The Cost of Uncertainty in Curing Epidemics

Jessica Hoffmann

The University of Texas at Austin
Austin, TX
hoffmann@cs.utexas.edu

Constantine Caramanis

The University of Texas at Austin
Austin, TX
constantine@utexas.edu

ACM Reference Format:

Jessica Hoffmann and Constantine Caramanis. 2018. The Cost of Uncertainty in Curing Epidemics. In *SIGMETRICS'18 Abstracts: ACM SIGMETRICS International Conference on Measurement & Modeling of Computer Systems Abstracts, June 18–22, 2018, Irvine, CA, USA*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3219617.3219622>

1 INTRODUCTION

Epidemic models are used across biological and social sciences, engineering, and computer science, and have had important impact in the study of the dynamics of human disease and computer viruses, but also trends rumors, viral videos, and most recently the spread of fake news on social networks. In this paper, we focus on epidemics propagating on a graph, as introduced by the seminal paper [5]. In particular, we consider so-called SI models (see below for a precise definition) where an infected node can only propagate the infection to its non-infected neighbor, as opposed to the fully mixed models considered in the early literature. This graph-based approach provides a more realistic model, in which the spread of the epidemic is determined by the connectivity of the graph, and accordingly some nodes may play a larger role than others in the spread of the infection.

At any point in time, the *state* of an SI-type epidemic on a graph is given by the list of nodes on the graph that are infected, and their relative topology (position) in the graph. Having a good estimate of the state is critical, as it determines the dynamics of the spread of the epidemic into the future. As a simple example, we can ask what the spreading rate is on an N -node line graph of an infection with $N/2$ infected nodes. If those nodes are contiguous, then it will take $O(N)$ time for the epidemic to spread to the entire graph. If every other node is infected, it will take $O(1)$ time.

The problem of curing an epidemic with a limited budget, *but with perfect observation* (i.e., perfect knowledge of the state at each point in time), has been recently considered in [1, 3]. Their budget, as we explain more precisely further below, is essentially a bound on the *curing effort* they can expend at a given time (as opposed to total curing effort over time). In this setting, the problem is to optimize the allocation of the curing budget across nodes at every point in time. They characterize the budget required for fast curing, as a function of a combinatorial property of the graph – its **CUTWIDTH** (we define this below).

The problem of curing an epidemic with a limited budget and partial observation of the state of the epidemic (i.e., which nodes are infected and which are not) introduces a fundamentally new element to the problem. Indeed, this interaction represents a fundamental tension: our estimate of the state of a node improves the longer we observe it, and so the longer we wait to cure a node, the less likely we are to waste precious curing resources on non-infected nodes. On the other hand, the longer an infected node remains untreated, the more the epidemic spreads. To the best of our knowledge, no work has successfully attacked the problem of curing an epidemic with a limited budget and partial observation of the state of the epidemic (i.e., which nodes are infected and which are not). Our work considers precisely this problem, and therefore, broadly speaking, is about the interaction of – specifically, simultaneous – learning and control.

By considering learning the state and controlling the epidemic simultaneously, we prove a lower bound that shows that partial information can have a dramatic impact on the resources (either time or budget) required to cure an infection: even with slightly imperfect/incomplete information, the time to cure a particular graph may increase exponentially, unless the budget is also significantly increased. Concretely, we show that if instead of receiving the state of each node at each point in time, we receive a slightly noisy (e.g., only 99% accurate) guess of the state, then there is no constant factor of the **CUTWIDTH** which is sufficient for *any algorithm* to cure the epidemic in linear (expected) time.

1.1 Related Work and Background

When we have perfect knowledge of the state of each node, the problem of curing graphs with a limited budget (which corresponds to the speed at which the nodes are cured) has been solved in [1, 2]. Their results show that there exists a threshold phenomenon: for any given graph, if the curing budget is lower than a combinatorial quantity of the graph called the **CUTWIDTH**, the curing time is exponential; if it is higher, they exhibit a strategy to cure any graph in sublinear time. The **CUTWIDTH** captures a key bottleneck in curing, and is important in our work as well. Therefore it is useful to define this precisely now.

