \frac{1}{z - \lambda}$ is often called the *Stieltjes transform* of the random matrix H (or of its eigenvalue distribution). For a more detailed introduction to random matrices and Stieltjes transform, we refer the reader to [27]. It remains to calculate s.

$$\lim_{n \to \infty} s = \lim_{n \to \infty} \frac{1}{n} 1^T \left((\gamma + \beta p)I - \beta G \right)^{-1} 1$$

$$= \lim_{n \to \infty} \frac{1}{\beta \sqrt{(p - p^2)n^3}} \left(1^T \left(\frac{\gamma + \beta p}{\beta \sqrt{(p - p^2)n}} I - H \right)^{-1} 1 \right)$$

Using Lemma 1, we get

$$\lim_{n \to \infty} s = \lim_{n \to \infty} \frac{\frac{\gamma + \beta p}{\beta \sqrt{(p - p^2)n}} - \sqrt{\frac{(\gamma + \beta p)^2}{\beta^2 (p - p^2)n} - 4}}{2 \beta \sqrt{(p - p^2)n}}$$

$$= \lim_{n \to \infty} \frac{\gamma + \beta p - \sqrt{(\gamma + \beta p)^2 - 4\beta^2 (p - p^2)n}}{2\beta^2 (p - p^2)n}. \tag{7}$$

Since M is stable *a.s.*, $\gamma > \beta n p$. Continuing from (7) and removing $o(n)$ and $o(p)$ terms, we get

$$\lim_{n \to \infty} s = \frac{\gamma - \sqrt{\gamma^2 - 4\beta^2 \langle w \rangle}}{2\beta^2 \langle w \rangle} = \frac{F}{\beta \sqrt{\langle w \rangle}} \quad a.s. \tag{8}$$

where $\langle w \rangle = np$ and finally

$$\lim_{n \to \infty} C_D(n) = \lim_{n \to \infty} \left(\frac{\alpha c_d s}{1 - \beta p n s} \right) = \frac{\alpha c_d}{\beta \sqrt{\langle w \rangle}} \left(\frac{F}{1 - \sqrt{\langle w \rangle} F} \right) \quad a.s. \tag{9}$$

□

3.2   Bounds for General n

The prior section provided exact calculation of the disease cost in the case of large n. We now provide bounds for finite n. Note that these bounds hold for all graphs, not only the random graphs described in Section 2.

Theorem 3. *For a graph with stable system matrix M, the cost of disease per node satisfies*

$$C_D \leq \frac{\alpha c_d}{1 - \lambda_{max}(M)} \tag{10}$$

Proof. From Perron-Frobenius theorem, it follows that

$$-1 < -\lambda_{max}(M) \leq \lambda_{min}(M) \leq \lambda_{max}(M) = |\lambda_{max}(M)| < 1 \tag{11}$$

Note that $(I - M)^{-1}$ is a positive definite matrix since all eigenvalues are positive. From (11), we have

$$\frac{1}{1 - \lambda(M)} \leq \frac{1}{1 - \lambda_{max}(M)}$$

and

$$1^T (I - M)^{-1} 1 \leq 1^T \left(\frac{I}{1 - \lambda_{max}(M)} \right) 1 \tag{12}$$

This follows from the fact that if A and B are positive definite matrices with eigenvalues $\lambda_A \leq \lambda_B$ for all eigenvalues λ_A and λ_B, then $x^T A x \leq x^T B x$. The bound follows.   □

We note that the necessary and sufficient condition required for the disease to die out and the social cost to converge are the same, i.e., $\lambda_{max}(M) < 1$. Also, the bound only depends on $\lambda_{max}(M)$ from the disease propagation model, which is popularly known as the disease threshold. It is interesting that the same parameter of the disease plays the central role in both tapering off the disease and its total cost.

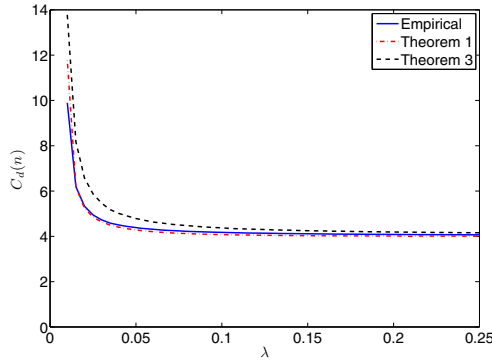**Comparison of Theorem 1 and Theorem 3**
Theorem 1 involves the entire eigen-distribution of M, whereas Theorem 3 is dependent only on the maximum eigenvalue of M. We show here, through simulation, that in the case of an exponential $p(w)$ and for an Erdös-Rényi network, the upper bound in Theorem 3 is rather tight.

