When Is a Network Epidemic Hard to Eliminate?

Kimon Drakopoulos,[a] Asuman Ozdaglar,[a] John N. Tsitsiklis [a]

[a] Laboratory for Information and Decision Systems; and Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology Cambridge, Massachusetts 02139

Contact: kimondr@mit.edu (KD); asuman@mit.edu (AO); jnt@mit.edu (JNT)

Abstract. We consider the propagation of a contagion process ("epidemic") on a network and study the problem of dynamically allocating a fixed curing budget to the nodes of the graph, at each time instant. For bounded degree graphs, we provide a lower bound on the expected time to extinction under any such dynamic allocation policy, in terms of a combinatorial quantity that we call the resistance of the set of initially infected nodes, the available budget, and the number of nodes n.

Specifically, we consider the case of bounded degree graphs, with the resistance growing linearly in n. We show that if the curing budget is less than a certain multiple of the resistance, then the expected time to extinction grows exponentially with n. As a corollary, if all nodes are initially infected and the CutWidth of the graph grows linearly, while the curing budget is less than a certain multiple of the CutWidth, then the expected time to extinction grows exponentially in n. The combination of the latter with our prior work establishes a fairly sharp phase transition on the expected time to extinction (sublinear versus exponential) based on the relation between the CutWidth and the curing budget.

1. Introduction

We study the dynamic *control* of contagion processes (from now on called *epidemics*) under limited curing resources. Specifically, we study *dynamic* allocation policies that use information on the underlying structure of contacts and on the *infection state* of individuals, and evaluate performance in terms of the expected time until the epidemic becomes extinct. In our main contribution, we provide an exponentially large lower bound on the expected time to extinction, under certain assumptions on the network and the available curing resources.

Our general motivation comes from infectious disease epidemics, although without aiming at a faithful representation of the details of real-world situations. One example is the recent outbreak of the Ebola virus, which causes an acute and serious illness, which is often fatal if untreated (WHO [23]). However, supplies of experimental medicines, e.g., the prototype drug ZMapp, are limited and "will not be sufficient for several months to come," as stated in WHO [22]. In view of the limited availability of treatment for the virus, Pollack [18] addresses the following question: Ebola drug could save a few lives. But whose? Apart from the above, contagion processes are also relevant in the context of information and influence propagation in social networks (Adar and Adamic [1], Aral and Walker [2], Kim et al. [12], Gomez Rodriguez et al. [10]), viral marketing (Leskovec et al. [14]), spread of computer viruses (Garetto et al. [9]), or diffusion of innovations (Rogers [20]).

1.1. Preview of the Model

The wide relevance and applicability of contagion processes has led to extensive work on modeling their evolution and on understanding the resulting dynamics. Many models have been proposed in the literature; see, e.g., Liggett [15] for an in-depth review of such models and main results. Our work involves an extension of the canonical susceptible-infected-susceptible (SIS) epidemic model: the epidemic spreads on the underlying network from an initial set of *infected* nodes to *healthy* nodes, and at the same time, infected nodes can be cured. Healthy nodes get infected at a constant and common *infection rate* by each of their infected neighbors. In contrast to the standard SIS model, which assumes a common curing rate for all infected nodes at all times, we assume instead a node and time-specific *curing rate*. A *curing policy*, to be applied by a central controller, is a choice, at each time instant, of the curing rates at each node, taking into account the history of the epidemic and the network structure, subject to a budget on the sum of the curing rates applied at each time. The resulting

process is a controlled finite Markov chain with a unique absorbing state: the state where all nodes are healthy. We say that the epidemic becomes *extinct* when that absorbing state is reached. Under mild assumptions on the curing budget and given any set of initially infected nodes, the epidemic becomes extinct in a random but finite amount of time. The main question here is: How much time will be needed?

We can draw an important qualitative distinction between networks in which (i) the spread of the epidemic is hard to stop with the given curing budget, so that the expected time to extinction grows exponentially with the number of nodes and (ii) networks for which the curing resources are adequate, so that the expected time to extinction grows slowly (polynomially or even subpolynomially) with the number of nodes. Our general objective is to develop criteria that allow us to distinguish between cases (i) (slow extinction) and (ii) (fast extinction). In this paper, we focus on graphs in which the maximum degree is bounded by some Δ (independent of n) and provide the answer for a particular "regime," namely, the case of graphs whose CutWidth grows linearly.

1.2. Main Contribution
In a companion paper Drakopoulos et al. [7], we have established that a certain combinatorial quantity, the CutWidth of the underlying graph, denoted by W, plays a central role. In particular, if the curing budget r satisfies $r \geq 4W$ and $r \geq 16\Delta \log_2 n$, where n is the number of nodes, then the expected time to extinction is small—in fact, bounded above by $26n/r$, and therefore sublinear in n. In the present paper, we establish a converse result, for graphs with large CutWidth, namely, for graphs whose CutWidth is bounded below by $c_\gamma n$, for some constant $c_\gamma > 0$. In particular, we show that if $r \leq c_r W$, where $c_r > 0$ is an absolute constant (depending only on the degree bound and on c_γ), then, for some initial states, the expected time to extinction is at least exponential, under any curing policy. In other words, for graphs whose CutWidth scales linearly with n, a curing budget that also scales linearly with n is necessary (from the results in this paper) and sufficient (from the results in Drakopoulos et al. [7]) for fast extinction. In an equivalent interpretation of our main result, we are establishing that, for the case of bounded degree graphs, fast extinction with a sublinear curing budget is possible if and only if the CutWidth grows sublinearly.

1.3. Related Literature
A similar problem, but in which the curing rate allocation is static (open loop) has been studied in Cohen et al. [6], Gourdin et al. [11], Chung et al. [5], Preciado et al. [19], but the proposed methods were either heuristic or based on mean-field approximations of the evolution process; see Nowzari et al. [16] for a survey. Closer to our work, the authors of Borgs et al. [4] let the curing rates be proportional to the degree of each node and independent of the current state of the network, which may actually result in having curing resources wasted on healthy nodes. For bounded degree graphs, the policy in Borgs et al. [4] achieves sublinear expected time to extinction, but requires a curing budget that is proportional to the number of nodes. In contrast, the dynamic policy in Drakopoulos et al. [7] achieves the same performance (sublinear expected time to extinction) for bounded degree graphs with small CutWidth, *more economically*, by properly allocating a sublinear curing budget, hence demonstrating the increased effectiveness of dynamic policies.

Regarding lower bounds for dynamic policies, Borgs et al. [4] establishes that for expander graphs, and with a sublinear curing budget, the expected time to extinction is at least exponential in the number of nodes, under any curing policy. Expander graphs have automatically a large CutWidth, and so our result is in the same flavor, but much more general and also much harder to establish. The argument behind the result in Borgs et al. [4] is essentially the following: for expander graphs, any sufficiently large set of infected nodes results in a large (linear) total infection rate, which cannot be countered by a sublinear curing budget. But more general graphs with a large CutWidth need not have such a property: it may be the case that the instantaneous total infection rate is large for only "a few" configurations (sets of infected nodes). For such graphs, one can still establish that the total infection rate will be larger than the curing rate at some point in time, but this may last, in principle, for only a small time interval, and this is not enough to establish a negative result, in the form of a strong lower bound. We finally mention Wagner and Anantharam [21], who also deal with dynamic policies, but for the special case of a line graph.

To establish our result, we have to argue that a large total infection rate (larger than the curing budget) will be encountered for a sufficiently longtime interval, and that this creates a barrier to the fast extinction of the epidemic. The argument involves an elaborate combinatorial analysis of the evolution of the set of infected nodes.