Definition 1.1. Given a graph $G = G(V, E)$, and any subset of the nodes, $S \subseteq V$, the **CUT** of S is the number of edges crossing from S to S^c . Given any sequence of $|V| + 1$ subsets $S_0, \dots, S_{|V|}$ such that $S_0 = \emptyset$, $S_{|V|} = V$, and S_k and S_{k+1} differ by the addition of a single node (called a *crusade* in [1, 2]), the cut of the sequence is the largest cut of any of the sets S_k . The **CUTWIDTH** of a graph is the minimum cut of any sequence satisfying the above properties.

Their strategy is based on two main ideas. The nodes are cured following an ordering which keeps the cut between the infected set and the non-infected set as low as possible. Then, as soon as

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGMETRICS'18 Abstracts, June 18–22, 2018, Irvine, CA, USA

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5846-0/18/06.

<https://doi.org/10.1145/3219617.3219622>

there is a new infection, the strategy switches to damage control, and focuses on returning to the ordering previously mentioned.

Our result hinges on the fact that the damage-control part of the strategy is exactly the part which is hard to accomplish with partial information. If the number of k -hop neighbors of a node grows exponentially, as is the case for the binary tree, detecting where the infection can have spread becomes a difficult task. Moreover, if we can detect such an escape path, but the infection has spread to a high number of nodes by the time we have enough information to try to prevent it, detection was useless. It is the tension between waiting less time and wasting budget on false alerts, or waiting too long and being unable to prevent the spread, which makes the problem of curing with partial information challenging.

2 MODEL AND MAIN CONTRIBUTIONS

2.1 The SI + curing model

In a standard SI (susceptible $\rightarrow$ infected) model, an epidemic spreads along edges from infected nodes to their neighbors according to an exponential spreading model: when a node becomes infected, it infects each uninfected neighbor according to an exponential random variable. SIS models are SI models where infected nodes can transition to susceptible, again at an exponential rate. Here, we consider the setting where the rate at which nodes transition from infected to susceptible is under our control, subject to a budget. How to optimally use this budget is the main question at hand.

We consider a discrete, Bernoulli approximation to these exponential rate models, by considering the dynamics evolving with discrete time steps τ ; we then take the time step τ to zero, hence recovering the continuous time dynamics. In particular, this model is a discretization of the exponential model of [3].

The dynamics of this controlled stochastic process evolve as follows. At each time t , for all N nodes of the graph, the decision-maker assigns a budget r_i^t , subject to the constraints $\sum_{i=1}^N r_i^t = r$. During a time step of length τ , each node i is cured with probability $\delta_i^t = 1 - e^{-r_i^t \tau}$ if it was infected, and nothing happens otherwise – the budget is wasted. Then, for every edge between an infected and a susceptible node, an infection occurs with probability $\mu = 1 - e^{-\tau}$. The number of infected nodes at time t is given by I_t . In particular, since the graph is completely infected at the beginning, we have $I_0 = N$.

2.2 Partial Information/Blind Curing

In the Complete Information setting, we assume that the state (infected or susceptible) of each node is known at each point in time. In what we call the Blind Curing model, we never have any information about the status of each node. The Blind Curing model is a technical tool we use en route to the final result. We introduce a Partial Information model that interpolates between these two extremes, and indeed is our main object of interest. Our model of partial information provides a stark tradeoff for the decision-maker: allocate resources to nodes whose status is very uncertain, and thus significantly raise the probability of wasting curing resources, or wait to collect more information and hence more certainty about the status of a node, running the risk that an infected node was allowed to infect neighbors unfettered.

Thus, in the Complete Information model, an infected node would raise a flag at each instant with probability 1, and an uninfected node would never raise a flag. In the **Partial Information** model, at each time step, each node, independently of all others, raises a flag with some probability. The probability of getting a flag is p if the node is infected, q if the node is susceptible, with $p > q$. By aggregating the information about a node over multiple time steps, we can use basic concentration inequalities to deduce its state, and thus more observation time corresponds to higher certainty about a node's state.

As noted above, $p = 1, q = 0$ recovers the **Complete Information** setting, and $p = q$ the **Blind Curing** setting.