Define $p(w) = \lambda e^{-\lambda w}$; i.e., $p(w)$ is exponentially distributed with parameter λ. According to Theorem 1, the cost of the disease in the limit of large n will be
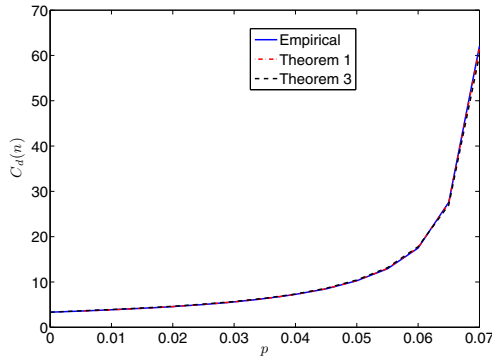
$$\lim_{n \to \infty} C_D(n) = \alpha c_d \left(\frac{1 + F^2}{\gamma} - \frac{F^2}{\gamma - \beta - \frac{\beta}{F}\sqrt{\frac{1}{\lambda}}} \right)$$

where

$$F = \int_1^\infty \frac{w \lambda e^{-\lambda w} dw}{\frac{\gamma}{\beta} \sqrt{\frac{1}{\lambda}} - F w},$$

(a) Exponential $p(w)$

(b) Erdös-Rényi

Fig. 1. Comparison of Theorem 1 and Theorem 3

as $\langle w \rangle = 1/\lambda$. Using the parameters $n = 1000$, $\alpha = 0.2$, $\beta = 0.002$, $\gamma = 0.5$, and $c_d = 10$, we solve for F numerically and show the cost of the disease in the limit of large n as a function of the graph parameter λ in Figure 1a. We also plot the upper bound from Theorem 3 for comparison. The maximum eigenvalue for each λ was calculated as the average maximum eigenvalue of 10 realizations of M.

Now let $p(w) = \delta(w - np)$, for an Erdös-Rényi network. Using the parameters $n = 1000$, $\alpha = 0.2$, $\beta = 0.008$, $\gamma = 0.6$, and $c_d = 10$, we calculate F and $C_d(n)$ according to Theorem 2 and plot the result as a function of the graph parameter p in Figure 1b. Again, we also plot the upper bound from Theorem 3 for comparison. We note that for any probability $p(n)$, $\lambda_{max} = (1+O(1)) \max{(\sqrt{\Delta}, np)}$, where Δ denotes the maximum degree, almost surely [9], [16].

Note that the upper bound is tight, regardless of the parameter λ or p, indicating that only the maximum eigenvalue of M, rather than the entire eigendistribution, plays the central role in determining the cost of the disease.

4   Minimizing Social Cost with Immunization

The key impact of the cost calculations in Section 3 is to allow decisions that trade off immunization costs with disease cost. To illustrate this, in this section we study a simple example of such an optimization.

   We define the "social cost" of an epidemic as the sum of the disease cost and the immunization cost. For simplicity, we assume that immunization or vaccination will occur at $t = 0$ and any immunized node will remain immune for all time, thus incurring a single immunization cost. This cost could represent the cost of a vaccine to an individual or the cost of quarantining. If the node is infected in the starting state and is selected to be immunized, it is quarantined. If it is susceptible, it is vaccinated. We assume that the cost of quarantining and the cost of vaccinating are the same, c_v.

   We can incorporate the immunization process into the random graph model itself. Recall the original graph generation model, where the distribution $p(w)$ is sampled n times to create the degree vector w. Let us first consider a randomized immunization procedure where πn nodes are chosen uniformly at random for immunization. Then, the random graph is formed by sampling $p(w)$ only for the nodes that remain after the immunization. The nodes that are immunized are simply removed from the adjacency matrix A describing the network and the system matrix M, yielding $\tilde{M}$ and $\tilde{A}$. Using this procedure gives $e[\dim \tilde{M}] = (1 - \pi)n := \tilde{n}$. For a degree-based immunization scheme, we can simply truncate the degree distribution and sample as before.