We finally note a related negative result (exponential expected time to extinction) that we have established in Drakopoulos et al. [8]. That result deals with a special case, namely, graphs for which the CutWidth is close to

the largest possible value for graphs of the given size. The result in Drakopoulos et al. [8] admits, however, a much simpler proof, because under the large CutWidth assumption, it is easier to identify a barrier (a situation where the instantaneous total infection rate is high) inside which the process must remain for a sufficiently long time.

1.4. Outline of the Paper

The rest of the paper is organized as follows. In Section 2, we present the epidemic propagation model and define the curing policies under consideration. In Section 3, we define the CutWidth, as well as a generalization of that concept, and develop some combinatorial preliminaries that will be needed later. Section 4 contains a statement of the main result, two key lemmas that comprise the core of the proof, and some discussion. Sections 5, 6, and 7 contain the proofs of the three key lemmas. Finally, Section 8 contains some concluding remarks.

2. The Model

The model that we use and the contents of this section are borrowed from Drakopoulos et al. [7, 8]. We consider a network, represented by an undirected graph $G = (V, E)$, where V denotes the set of nodes and E denotes the set of edges. We use n to denote the number of nodes. Two nodes $u, v \in V$ are *neighbors* if $(u, v) \in E$. We restrict to graphs for which the node degrees are bounded above by Δ, which we take to be a given constant throughout the paper.

We let $I_0 \subseteq V$ be a set of intially infected nodes, and assume that the infection spreads according to a controlled contact (or SIS) process, where the rate at which infected nodes get cured is determined by a network controller. Specifically, each node can be in one of two states: *infected* or *healthy*. The controlled contact process is a right continuous, continuous-time Markov process $\{I_t\}_{t \geq 0}$ on the state space $\{0, 1\}^V$, where I_t stands for the set of infected nodes at time t. We refer to I_t as the *infection process*. We will sometimes use I_{t_-} as a shorthand for the value $\lim_{s \uparrow t} I_s$ just before time t.

At any point in time, state transitions at each node occur independently, according to the following rates. (These rates essentially define the generator matrix of the continuous-time Markov process under consideration.)

(a) The process is initialized at the given initial state I_0.

(b) If a node v is healthy, i.e., if $v \notin I_t$, the transition rate associated with a change of the state of that node to being infected is equal to a positive infection rate β times the number of infected neighbors of v, that is,

$$\beta \cdot \big|\{(u, v) \in E : u \in I_t\}\big|,$$

where we use $|\cdot|$ to denote the cardinality of a set. Any transition of this type will be referred to as an *infection*. By rescaling time, we can and will assume throughout the paper that $\beta = 1$.

(c) If a node v is infected, i.e., if $v \in I_t$, the transition rate associated with a change of the state of that node to being healthy is equal to a curing rate $\rho_v(t)$ that is determined by the network controller, as a function of the current and past states of the process. We are assuming here that the network controller has access to the entire history of the process. Any transition of this type will be referred to as a *recovery*.

We impose a *budget constraint* of the form

$$\sum_{v \in V} \rho_v(t) \leq r \tag{1}$$

for each time instant t, reflecting the fact that curing is costly. A *curing policy* is a mapping, which at any time t maps the past history of the process to a curing vector $\rho(t) = \{\rho_v(t)\}_{v \in V}$ that satisfies (1).

We define the *time to extinction* as the first time when the process first reaches the absorbing state where all nodes are healthy:

$$\tau = \min\{t \geq 0 : I_t = \varnothing\}.$$

In this paper, we focus on the *expected time to extinction* (the expected value of τ) as the performance measure of interest.

Without loss of generality, we can and will restrict to curing policies that allocate the entire budget r to infected nodes, as long as such nodes exist; this is because having unused curing resources or allocating them to healthy nodes would be wasteful. Under this restriction, the empty set (all nodes being healthy) is a unique absorbing state, and therefore the time to extinction is finite with probability 1.

Finally, we can and will restrict to policies that at any point in time allocate the entire budget to a single infected node, if one exists. We can do this because it is not hard to show that there exist optimal policies (i.e., policies that minimize the expected time to extinction) with this property.[1]

3. Graph-Theoretic Preliminaries

In this section, after some elementary definitions and notation, we focus on a deterministic version of the problem under consideration. Variants of such deterministic problems have been studied in the literature LaPaugh [13], Parsons [17], and involve the concept of the *CutWidth* of a graph. Loosely speaking, the CutWidth is the maximum cut encountered during the *deterministic* extinction of an epidemic on a graph, starting from all nodes infected, in the absence of any reinfections of nodes that have become healthy, and under the best-possible sequence with which nodes are cured. (A formal definition will be given shortly.)

We also introduce and study a natural extension of the concept of the CutWidth, for the case where only a subset of the nodes is initially infected; we refer to it as the *resistance* of the subset. The resistance turns out to contain important information about the evolution of an epidemic, starting from the corresponding subset, and will serve as a low dimensional summary of the state of an infection process. In the subsections that follow, we introduce those two concepts and study the properties of the latter.

3.1. Notation and Terminology

For convenience, we use the term *bag* to refer to a "subset of V." For any bags A, B, and any node v, we define

$$A \backslash B = \{v \in A : v \notin B\},$$

which is the set of nodes that belong in A but not in B, and

$$A \triangle B = (A \backslash B) \cup (B \backslash A),$$

which is the set of nodes at which A and B differ. Finally, we write

$$A + v = A \cup \{v\}, \qquad A - v = A \backslash \{v\}.$$

We next define the concept of a crusade from A to B as a sequence of bags that starts at A and ends at B, with the restriction that at each step of this sequence, arbitrarily many nodes may be added to the previous bag, but at most one can be removed. The formal definition follows.

Definition 1. For any two bags A and B, an $(A$–$B)$-*crusade* ω is a sequence $(\omega_0, \omega_1, \ldots, \omega_k)$ of bags, of length $k + 1$, with the following properties:
  (i) $\omega_0 = A$,
  (ii) $\omega_k = B$, and
  (iii) $|\omega_i \backslash \omega_{i+1}| \le 1$ for $i = 0, \ldots, k - 1$.
We use the notation $\Omega(A)$ to refer to the set of all $(A$–$\varnothing)$-crusades, i.e., crusades that start with a bag A and eventually end up with the empty set.

Property (iii) states that at each step of a crusade, arbitrarily many nodes can be added to, but *at most one* node can be removed from the current bag. Note that the definition of a crusade allows for *nonmonotone* changes, since a bag at any step can be a subset, a superset, or not comparable to the preceding bag. For this reason, crusades, as defined here, are different from the *monotone* crusades that were introduced in Drakopoulos et al. [7].

3.2. Cuts, Cutwidth, and Resistance

The number of edges connecting a bag A with its complement will be called the cut of the bag. Its importance lies in that it is equal to the total rate at which new infections occur, when the set of currently infected nodes is A.

Definition 2. For any bag A, its *cut*, $c(A)$, is defined as the cardinality of the set of edges

$$\{(u, v) : u \in A, v \in A^c\}.$$

In Lemma 1, we record, without proof, some elementary properties of cuts.

Lemma 1. *For any two bags A and B, we have*
  (i) $|c(A) - c(B)| \le \Delta \cdot |A \triangle B|$.
  (ii) *If $A \subseteq B$ and $v \in A$, then*
$$c(A - v) - c(A) \le c(B - v) - c(B).$$

Note that Lemma 1(ii) states the well-known submodularity property of the function $c(\cdot)$, and thus of the infection rate.

We now define the width of a crusade as the maximum cut that it encounters.

Definition 3. The *width* $z(\omega)$ of an $(A$–$B)$-crusade $\omega = (\omega_0, \ldots, \omega_k)$ is defined by

$$z(\omega) = \max_{1 \le i \le k}\{c(\omega_i)\}.$$

Note that in the above definition, the maximization starts at the first step of the crusade, i.e., we exclude ω_0 from consideration. The reason is the important *monotonicity* property in Lemma 2(i), in the next subsection, which would otherwise fail to hold.