In order to recover the continuous time dynamics, we let $\tau \rightarrow 0$. The key quantity that measures the amount of information per fixed unit time, is given by the rate function from Sanov's theorem, normalized by the time step: $\frac{\mathcal{D}(p||q)}{\tau}$, where $\mathcal{D}(p||q)$ is the Kullback-Leibler distance between p and q . To understand this intuitively, this says that when $\frac{\mathcal{D}(p||q)}{\tau}$ is a constant, observing a node for a fixed period of time corresponds to administering a test with a nonzero false positive and false negative probability. That is, we can know the state of a node with constant probability of error by observing this node over a constant amount of time, which is what one expects from a real-world source of information. Note that as $\tau \rightarrow 0$, if $p - q$ is constant (or, more generally, if $\mathcal{D}(p||q)$ goes to zero sublinearly) then we recover the Complete Information setting. Hence, the setting of interest is where $(p - q) \rightarrow 0$ as $\tau \rightarrow 0$, and the critical scaling is controlled by $\mathcal{D}(p||q)/\tau$.

2.3 Main contributions

Our main result consists of two parts [4]. First, we show that there exist graphs that cannot be cured in polynomial time in the Blind Curing model. We then use this result to obtain a lower bound for the cost of lack of information in the Partial Information model. We obtain an expression for the lower bound that shows the required tradeoff between $\frac{\mathcal{D}(p||q)}{\tau}$ (the information available per unit of time), and the budget, r .

THEOREM 2.1. *A Partial Information impossibility result.*

We consider the task of curing a fully infected complete balanced binary tree with N nodes. Let $\frac{\mathcal{D}(p||q)}{\tau}$ be a measure of the amount of information we get per time step, and r be the budget (curing rate) of our curing process. If

$$\frac{\mathcal{D}(p||q)}{\tau} \leq O\left(\frac{\log(N)\sqrt{\log(r)}}{r}\right), \quad (1)$$

as $\tau \rightarrow 0$, then it is fundamentally impossible for any algorithm (of any computational complexity) to cure the complete binary tree in polynomial expected time with budget $r = O(W^\alpha)$, where W is the CUTWIDTH of the graph and α is any constant.

For the Blind Curing case, we also have the following upper bound.

THEOREM 2.2. *For all $c > 0$, we can always cure the binary tree in expected linear time with budget $O(e^{4/c} N^c)$. In particular, our strategy does not require any information about the state of the nodes.*

Interpreting the result. Suppose that if a node is observed for a fixed period of time, we can estimate its state (infected or not) with probability $1 - \delta$ for δ some constant. Fixing $\alpha = 1$ in the statement above, our result implies that regardless of what this constant is, e.g., even if we have a test that returns a result that is 99% accurate, then polynomial time curing is impossible, for budget any multiple of the CUTWIDTH. Indeed, as explained above, a constant-error estimate in a fixed unit of time corresponds to $D(p||q)/\tau$, the left-hand side of (1), being a constant. On the other hand, if the budget is any multiple of the CUTWIDTH, the right-hand side of (1) grows like $\sqrt{\log \log(N)}$, and in particular is larger than any constant. In contrast, with complete (and instantaneous) certainty of the state of each node (which would correspond to an infinite left-hand side of (1)), [1] proves that every graph can be cured in linear expected time with budget higher than the CUTWIDTH.

For the Blind Curing setting, Theorem 2.1 says that for budget of any polynomial of $\log(N)$, curing takes superpolynomial time. Theorem 2.2 gives an upper bound that shows that this lower bound is not too far off; it says that a budget of N^c is sufficient, for any $c > 0$.

Our result focuses on the binary tree. Since our main result is a *lower bound*, this specific example is sufficient to resolve the question of whether the CUTWIDTH (or something proportional to it) is the right quantity to focus on to build a curing strategy robust to noise in our node estimates. In addition to this, we note that many graphs contain trees as subgraphs. Since adding nodes and edges only makes curing more difficult, our results can be seen to apply to any graph structure with a binary tree as a subgraph (as long as adding edges does not dramatically change the CUTWIDTH of the graph).

3 PROOF SKETCH

We first prove that polynomial curing is impossible in the Blind Curing setting if the budget is polynomial in the CUTWIDTH. We then show that in the Partial Information setting, we do not obtain enough information to detect threats of reinfection, and thus cannot prevent them: we are "blind" to the threats until it is too late.