   In general, any one-shot immunization scheme simply results in a transformation of $p(w)$ and the social cost calculations in Section 3 still apply. For simplicity, in this paper, we focus on the case of random immunization of a fraction π of the nodes. The question is how to optimally choose π.

4.1   Social Cost Calculation

To determine the optimal immunization probability, we first need to incorporate the cost of vaccination into the calculations from Section 3. Define the per node social cost for a particular graph as:

$$S(M, \tilde{M}) = \frac{1}{n} \left[(\dim M - \dim \tilde{M})c_v + \left(1^T (I - \tilde{M})^{-1} 1 \right) \alpha c_d \right] \qquad (13)$$
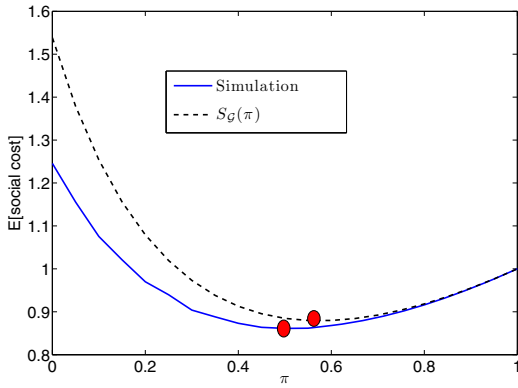
Considering a random immunization on a class of graphs $\mathcal{G}$, we define the expected social cost $S_{\mathcal{G}}(\pi)$ as a function of the fraction π of nodes removed as follows:

$$S_{\mathcal{G}}(\pi) = \pi c_v + (1 - \pi)C_D\left(n(1 - \pi) \right) \qquad (14)$$
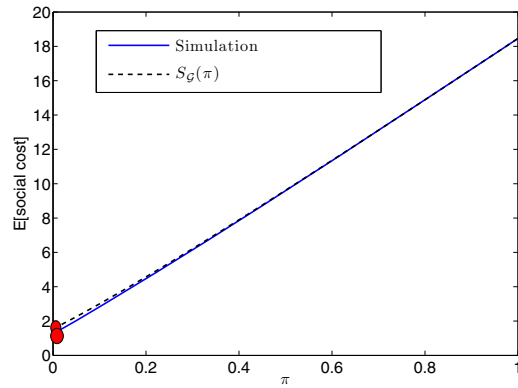
Thus, applying the results from Section 3, we obtain results for $S_{\mathcal{G}}(\pi)$. We illustrate this with the Erdös-Rényi case. Using Theorem 2 for $\mathcal{G} = G_{n, \delta(w - np)}$, we have that, as $n \to \infty$,
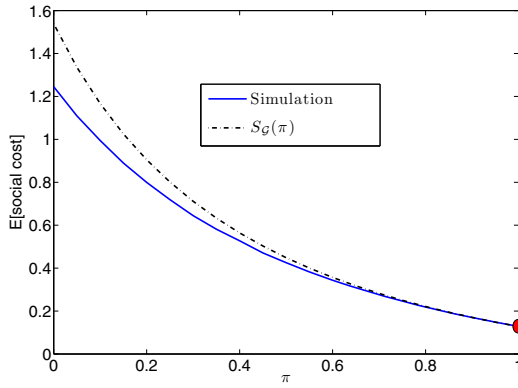
(a) Small C

(b) Medium C

(c) Large C

Fig. 2. Social cost simulations for different values of C on Erdös-Rényi network as a function of the immunization probability π. The optimal immunization probability in each case is highlighted with a red circle.

$$S_{\mathcal{G}}(\pi) \rightarrow \pi c_v + \frac{\alpha c_d}{\beta\sqrt{(1-\pi)np}} \left(\frac{F}{1 - \sqrt{(1-\pi)np}F} \right) \quad \text{a.s.} \tag{15}$$

$$\text{where } F = \frac{\gamma - \sqrt{\gamma^2 - 4\beta^2(1-\pi)np}}{2\beta\sqrt{(1-\pi)np}}.$$

Using $\sqrt{1-x} \approx 1 - \frac{x}{2}$, we get $F \approx \frac{\beta\sqrt{(1-\pi)np}}{\gamma}$. Thus

$$S(\pi) \approx \pi c_v + \frac{\alpha(1-\pi)c_d}{\gamma - (1-\pi)\beta np} \tag{16}$$

4.2   Optimal Immunization

To determine the optimal choice of π in the random immunization scheme for a large network, we use the social cost as calculated in Section 4.1. We illustrate this through the Erdös-Rényi case. For convenience, we normalize $c_d = 1$ and $c_v = C$, and consider the per node cost as given in equation ((16)).