We finally define the *resistance*[2] of a bag A as the minimum crusade width, over all $(A$–$\varnothing)$-crusades. Intuitively, this is the maximum cut encountered after the first step, during a crusade that "cures" all nodes in A in an "optimal" manner.

Definition 4. The *resistance* $\gamma(A)$ of a bag A is defined by

$$\gamma(A) = \min_{\omega \in \Omega(A)} z(\omega).$$

For the special case where A is the set V of all nodes, the corresponding resistance $\gamma(V)$ is called the *CutWidth* of the graph and is denoted by W.

We remark that the more common definition of the CutWidth, and which was the one used in Drakopoulos et al. [7], takes the minimum over monotone crusades, i.e., over crusades that only remove nodes. Nevertheless, the two definitions are equivalent, as has been established in Bienstock and Seymour [3], LaPaugh [13].

We close this section by observing that the resistance of a bag A satisfies the Bellman equation

$$\gamma(A) = \min_{|A \setminus B| \le 1} \{\max\{c(B), \gamma(B)\}\}. \tag{2}$$

3.3. Properties of the Resistance

This section develops some properties of the resistance. Lemma 2(i) states that if A and B are two bags with $A \subseteq B$, then $\gamma(A) \le \gamma(B)$. Intuitively, this is because one can construct a crusade from A to $\varnothing$ as follows. The crusade starts from A, then continues to the first bag encountered by a B-optimal crusade ω^B, and then follows ω^B. The constructed crusade and ω^B are the same except for the respective initial bags. By the definition of the resistance, the initial bag does not affect the maximization, and thus the width of the new crusade is equal to $\gamma(B)$. An optimal crusade from A can do no worse.

Lemma 2(ii) states that if two bags A and B differ by only m nodes, then the corresponding resistances are at most $m\Delta$ apart. Intuitively, this is because if $m = 1$ and $A \triangle B = \{v\}$, one can attach node v to the optimal crusade for the smaller of the two bags, thus obtaining a crusade that starts at the larger bag and encounters a maximum cut, which is at most Δ different from the original. The result for general m is obtained by moving from A to B by adding or removing one node at a time.

The formal proof of Lemma 2 follows the above outlined intuitive argument, and is given in the appendix so as not to disrupt continuity.

Lemma 2. *Let A and B be two bags.*
(i) *(Monotonicity). If $A \subseteq B$, then $\gamma(A) \le \gamma(B)$.*
(ii) *(Smoothness). We have that $|\gamma(A) - \gamma(B)| \le \Delta \cdot |A \triangle B|$.*

An immediate corollary of Lemma 2(i) is that for any bag A, we have $\gamma(A) \le W$.

Example. Consider a line graph with n nodes. Its CutWidth is easily seen to be equal to 1: if all nodes are initially infected, we can cure them one at a time, starting from the left; the cuts encountered along the way are all equal to 1. On the other hand, if all even nodes are initially infected, the corresponding cut is large, equal to $n - 1$. However, the cut being large does not acurately convey the difficulty of curing those nodes. For example, we might artificially infect the healthy nodes (or, in the stochastic SIS model, simply wait until they all get infected—this would happen in time which is sublinear in n), and then follow a curing policy for the case of a fully infected initial graph. Thus the expected time to extinction will be comparable to the one for the case where all nodes are initially infected. Note that the resistance of any nonempty bag is equal to 1: an optimal nonmonotone crusade can start by infecting all nodes (except possibly one of the end nodes, if n is even), and then curing the nodes one at a time. Thus the resistance, rather than the cut, is a better reflection of the difficulty of curing a given initial set of infected nodes. As a sidenote, these considerations suggest that some additional infections in the beginning (as allowed in our definition of the resistance) can be beneficial, and this is one of the reasons why our line of argument is based on nonmonotone (as opposed to monotone) crusades.

3.4. Relating Cuts to the Resistance

As illustrated in the preceding example, a large value of $c(I_0)$ does not mean that the infection is hard to extinguish; the resistance is more relevant. On the other hand, the proof of any negative result (i.e., a lower bound on the expected time to extinction) has to argue that at certain times the cut will be large and will present a barrier to the extinction of the epidemic. For this reason, we need a way of pinpointing certain times, at which a large resistance implies a large cut. This is accomplished by the next lemma, which establishes a connection between cuts and resistances at those times that the resistance is reduced. It shows that whenever the resistance is high and is reduced, the total infection rate is also high. This observation will play a central role in the proof of our main result.

Lemma 3. *Let A be a bag and suppose that $\gamma(A - v) < \gamma(A)$ for some $v \in A$. Then,*

$$c(A - v) \geq \gamma(A).$$

Proof. Let $B = A - v$. Since $|A \backslash B| = 1$, Equation (2) implies that

$$\gamma(A) \leq \max\{c(B), \gamma(B)\}. \tag{3}$$

Having assumed that $\gamma(B) < \gamma(A)$, Equation (3) implies that $\gamma(A) \leq c(B)$. □

4. The Main Result and the Core of its Proof

In this section, we state our main result and provide the key elements of its proof in the form of two lemmas. Loosely speaking, the result states that if the resistance of the initial bag scales linearly with the number n of nodes, and the budget scales only as a small constant multiple of n, then the expected time to extinction is exponentially large.

Theorem 1. *Consider a graph with n nodes and a set I_0 of initially infected nodes, and suppose that for some constant c_γ,*

$$\gamma(I_0) \geq c_\gamma n.$$

Suppose, furthermore, that all node degrees are bounded above by Δ. Then, there exist positive constants c_r and c, which only depend on c_γ and Δ such that if

$$r \leq c_r n,$$

Then,

$$\mathbb{P}(\tau \geq c e^{cn}) \geq \tfrac{1}{2},$$

under any policy, and for all large enough n. In particular,

$$\mathbb{E}[\tau] \geq \tfrac{1}{2} c e^{cn}.$$

Remark. An immediate corollary of Theorem 1 is obtained by letting $I_0 = V$, so that $\gamma(I_0)$ coincides with the CutWidth W: if the CutWidth scales linearly in n, and the curing budget is less than a certain multiple of the CutWidth, then the expected time to extinction grows exponentially in n. As a further corollary, if the curing budget grows sublinearly with n, fast extinction is possible only if the CutWidth grows sublinearly in n. This is a converse to the results of Drakopoulos et al. [7], which establish that if the CutWidth grows sublinearly in n, then fast extinction is possible with a sublinear budget.

The proof of our main result involves the following line of argument:

(a) In the first, deterministic, part of the proof (Lemma 4), we show that for graphs with large CutWidth, the time interval until the extinction of the epidemic must contain a substantially long subinterval during which the expected total infection rate is significantly larger than the budget, yet the realized ratio of infections to recoveries is relatively small, and, in particular, fairly different than the ratio of the corresponding expected rates.

(b) In the second, stochastic, part of the proof (Lemma 5), we argue that for a given time interval to have the properties in (a), a "large deviations" event with exponentially small (in n) probability, must occur. This is used to conclude that, with significant probability, it will take an exponentially long amount of time until an interval with the properties in (a) emerges.

Proof of Theorem 1. We start the proof by fixing a graph with n nodes, and the initial set I_0 of infected nodes. For convenience, from now on, we will use the shorthand notation γ instead of $\gamma(I_0)$. We assume that c_γ and Δ have been fixed, and that $\gamma \geq c_\gamma n$. Note that for sufficiently large n, γ will be much larger than Δ, so that we can use freely inequalities such as $\Delta < \gamma/4$ or $\gamma/(4\Delta) > 1$. To keep notation simple and avoid the use of ceilings and floors, we will also assume from now on that $\gamma/(4\Delta)$ is an integer. The proof for the general case is essentially the same.