Our proof in the Blind Curing setting focuses on a subprocess which is bound to happen for any curing strategy. We consider the last $\frac{N}{r^4}$ infected nodes. We show that by the time we cure these last remaining infected nodes, a new set of $\frac{N}{r^4}$ nodes becomes infected with high probability. Trying to cure the whole graph is then similar to playing a very long game of whack-a-mole with superpolynomial expected end time.

3.1 Blind Curing setting

Step 1: We first show that if a strategy allows the cut between the infected and susceptible set to be much higher than the available budget r , the infection set grows. In this case, the infection rate exceeds the curing rate, and the reinfection would be inevitable even if we had complete knowledge about the infection state of each node at each time (*i.e.* this happens even in the Complete Information setting). In particular, if $\frac{N}{r^4}$ nodes are infected and the cut is above r^4 , the drift

of the curing process is dominated by the infections. We can then use random walk results, such as Wald's Inequality, to prove that after a few time steps, we end up with at least as many infected nodes, but a cut below r^3 (we actually end up with many more infected nodes, but as many is enough for the proof). We can therefore focus on analyzing the situation in which $\frac{N}{r^4}$ nodes remain infected with a cut lower than r^4 .

Step 2: Due to the topology of the binary tree, a cut below r^4 implies that there exists an infected node which is close to the root. This makes it easy for the infection to escape through the root, and reach a large number of susceptible nodes. One key point of the proof is that this node will remain infected (and therefore potentially infecting) for a very long time, and an infection can start at any time step during this period.

Step 3: Since the infection escapes through the root, the number of uninfected nodes easily accessible is very large, and specifically, larger than the budget. This makes it impossible to cover all the potential escape routes. Notice that this is very specific to the Blind Curing setting: if we knew in which direction the infection was escaping, we could prevent it as in [3]. It is because the number of potential infected nodes is exponentially higher than the number of actual nodes infected, and because we do not know where the infection actually is, that we end up wasting considerable curing budget on uninfected nodes. Therefore, the infection is very likely to escape, and a new set of $\frac{N}{r^4}$ nodes becomes infected again.

3.2 Partial Information setting

To extend this result to the Partial Information setting, we notice that as soon as the cut of the new infection reaches $3r$, we can use Gambler's Ruin results to show that at least $\frac{N}{r^4}$ nodes will become infected with constant probability. If we cannot detect the infection escaping until a cut of $3r$ is reached, we therefore cannot prevent the reinfection with constant probability. Using Sanov's Theorem, we show that the uncertainty in our state estimation for any node does not resolve itself quickly enough (in particular, with respect to how fast the neighborhoods of the binary tree grow). Specifically, the infection remains undetected with constant probability until a cut of $3r$ is attained. This allows us to extend the result from the Blind Curing Setting to the Partial Information setting.

REFERENCES

- [1] Kimon Drakopoulos, Asuman Ozdaglar, and John N. Tsitsiklis. 2014. An efficient curing policy for epidemics on graphs. *arXiv preprint arXiv:1407.2241* December (2014), 1–10. <https://doi.org/10.1109/TNSE.2015.2393291> arXiv:arXiv:1407.2241v1
- [2] Kimon Drakopoulos, Asuman Ozdaglar, and John N. Tsitsiklis. 2015. A lower bound on the performance of dynamic curing policies for epidemics on graphs. 978 (2015), 3560–3567. <https://doi.org/10.1109/CDC.2015.7402770> arXiv:1510.06055
- [3] Kimon Drakopoulos, Asuman Ozdaglar, and John N. Tsitsiklis. 2015. When is a network epidemic hard to eliminate? (2015), 1–17. arXiv:1510.06054 <http://arxiv.org/abs/1510.06054>
- [4] Jessica Hoffmann and Constantine Caramanis. 2018. The Cost of Uncertainty in Curing Epidemics. In *Proceedings of the ACM on Measurement and Analysis of Computing Systems*.
- [5] Mark E. J. Newman. 2002. Spread of epidemic disease on networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 66, 1 (2002). <https://doi.org/10.1103/PhysRevE.66.016128> arXiv:cond-mat/0205009