Minimizing $S(\pi)$ over π, we obtain the approximate optimal random immunization, π^*. Define $a = \frac{\alpha}{\gamma}$ and $b = \frac{\alpha\gamma}{(\gamma - \beta np)^2}$, and note that $a < b$ always.

$$\pi^* = \begin{cases} 1 & C \leq a < b \\ 1 - \frac{\gamma - \sqrt{\gamma\alpha/C}}{\beta np} & a < C < b \\ 0 & a < b \leq C \end{cases} \tag{17}$$

To illustrate the above, we simulate a disease propagating on an Erdös-Rényi graph, with $n = 100000$, $p = 1.27 \times 10^{-4}$, $\alpha = 0.2$, $\beta = 0.02$, and $\gamma = 0.39$. We use a low $C = 0.1282$, medium $C = 1$, and high $C = 18.46$. The simulated cost as a function of π is shown in Figure 2, together with the approximate calculated cost as given in ((16)). Note that the simulation is of the disease propagating according to the SIS model, whereas $S_{\mathcal{G}}(\pi)$ as shown is an approximation on the per node social cost assuming infinite size graphs. In addition, the infection propagation model makes several assumptions explored further in [3]. However, note that the optimal π for both the bound and the simulation is very close, indicating the usefulness of the model for immunization decisions.

5   Conclusion and Future Work

In this paper, rather than taking the approach of previous work, which has typically focused on determining the optimal strategy to limit the lifetime of a disease, without considering the cost of such strategies, we look at the total "social cost" of an epidemic, including the cost of immunization and the cost to an individual being infected for a finite amount of time. Determining the social cost

of an epidemic is a harder problem than estimating its lifetime, as it depends on the transient behavior of the epidemic, not just its steady state. This approach can help quantify the efficacy of immunization decisions from an economic standpoint for decision-making bodies like the CDC. It can also be applied to the dual problem, where the desire is to maximize the spread of information while keeping advertising costs low.

We model the physical interaction of individuals (and hence the transmission of disease) as an undirected random graph using a new random graph model and a dynamical system based on the popular SIS model. Within this framework, we derive the exact cost of an epidemic in the large graph limit (Theorem 1) and bounds on the cost of an epidemic for finite graphs (Theorem 3). The exact cost depends on the entire eigendistribution of the system matrix, whereas the upper bound depends only on the largest eigenvalue. Despite its simpler form, the upper bound appears to be tight, as shown by our simulations of Erdös-Rényi networks and graphs with exponential degree distributions. Our analysis makes use of ideas and techniques from random matrix theory, differentiating this work from previous work on the spread of epidemics. The key impact of our theorems calculating and bounding the disease cost is to allow decisions that trade off immunization costs with disease cost - as shown by our calculation of the social cost based on a one-shot random immunization strategy for an Erdös-Rényi network.

We note that the graph model proposed in Section 2 is a generalization of existing models and can capture any class of random graphs defined by an expected degree distribution. As such, it may be applicable to other areas outside of the scope of this paper and is worth further investigation.

To extend this work, we would like to investigate both theoretical and practical refinements of our model. We note that the linear model of disease propagation used here requires independence assumptions that do not hold in general and it would be worthwhile to explore other models that describe infection more accurately. We can also extend our model to include other classic disease models, such as the SIR model and more complicated models of infection. With regard to immunization, we would like to examine a more dynamic strategy where individuals vaccinate themselves at different points in time based on their currently available information (fraction of friends infected, etc.). Moreover, a case study with real data from communicable diseases like influenza or herpes would provide more insight into the accuracy and predictive power of these models in real-world scenarios, as well as help refine the cost of disease which we have assumed is a single parameter, c_d.