The first part of the proof corresponds to the following lemma.

Lemma 4. *Consider a sample path for which $\tau < \infty$. For that sample path, there exist times t' and t'' with $0 \leq t' \leq t'' \leq \tau$ such that*:
  (i) $c(I_t) \geq \gamma/4$ for all $t \in [t', t'']$;
  (ii) we have $b = \gamma/(4\Delta) - 1$ recoveries during the interval $[t', t'']$;
  (iii) we have no more than $n + b$ infections during the interval $[t', t'']$.

The times t' and t'' in the preceding lemma are random variables (they depend on the sample path). However, they are not necessarily stopping times of the underlying stochastic process.

Note that it suffices to prove the existence of a time interval $[t', t'']$ with just properties (i) and (ii). This is because there are only n nodes in the graph. If we have b recoveries during a time interval, the number of infections cannot exceed $n + b$, and property (iii) follows automatically.

For the stochastic part of the proof, let us introduce some notation: for any $c > 0$, we define B_c to be the event that there exist times t', t'', with the properties in Lemma 4, together with the additional property $t'' \leq ce^{cn}$.

Lemma 5. *Having fixed c_γ and Δ, there exist small enough positive constants c_r and c such that if $r \leq c_r n$, then,*

$$\mathbb{P}(B_c) \leq \tfrac{1}{2}$$

for all large enough n.

Lemmas 4 and 5 immediately imply Theorem 1. To see this, Lemma 4 implies that t'' is well defined for any sample path. For any sample path that satisfies $\tau \leq ce^{cn}$, we must also have $t'' \leq ce^{cn}$. Thus the event $\{\tau \leq ce^{cn}\}$ is a subset of the event B_c. Using Lemma 5, we conclude that $\mathbb{P}(\tau \leq ce^{cn}) \leq \mathbb{P}(B_c) \leq 1/2$, as long as c_r and c are suitably chosen.

5. Proof of Lemma 4

Lemma 4 is the central—and least obvious—part of the proof. Before continuing with a formal argument, we provide a high-level informal overview, intended to enhance comprehension. The overall plan is to argue that $\gamma(I_t)$, whose initial value is γ, must eventually (at some time T) drop to $\gamma/2$, and that while the value $\gamma/2$ is approached, there must be a sufficiently long interval during which $c(I_t)$ is at least $\gamma/4$. Indeed, if $c(I_t) \geq \gamma/4$ for all times in $[0, T]$ (this is Case 1 below), the cut remains relatively large (and larger than the budget), which implies that the process is moving in a direction opposite to its drift; in particular, the probability of this happening is small.

Recall now that the cut is approximately equal to the resistance at those times that the resistance drops. Thus, $c(I_T)$ is approximately equal to $\gamma/2$. If $c(I_t)$ drops below $\gamma/4$ before time T (this is Case 2 below), there must exist an interval $[T', T]$ during which $c(I_t) \geq \gamma/4$, and during which the cut increases from $\gamma/4$ to $\gamma/2$. We want to argue that such an increase must be accompanied by a large number of recoveries (which will consist of a low probability event). The difficulty is that cut increases may be caused by either recoveries or infections. To isolate the effects of recoveries, we look at a "bottleneck process" Θ_t that starts I_t at time T', and which keeps track of the recoveries in I_t, while ignoring the infections. Similar to I_t, there will be a time at which the resistance of Θ_t will drop to $\gamma/2$ (this is because of $\Theta_t \subset I_t$, and monotonicity), and at that time, $c(\Theta_t)$ will be roughly equal to $\gamma/2$. Thus $c(\Theta_t)$ also increases from $\gamma/4$ to $\gamma/2$. However, because Θ_t only changes whenever the process I_t has a recovery, it follows that there must be $O(\gamma)$ recoveries in the process Θ_t and, therefore, for the process I_t as well (Lemma 6).

We can now start with the formal proof. Let us fix a particular sample path for which $\tau < \infty$. Let T be the first time that $\gamma(I_t)$ drops to a value of $\gamma/2$ or less:

$$T = \inf\{t \geq 0 : \gamma(I_t) \leq \gamma/2\}.$$

Given that $\gamma(I_\tau) = \gamma(\varnothing) = 0$, it is clear that such a time T exists and satisfies $0 \leq T \leq \tau$.

We distinguish between two cases:

Case 1. Suppose that throughout the interval $[0, T]$, we also have $c(I_t) \geq \gamma/4$. Because of the monotonicity property of $\gamma(\cdot)$ (Lemma 2(i)), $\gamma(I_t)$ decreases only when the set I_t decreases, that is, only when there is a recovery. Furthermore, using the *smoothness* property in Lemma 2(ii), each time that there is a recovery, $\gamma(I_t)$ can drop by at most Δ. Therefore the number of recoveries during the time interval $[0, T]$ is at least

$$\frac{\gamma(I_0) - \gamma(I_T)}{\Delta} \geq \frac{\gamma - \gamma/2}{\Delta} = \frac{\gamma}{2\Delta}.$$

We can then find some $\hat{T} \leq T$ such that during the time interval $[0, \hat{T}]$, we have exactly $\gamma/(4\Delta) - 1$ recoveries, and properties (i) and (ii) in the statement of Lemma 4 are satisfied by letting $t' = 0$ and $t'' = \hat{T}$.

Case 2. Suppose now that there exists some $t \in [0, T]$ with $c(I_t) < \gamma/4$, which is the more difficult case. Note that just before time T, we have $\gamma(I_{T-}) > \gamma/2$. Furthermore, $\gamma(I_T) \leq \gamma/2$. With our continuous-time Markov chain model, only one event (infection or recovery) can happen at any time. Since $\gamma(I_T) < \gamma(I_{T-})$, and since $\gamma(\cdot)$ is monotonic, it follows that we had a recovery, and therefore $I_T = I_{T-} - v$ for some node v. Lemma 3 applies with $A = I_{T-}$ and $A - v = I_T$, and we obtain

$$c(I_T) \geq \gamma(I_{T-}) > \frac{\gamma}{2}.$$

We now define

$$T' = \sup\left\{t \leq T: c(I_t) < \frac{\gamma}{4}\right\},$$

so that $c(I_{T'-}) < \gamma/4$. Furthermore, since $c(I_t)$ can change by at most Δ at each transition (Lemma 1), we must actually have

$$c(I_{T'}) < \frac{\gamma}{4} + \Delta, \tag{4}$$

which implies that $T' \neq T$ and $0 \leq T' < T$.

We will show that the interval $[t', t'']$ with $t' = T'$ and $t'' = T$ has properties (i) and (ii) in the statement of Lemma 4. Indeed, the definition of T' implies that

$$c(I_t) \geq \frac{\gamma}{4}, \quad \forall t \in [T', T],$$

which is property (i). The proof of Lemma 4 is completed by showing property (ii), namely, that the increase in $c(I_t)$, from a value smaller than $(\gamma/4) + \Delta$ (at time T'), to a value above $\gamma/2$ (at time T) together with a drop of the resistance from a value above $\gamma/2$ (at time T') to a value below $\gamma/2$ (at time T), must be accompanied by at least $\gamma/(4\Delta) - 1$ recoveries. This is the content of the next lemma.

Lemma 6. *The number of recoveries during the time interval $[T', T]$ is at least*

$$\frac{\gamma}{4\Delta} - 1.$$

Lemma 6 is a rather simple statement, but we are not aware of a simple proof or of a transparent intuitive explanation. Our proof relies on an auxiliary process, the *bottleneck process*, coupled with I_t, which is introduced and analyzed in the next section.

6. Proof of Lemma 6

The first step in proving Lemma 6 is the construction of a process, which is coupled with the infection process. Observe that a sample path of the infection process defines a crusade in which, at each step, a single node is added to or removed from the current bag. To any such crusade, we associate a *bottleneck sequence*, which is a sequence of bags consisting of subsets of the bags in the original crusade with several important properties. Consider a crusade $\omega = (A_0, A_1, \ldots, A_l)$ in which $|A_i \bigtriangleup A_{i-1}| = 1$ for $i = 1, \ldots, l$. In particular, we always have $A_i \subset A_{i-1}$ or $A_i \supset A_{i-1}$. We associate with ω a related sequence of bags $(\Theta_0, \ldots, \Theta_l)$ by letting

$$\Theta_i = \bigcap_{k=0}^{i} A_k, \quad i = 0, \ldots, l. \tag{5}$$

It is clear from our construction that Θ_i is always a subset of A_i, and that $\Theta_i \supseteq \Theta_{i-1}$. We have the following interpretation: Θ_0 starts the same as A_0. Whenever a node is removed from a bag in the original sequence, the same is done in the bottleneck sequence, as long as this is possible. On the other hand, whenever a node is added to a bag in the original sequence, nothing is done in the bottleneck sequence.

Lemma 7. *Consider a sequence $(A_0, A_1, \ldots, A_l)$ of bags such that $|A_i \,\triangle\, A_{i-1}| = 1$ for $i = 1, \ldots, l$, and the associated bottleneck sequence $(\Theta_0, \ldots, \Theta_l)$. The following hold*:

(i) $\Theta_i \subseteq A_i$,

(ii) *if* $c(\Theta_i) > c(\Theta_{i-1})$, *then* $A_i \subset A_{i-1}$, *and*

(iii) $c(\Theta_i) - c(\Theta_{i-1}) \le \Delta$.

Proof. (i) Follows directly from the definition.

(ii) Suppose that $c(\Theta_i) > c(\Theta_{i-1})$. Then, $\Theta_i \ne \Theta_{i-1}$. From the definition of the bottleneck sequence, we see that it if $A_i \supset A_{i-1}$, then $\Theta_i = \Theta_{i-1}$. Therefore we must have that $A_i \subset A_{i-1}$.

(iii) If $A_i \supset A_{i-1}$, then, $\Theta_i = \Theta_{i-1}$, and $c(\Theta_i) - c(\Theta_{i-1}) = 0$. On the other hand, if $A_i \subset A_{i-1}$, and using the assumption $|A_i \,\triangle\, A_{i-1}| = 1$, we write $A_i = A_{i-1} - v$ for some $v \in A_{i-1}$, and from Equation (5), we obtain $\Theta_i = \Theta_{i-1} - v$. The result then follows from Lemma 1(i). $\square$

We now complete the proof of Lemma 6. Let $A_0, \ldots, A_l$ be the sequence of bags that arise during the evolution of I_t, between times T' and T. In particular, $A_0 = I_{T'}$ and $A_l = I_T$. Let $\Theta_0, \ldots, \Theta_l$ be the corresponding bottleneck sequence, so that $\Theta_0 = A_0 = I_{T'}$. Using property (i) in Lemma 7, we have $\Theta_i \subseteq A_i$ for all i. Using the nonotonicity of $\gamma(\cdot)$, we obtain $\gamma(\Theta_i) \le \gamma(A_i)$ for all i. In particular,

$$\gamma(\Theta_l) \le \gamma(A_l) = \gamma(I_T) \le \frac{\gamma}{2} < \gamma(I_{T'}) = \gamma(\Theta_0).$$

(The second and third inequalities follow from the definition of T and the fact $T' < T$, respectively.) This implies that there exists some $i \in \{1, \ldots, l\}$ for which

$$\gamma(\Theta_i) \le \frac{\gamma}{2} < \gamma(\Theta_{i-1}).$$

We apply Lemma 3 and obtain that $c(\Theta_i) \ge \gamma(\Theta_{i-1}) > \gamma/2$. Thus the bottleneck sequence starts with $c(\Theta_0) = c(I_{T'}) < \gamma/4 + \Delta$ (cf. Equation (4)) and eventually its cut rises to a value above $\gamma/2$. From part (iii) of Lemma 7, $c(\Theta_i)$ can increase only when there is a recovery. From part (iii) of Lemma 7, $c(\Theta_i)$ can increase by at most Δ at each recovery. Thus, to obtain an increase from $\gamma/4 + \Delta$ to $\gamma/2$, we must have had at least $\gamma/(4\Delta) - 1$ recoveries in the process I_t between times T' and T.

A schematic summary of the two cases introduced in Section 5 is provided in Figure 1.

7. Proof of Lemma 5

Lemma 5 is a fairly routine "large deviations" result. It is useful to provide some intuition by considering the special case in which the times t' and t'' are fixed (not random), and $c(I_t) = \gamma/4$ throughout the interval $[t', t'']$ (as opposed to $c(I_t) \ge \gamma/4$). In this case, we have a Poisson process (recoveries) with rate r and an independent Poisson process (infections) with rate $\gamma/4$; their ratio is $4r/\gamma$. For properties (i) and (ii) in Lemma 4 to hold, the empirical ratio of observed recoveries to infections must be at least

$$\frac{b}{n+b} = \frac{\gamma/(4\Delta) - 1}{n + \gamma/(4\Delta) - 1},$$

where b is defined in Lemma 4. When r is small compared to $\gamma/4$, which is the case if we choose c_r small enough, we have an empirical ratio of recoveries to infections, which is above the theoretical ratio by a constant factor. Large deviations theory implies that this event has exponentially small probability. We then argue that within the time horizon of interest, $[0, ce^{cn}]$, there are only $O(ne^{cn})$ intervals that need to be considered. By choosing c small enough and using the union bound, the overall probability that there exist t' and t'' with the desired properties can be made small.

The proof for the general case runs along the same lines but involves a coupling argument to show that when $c(I_t)$ can exceed $\gamma/(4\Delta)$, then the event of interest (relatively few infections or, equivalently, too many recoveries) is even less likely to occur.

7.1. Decomposing the Event of Interest

Let c be a small enough constant—how small it needs to be will be seen at the end of the proof. Let $t^* = ce^{cn}$, which is the time horizon of interest in Theorem 1. Recall our definition of the event B_c in §4: event B_c occurs if and only if there exists a time interval $[t', t'']$ with $t'' \le ce^{cn} = t^*$, with exactly $b = \gamma/(4\Delta) - 1$ recoveries, with at most $n + b$ infections, and during which $c(I_t) \ge \gamma/4$.