References

1. Barthelemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A.: Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks. Phys. Rev. Lett. 92(17) (2004)
2. Boguna, M., Pastor-Satorras, R., Vespignani, A.: Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations. Phys. Rev. Lett. 90(2) (2003)
3. Bose, S., Bodine-Baron, E., Wierman, A., Hassibi, B.: Cost of disease in large networks. Working paper (2010)
4. Bownds, L., Lindekugel, R., Stepak, P.: Economic impact of a Hepatitis A epidemic in a mid-sized urban community: The case of Spokane. Journal of Community Health 38(4) (2003)
5. Cha, M., Mislove, A., Gummadi, K.P.: A measurement-driven analysis of information propagation in the flickr social network. In: Proc. 18th WWW Conf., pp. 721–730 (2009)
6. Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., Faloutsos, C.: Epidemic thresholds in real networks. ACM Trans. on Info. and System Security (TISSEC) 10(4) (2008)
7. Chen, H., Jin, H., Sun, J., Han, Z.: Efficient Immunization Algorithm for Peer-to-Peer Networks. In: Bougé, L., Prasanna, V.K. (eds.) HiPC 2004. LNCS, vol. 3296, pp. 232–241. Springer, Heidelberg (2004)
8. Chen, P., David, M., Kempe, D.: Better vaccinations for better people. In: Proc. EC 2010 (2010)
9. Chung, F., Lu, L.: Complex Graphs and Networks. AMS (2006)
10. Cohen, R., Havlin, S., ben Avraham, D.: Efficient immunization strategies for computer networks and populations. Phys. Rev. Lett. 91(24) (2003)
11. Daley, D., Gani, J.: Epidemic Modeling: An Introduction. Cambridge University Press (2005)
12. Huang, L., Park, K., Lai, Y.C.: Information propagation on modular networks. Phys. Rev. E 73(3) (March 2006)
13. Jackson, M.O.: Social and Economic Networks. Princeton University Press (2008)
14. Jacquet, P., Mans, B., Rodolakis, G.: Information propagation speed in mobile and delay tolerant networks. IEEE Trans. on Info. Theory 56(10), 5001–5015 (2010)
15. Kermack, W.O., McKendrick, A.G.: A Contribution to the Mathematical Theory of Epidemics. Proc. Roy. Soc. Lond. A 115, 700–721 (1927)
16. Krivelevich, M., Sudakov, B.: The largest eigenvalue of sparse random graphs. Combinatorics, Probability, and Computing 12, 61–72 (2003)
17. Madar, N., Kalisky, T., Cohen, R., ben Avraham, D., Havlin, S.: Immunization and epidemic dynamics in complex networks. Euro. Phys. Journal B - Condensed Matter and Complex Systems 38, 269–276 (2004)
18. Miller, J., Hyman, J.: Effective vaccination strategies for realistic social networks. Physica A: Statistical Mechanics and its Applications 386(2), 780–785 (2007)
19. Newman, M.E.J.: Random graphs as models of networks. In: Handbook of Graphs and Networks (2005)
20. Pastor-Satorras, R., Vespignani, A.: Epidemic Spreading in Scale-Free Networks. Phys. Rev. Lett. 86, 3200–3203 (2001)
21. Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics in finite size scale-free networks. Phys. Rev. E 65 (2002)
22. Pastor-Satorras, R., Vespignani, A.: Immunization of complex networks. Phys. Rev. E 65 (2002)

23. Peng, C., Jin, X., Shi, M.: Epidemic threshold and immunization on generalized networks. Physica A: Statistical Mechanics and its Applications 389(3), 549–560 (2010)
24. Phelps, J., Lewis, R., Mobilio, L., Perry, D., Raman, N.: Viral marketing or electronic word-of-mouth advertising: Examining consumer responses and motivations to pass along email. Journal of Advertising Research, 44 (2004)
25. Richardson, M., Domingos, P.: Mining knowledge-sharing sites for viral marketing. In: Proc. 8th ACM SIGKDD, pp. 61–70 (2002)
26. Rubin, R., Harrington, C., Poon, A., Dietrich, K., Greene, J., Moiduddin, A.: The Economic Impact of Staphylococcus aureus Infection in New York City Hospitals. Emerging Infectious Diseases 5(1) (1999)
27. Tulino, A.M., Verdú, S.: Random matrix theory and wireless communications. Comm. Inf. Theory 1, 1–182 (2004)
28. Wang, C., Knight, J.C., Elder, M.C.: On computer viral infection and the effect of immunization. In: Proc. ACSAC 2000, pp. 246–256 (2000)
29. Wang, Y., Chakrabarti, D., Wang, C., Faloutsos, C.: Epidemic spreading in real networks: An eigenvalue viewpoint. In: Proc. Symp. Reliable Dist. Systems (2003)