Figure 1. *Case* 1. In the first case, $c(I_t)$ remains at least $\gamma/4$ throughout the interval $[0, T]$. Moreover, since the resistance drops from γ to $\gamma/2$, at least $\gamma/2$ recoveries must occur. *Case* 2. In the second case, $c(I_t)$ drops below $\gamma/4$. The last time that it does so (time T'), the resistance is above $\gamma/2$ and needs to drop to a value below $\gamma/2$. Therefore $c(I_t)$ needs to grow above (roughly) $\gamma/2$. In principle, this increase may happen through infections and not only through recoveries. This is why we define the auxiliary process Θ_i, whose cut also needs to increase to $\gamma/2$ but can only increase through recoveries, implying that at least (roughly) $\gamma/4\Delta$ recoveries occur.
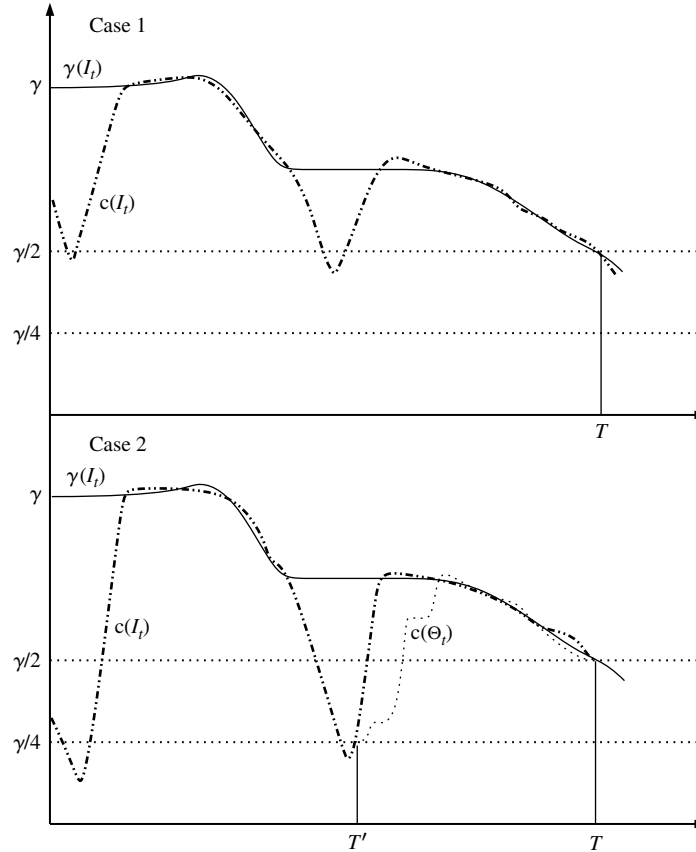
Our first step is to show that only a finite number of intervals $[t', t'']$ need to be considered. The recovery process behaves as a Poisson process with rate r, as long as the absorbing state has not been entered. To simplify the presentation, let us redefine the process, so that recoveries take place forever, according to a Poisson process. Any recovery that occurs after the extinction time τ is "dummy" and has no effect on the process $\{I_t\}$.

For $i \geq 1$, let t_i be the time of the ith recovery (actual or dummy). We consider the time interval $[t_i, t_{i+b-1}]$, which is the interval until $b-1$ new recoveries are observed, after the time t_i of the ith recovery.

For $i \geq 1$, we define B_i as the event that throughout the interval $[t_i, t_{i+b-1}]$, we have $c(I_t) \geq \gamma/4$ and at most $n + b$ infections.

Lemma 8. $B_c \subseteq \bigcup_{i=1}^{\infty} B_i$.

Proof. Consider a sample path that belongs to B_c, so that there exists an interval $[t', t'']$ with the properties in the definition of B_c. In particular, there exists some $i \geq 1$ such that the interval $[t', t'']$ contains the times $t_i, \ldots, t_{i+b-1}$, i.e.,

$$t' \leq t_i \leq t_{i+b-1} \leq t'';$$

furthermore, $c(I_t) \geq \gamma/4$ during that interval, and we have at most $n + b$ infections. But in that case, the interval $[t_i, t_{i+b-1}]$ has all of the properties that are required for event B_i to hold. $\square$

Let K be the total number of recoveries (real or dummy) during the time interval $[0, t^*]$. Using Lemma 8 and the union bound, we obtain

$$\mathbb{P}(B_c) \leq \sum_{i=1}^{4rt^*} \mathbb{P}(B_i) + \mathbb{P}(K > 4rt^*) \leq \sum_{i=1}^{4rt^*} \mathbb{P}(B_i) + \tfrac{1}{4}, \tag{6}$$

where the last inequality is obtained from the fact that K is a (Poisson) random variable with mean rt^*, and the Markov inequality.

It remains to bound the sum of the $\mathbb{P}(B_i)$. Since t^* grows exponentially with n, we are looking for an exponentially small upper bound on each B_i. This is the subject of the next subsection.

7.2. Bounding $\mathbb{P}(B_i)$

The main obstacle in characterizing $\mathbb{P}(B_i)$ is that the infection process has a time-varying rate. We will handle this issue through a coupling with a Poisson process that has a constant rate.

For $t \geq t_i$, let $M_i(t)$ be the number of infections during the interval $[t_i, t]$. Let also

$$C_i(t) = \{c(I_t) \geq \gamma/4, \, \forall \, t \in [t_i, t]\},$$

which is the event that $c(I_t)$ remains "large" during the interval $[t_i, t]$. Then, the event B_i can be expressed as

$$B_i = \{M_i(t_{i+b-1}) \leq n + b\} \cap C_i(t_{i+b-1}).$$

For the remainder of the proof, we assume that c_r is chosen (based only on c_γ and Δ, as in the statement of the theorem) so that

$$c_r < \frac{c_\gamma^2}{40\Delta}. \tag{7}$$

By rearranging terms, it is then seen that we can fix a constant $\bar{t}$ that again depends only on c_γ and Δ, which satisfies

$$c_r \bar{t} < \frac{c_\gamma}{5\Delta} \quad \text{and} \quad \frac{c_\gamma}{4}\bar{t} > 2. \tag{8}$$

For some interpretation and an outline of the rest of the argument, $\bar{t}$ is chosen so that, with high probability, the interval $[t_i, t_i + \bar{t}]$ has fewer than $b - 1$ recoveries, but more than $n + b$ infections if the cut remains "large." As will be seen, this property of $\bar{t}$ implies that, with high probability, the event B_i does not occur.

We define the event $\bar{B}_i$ by

$$\bar{B}_i = \{t_{i+b-1} < t_i + \bar{t}\} \cup \big(\{M_i(t_i + \bar{t}) \leq n + b\} \cap c(t_i + \bar{t})\big).$$

We will now show that $B_i \subseteq \bar{B}_i$. Consider a sample path in B_i. If that sample path also satisfies $t_{i+b-1} < t_i + \bar{t}$, then it is also an element of $\bar{B}_i$. Suppose now that the sample path satisfies $t_{i+b-1} \geq t_i + \bar{t}$. Using the monotonicity of the counting process $M_i(\cdot)$, we obtain $M_i(t_i + \bar{t}) \leq M_i(t_{i+b-1}) \leq n + b$, where the last inequality holds because the sample path belongs to B_i. Furthermore, since the sample path belongs to B_i, it must belong to $C_i(t_{i+b-1})$, which implies that it must also belong to $C_i(t_i + \bar{t})$. Thus the sample path belongs to $\{M_i(t_i + \bar{t}) \leq n + b\} \cap c(t_i + \bar{t})$, and is therefore an element of $\bar{B}_i$. This concludes the proof that $B_i \subseteq \bar{B}_i$. It then follows, using the union bound, that

$$\mathbb{P}(B_i) \leq \mathbb{P}(\bar{B}_i) \leq \mathbb{P}(t_{i+b-1} < t_i + \bar{t}) + \mathbb{P}\big(\{M_i(t_i + \bar{t}) \leq n + b\} \cap c(t_i + \bar{t})\big). \tag{9}$$

Our next step is to derive an upper bound for each of the two terms on the right-hand side of Equation (9), in terms of the Poisson distribution. For the first term, this is simple. The event $\{t_{i+b-1} < t_i + \bar{t}\}$ is the event that starting from time t_i, at least $b - 1$ recoveries occur within $\bar{t}$ time units. Since the recovery process is Poisson with rate r, we have

$$\mathbb{P}(t_{i+b-1} < t_i + \bar{t}) = \mathbb{P}(R > b - 1), \tag{10}$$

where R is a Poisson random variable with mean $r\bar{t}$.

To study the second term, we use $\mathbb{1}_C$ to denote the indicator function of the event $C_i(t_i + \bar{t})$. For those sample paths that belong to $C_i(t_i + \bar{t})$, and during the interval $[t_i, t_i + \bar{t}]$, the counting process $M_i(\cdot)$ maintains a rate that is larger than or equal to $\gamma/4$. Thus, on that time interval, $M_i(\cdot)$ can be coupled with a Poisson process $\bar{M}(\cdot)$ with rate equal to $\gamma/4$, in a way that guarantees that

$$M_i(t_i + \bar{t})\mathbb{1}_C \geq \bar{M}_i(t_i + \bar{t})\mathbb{1}_C,$$

for every sample path. Using this dominance relation, we obtain

$$\mathbb{P}\big(\{M_i(t_i + \bar{t}) \leq n + b\} \cap c(t_i + \bar{t})\big) = \mathbb{P}\big(\{M_i(t_i + \bar{t})\mathbb{1}_C \leq n + b\} \cap c(t_i + \bar{t})\big)$$
$$\leq \mathbb{P}\big(\{\bar{M}_i(t_i + \bar{t})\mathbb{1}_C \leq n + b\} \cap c(t_i + \bar{t})\big)$$

$$
\begin{aligned}
&= \mathbb{P}\big(\{\bar{M}_i(t_i + \bar{t}) \le n + b\} \cap \mathrm{c}(t_i + \bar{t})\big)\\
&\le \mathbb{P}\big(\bar{M}_i(t_i + \bar{t}) \le n + b\big)\\
&= \mathbb{P}(\bar{M} \le n + b),
\end{aligned}
\tag{11}
$$

where $\bar{M}$ is a Poisson random variable with mean $\gamma \bar{t}/4$.

We are now ready to apply large deviations results for Poisson random variables. Note that a Poisson random variable with mean λn can be viewed as a sum of n independent Poisson random variables with mean λ, and therefore, by the Chernoff bound, the probability of deviating from the mean by a constant factor falls exponentially with n. We record this fact in the lemma that follows, which just asserts the fact that we have a positive large deviations exponent.

Lemma 9. *There exists a function $\epsilon(\lambda, \lambda')$, defined for positive λ and λ', and which is positive whenever $\lambda \neq \lambda'$, with the following properties.*

(i) Let X be a Poisson random variable with mean bounded above by λn. If $\lambda' > \lambda$, then

$$
\mathbb{P}(X \ge \lambda' n) \le e^{-\epsilon(\lambda, \lambda')n}, \quad \forall n.
$$

(ii) Let X be a Poisson random variable with mean bounded below by λn. If $\lambda' < \lambda$, then

$$
\mathbb{P}(X \le \lambda' n) \le e^{-\epsilon(\lambda, \lambda')n}, \quad \forall n.
$$

The random variable R in Equation (10) is Poisson with mean $r\bar{t} \le c_r \bar{t} n$. Note that, for large enough n, we have $b - 1 = \gamma/(4\Delta) - 2 \ge \gamma/(5\Delta) \ge (c_\gamma/(5\Delta))n$, where the last inequality follows from the fact that $\gamma \ge c_\gamma n$. We apply Lemma 9(i), with $\lambda = c_r \bar{t}$ and $\lambda' = c_\gamma/(5\Delta)$:

$$
\mathbb{P}(R > b-1) \le \mathbb{P}(R > (c_\gamma/(5\Delta))n) \le e^{-\epsilon_1 n}.
$$

Because of our assumptions on c_r and $\bar{t}$ (cf. Equation (8)), we have $\lambda > \lambda'$, and ϵ_1 is a positive number determined by c_r, c_γ, and Δ.

Similarly, the random variable $\bar{M}$ in Equation (11) is Poisson with mean $\gamma \bar{t}/4 \ge (c_\gamma \bar{t}/4)n$. For any graph, γ is bounded above by $n\Delta$, and this implies that $b = \gamma/(4\Delta) - 1 \le n$. We apply Lemma 9(ii), with $\lambda = c_\gamma \bar{t}/4$ and $\lambda' = 2$:

$$
\mathbb{P}(\bar{M} \le n + b) \le \mathbb{P}(\bar{M} \le 2n) \le e^{-\epsilon_2 n}.
$$

Because of our assumptions on c_γ and $\bar{t}$ (cf. Equation (8)), we have $\lambda' < \lambda$, and ϵ_2 is a positive number determined by c_γ.

We have therefore established that each of the two terms on the right-hand side of Equation (9) is bounded above by an exponentially decaying term. By letting $\epsilon = \min\{\epsilon_1, \epsilon_2\} > 0$, we obtain that

$$
\mathbb{P}(B_i) \le 2e^{-\epsilon n}.
\tag{12}
$$

7.3. Completing the Proof of Lemma 5

For the given c_γ and Δ, we choose a suitably small c_r as in Equation (7). This allows us to set $\bar{t}$ as in Equation (8), leading to a positive ϵ in Equation (12). We then use Equation (12) to bound the terms $\mathbb{P}(B_i)$ in the inequality (6), and also make use of the facts that $t^* = ce^{cn}$ and $4rt^* \le 4c_r nce^{cn}$ to obtain

$$
\mathbb{P}(B_c) \le 4c_r nce^{cn}2e^{-\epsilon n} + \tfrac{1}{4} \le \tfrac{1}{2},
$$

provided that c is small enough (it just needs to be chosen a little smaller than ϵ) and n is large enough. This concludes the proof of Lemma 5.

8. Conclusions

We have considered the control of an epidemic (contagion process) given a limited curing budget, and provided an exponential lower bound on the expected time to extinction for bounded degree graphs. For the interesting (and least favorable) case where all nodes are initially infected, our assumption was that the CutWidth of the graph scales linearly with the number of nodes, and that the curing budget is bounded above by a small enough multiple of the number of nodes.

This result complements the results in Drakopoulos et al. [7], which show that when the ratio of the curing budget to the CutWidth is large enough, then the expected time to extinction is sublinear in the number of nodes. These results, taken together, show that for graphs with a large CutWidth, the ratio of the curing resources to the CutWidth is the key factor that distinguishes between slow and fast extinction. Our proof was based on a generalization of the CutWidth, the "resistance," which captures the difficulty of extinguishing an epidemic, starting from an arbitrary set of infected nodes.

It remains an open problem to develop lower bounds for more general bounded degree graphs, whose CutWidth scales sublinearly with the number of nodes. In some cases, this is easy. For example, for a square mesh with n nodes, the CutWidth is of order $O(\sqrt{n})$. Using the fact that any subset of the mesh with $\Theta(n)$ nodes has a cut of size $\Omega(\sqrt{n})$, one can show that a curing budget that scales at least as fast as $\sqrt{n}$ is necessary for fast extinction. The same argument applies whenever we deal with families of graphs that satisfy suitable isoperimetric inequalities. We conjecture that a similar result is always true: that is, unless the curing budget scales in proportion with the CutWidth, the expected time to extinction will be exponential. However, some new tools may have to be developed.

The proof of Theorem 1 and, in particular, Equation (7), shows that the exponential lower bound holds when c_r is smaller than a constant multiple of c_γ^2. We conjecture that a similar lower bound can be established under the assumption that c_r is smaller than a constant multiple of c_γ. If this is true, the deciding factor will be the ratio between the resistance and the recovery rate in a very concrete sense. However, the proof of this conjecture, if true, will require a much more refined argument.

Finally, the problem of controlling contagion processes on networks gives rise to a broader family of interesting research directions, such as control under partial information on the state of each node, combining inference and control, etc.

Appendix. Proof of Lemma 2

Recall that $\Omega(A)$ stands for the set of all $(A–\varnothing)$-crusades. Let also Ω^A be the set of all such crusades that achieve the minimum in the definition of the resistance, i.e.,

$$\Omega^A = \{\omega \in \Omega(A): z(\omega) = \gamma(A)\}.$$

(i) Suppose that $A \subseteq B$. Let $\omega^B = (\omega_0^B, \dots, \omega_k^B) \in \Omega^B$. Consider the sequence $\hat{\omega} = (\hat{\omega}_0, \dots, \hat{\omega}_k)$ of bags with $\hat{\omega}_0 = A$, and $\hat{\omega}_i = \omega_i^B$ for $i = 1, \dots, k$. We claim that $\hat{\omega}$ is a crusade $\hat{\omega} \in \Omega(A)$. Indeed,

    (a) $\hat{\omega}_0 = A$;
    (b) $\hat{\omega}_k = \omega_k^B = \varnothing$;
    (c) $|\hat{\omega}_0 \backslash \hat{\omega}_1| = |A \backslash \hat{\omega}_1| \leq |B \backslash \omega_1^B| = |\omega_0^B \backslash \omega_1^B| \leq 1$, where the first inequality follows from $A \subseteq B$ and $\hat{\omega}_1 = \omega_1^B$. Moreover, for $i = 0, \dots, k-1$, we have $|\hat{\omega}_i \backslash \hat{\omega}_{i+1}| = |\omega_i^B \backslash \omega_{i+1}^B| \leq 1$. Clearly,

$$z(\hat{\omega}) = \max_{1 \leq i \leq k}\{c(\hat{\omega}_i)\} = \max_{1 \leq i \leq k}\{c(\omega_i^B)\} = \gamma(B).$$

Using the definition of $\gamma(A)$, and the fact that $\hat{\omega} \in \Omega(A)$, we conclude that

$$\gamma(A) = \min_{\omega \in \Omega(A)} z(\omega) \leq z(\hat{\omega}) = \gamma(B).$$

(ii) If $|A \triangle B| = m$, we can go from bag A to bag B in a sequence of m steps, where at each step, we add or remove a single node. It thus suffices to show that each one of these steps can change the resistance by at most Δ. Accordingly, we only need to consider the case where $B = A + v$ for some $v \notin A$.

Let $\omega^A = (\omega_0^A, \dots, \omega_k^A) \in \Omega^A$. Consider the sequence $\hat{\omega} = (\hat{\omega}_0, \dots, \hat{\omega}_{k+1})$ of bags with $\hat{\omega}_i = \omega_i^A + v$ for $i = 0, \dots, k$, and $\hat{\omega}_{k+1} = \varnothing$. Clearly, $\hat{\omega}$ is a crusade in $\Omega(B)$, and therefore

$$\gamma(B) \leq z(\hat{\omega}) = \max_{1 \leq i \leq k}\{c(\omega_i^A + v)\} \leq \max_{1 \leq i \leq k}\{c(\omega_i^A)\} + \Delta = \gamma(A) + \Delta,$$

where the second inequality follows because the addition of one node can change the cut by at most Δ (Lemma 1(i)).

Endnotes

[1] A formal proof of this statement (which we only outline) goes as follows. We write down the Bellman equation for the problem of minimizing the expected time to extinction and observe that the right-hand side of Bellman's equation is linear in $\rho(t)$. We then recall that $\rho(t)$ is constrained to lie in a certain simplex, and conclude that we can restrict, without loss of optimality, to the vertices of that simplex. Any such vertex corresponds to allocating the entire budget to a single infected node.

[2] Note that in Drakopoulos et al. [7], we used a related notion, where the minimization took place with respect to monotone crusades.

References

[1] Adar E, Adamic LA (2005) Tracking information epidemics in blogspace. *Proc. 2005 IEEE/WIC/ACM Internat. Conf. Web Intelligence, WI '05* (IEEE Computer Society, Washington, DC).

[2] Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341.

[3] Bienstock D, Seymour P (1991) Monotonicity in graph searching. *J. Algorithms* 12(2):239–245.

[4] Borgs C, Chayes J, Ganesh A, Saberi A (2010) How to distribute antidote to control epidemics. *Random Structures and Algorithms* 37(2):204–222.

[5] Chung FRK, Horn P, Tsiatas A (2009) Distributing antidote using pagerank vectors. *Internet Math.* 6(2):237–254.

[6] Cohen R, Havlin S, Ben-Avraham D (2003) Efficient immunization strategies for computer networks and populations. *Physical Rev. Lett.* 91:247901.

[7] Drakopoulos K, Ozdaglar A, Tsitsiklis J (2014) An efficient curing policy for epidemics on graphs. *IEEE Trans. Network Sci. Engrg.* 1(2):67–75.

[8] Drakopoulos K, Ozdaglar A, Tsitsiklis JN (2015) A lower bound on the performance of dynamic curing policies for epidemics on graphs. *Proc. 54th IEEE Conf. Decision and Control, CDC '05* (IEEE, Piscataway, NJ), 3560–3567.

[9] Garetto M, Gong W, Towsley D (2003) Modeling malware spreading dynamics. *Proc. 22nd Annual Joint Conf. IEEE Comput. Comm. Soc., INFOCOM '03* (IEEE, Piscataway, NJ), 1869–1879.

[10] Gomez Rodriguez M, Leskovec J, Krause A (2010) Inferring networks of diffusion and influence. *Proc. 16th ACM SIGKDD Internat. Conf. Knowledge Discovery and Data Mining, KDD '10* (ACM, New York), 1019–1028.

[11] Gourdin E, Omic J, Van Mieghem P (2011) Optimization of network protection against virus spread. *8th Internat. Workshop on the Design of Reliable Comm. Networks, DRCN '11* (IEEE, Piscataway, NJ), 86–93.

[12] Kim L, Abramson M, Drakopoulos K, Kolitz S, Ozdaglar A (2014) Estimating social network structure and propagation dynamics for an infectious disease. Kennedy WG, Agarwal N, Yang SJ, eds. *Social Computing, Behavioral-Cultural Modeling and Prediction*, Lecture Notes in Computer Science, Vol. 8393 (Springer International, Cham, Switzerland).

[13] LaPaugh AS (1993) Recontamination does not help to search a graph. *J. ACM* 40(2):224–245.

[14] Leskovec J, Adamic LA, Huberman BA (2007) The dynamics of viral marketing. *ACM Trans. Web* 1(1): Article 5.

[15] Liggett TM (1985) *Interacting Particle Systems* (Springer, Berlin).

[16] Nowzari C, Preciado VM, Pappas GJ (2016) Analysis and Control of Epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems* 36(1): 26–46.

[17] Parsons TD (1978) The search number of a connected graph. *Proc. Ninth Southeastern Conf. Combinatorics, Graph Theory, Comput.* (Utilitas Mathematica, Winnipeg, Canada), 549–554.

[18] Pollack A (2014) Ebola drug could save a few lives. But whose? *New York Times* (August 9) A1.

[19] Preciado VM, Zargham M, Enyioha C, Jadbabaie A, Pappas GJ (2013) Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. arXiv:1303.3984.

[20] Rogers E (2003) *Diffusion of Innovations*, 5th ed. (Simon and Schuster, New York).

[21] Wagner AB, Anantharam V (2005) Designing a contact process: The piecewise-homogeneous process on a finite set with applications. *Stochastic Processes Appl.* 115(1):117–153.

[22] World Health Organization (2014) Statement on the WHO consultation on potential Ebola therapies and vaccines. (September), Media center, Geneva, Switzerland, http://www.who.int/mediacentre/news/statements/2014/ebola-therapies-consultation/en/.

[23] World Health Organization (2016) Ebola virus disease. Fact Sheet N° 103 (September), Media center, Geneva, Switzerland, http://www.who.int/mediacentre/factsheets/fs103/en/.