

Information Theoretic Adaptive Tracking of Epidemics in Complex Networks

Patrick L. Harrington Jr.
Bioinformatics Graduate Program
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
plhjr@umich.edu

Alfred O. Hero III.
Department of EECS
Department of Statistics
University of Michigan
Ann Arbor, MI 48109
hero@umich.edu

Abstract—Adaptively monitoring the states of nodes in a large complex network is of interest in domains such as national security, public health, and energy grid management. Here, we present an information theoretic adaptive tracking and sampling framework that recursively selects measurements using the feedback from performing inference on a dynamic Bayesian Network. We also present conditions for the existence of a network specific, observation dependent, phase transition in the updated posterior of hidden node states resulting from actively monitoring the network. Since traditional epidemic thresholds are derived using observation independent Markov chains, the threshold of the posterior should more accurately model the true phase transition of a network. The adaptive tracking framework and epidemic threshold should provide insight into modeling the dynamic response of the updated posterior to active intervention and control policies while monitoring modern complex networks.

I. INTRODUCTION

This paper treats the important problem of monitoring the states of nodes in large computer, social, or power networks where these states dynamically change due to viruses, rumors, or failures that propagate according to the graph topology [4], [7], [10]. This class of network dynamics has been extensively modeled as a percolation phenomenon, where nodes on a graph can randomly “infect” their neighbors.

Percolation across networks has a rich history in the field of statistical physics, computer science, and mathematical epidemiology [10], [15], [17]. Here, researchers are typically confronted with a network, or a distribution over the network topology, and extract fixed point attractors of node configurations, thresholds for phase transitions in node states, or distributions of node state configurations [2], [6], [11]. In the field of fault detection, the nodes or edges can “fail”, and the goal is to activate a subset of sensors in the network which yield high quality measurements that identify these failures [14], [18]. While the former field of research concerns itself with extracting *offline* statistics about properties of the percolation phenomenon on networks, devoid of any measurements, the latter field addresses *online* measurement selection tasks.

Here, we propose a methodology that actively tracks a causal Markov process across a complex network (such as the one in Figure 3(a)), represented as a dynamic Bayesian network, where measurements are adaptively selected using feedback from the updated posterior distribution. We establish conditions such that the updated posterior probability of all nodes “infected” is driven to one as the number of time samples goes to infinity. The proposed epidemic/percolation threshold on the *updated* posterior distribution over the hidden states is a function of structural properties of the network, epidemiological parameters, and sensor likelihoods corresponding to those nodes that were sampled.

The proposed percolation threshold should more accurately reflect the true conditions that cause a phase transition in a network, e.g., node status changing from healthy/normal to infected/failed, than traditional thresholds derived from conditions on predictive distributions which are devoid of any measurements. As the conditions of a threshold are extracted by inspecting the dominant mode of decay of the updated posterior, this permits specification of best and worst case convergence rates of that a network clears the infection. Additionally, the decay dynamics of the updated posterior can yield insight into the asymptotic detection performance of the system for a given false alarm rate.

Since most practical networks of interest are large, it is usually infeasible to sample all nodes continuously and exhaustively. Given sampling constraints, we present an information theoretic sampling strategy that selects specific nodes that will yield the largest information gain, and thus, better detection performance.

The proposed sampling strategy balances the trade-off between trusting the *predictions* from the assumed model dynamics and expending precious resources to select a set of nodes for measurement.

We present the adaptive measurement selection problem and give two tractable approximations to this subset selection problem based upon the joint and marginal posterior

distribution, respectively. A set of decomposable Bayesian filtering equations are presented for this adaptive sampling framework and the tractable inference algorithms for complex networks are discussed. We present analytical worst case performance bounds for our adaptive sampling performance, which can serve as sampling heuristics for the activation of sensors or trusting predictions generated from previous measurements.

We believe that this is the first attempt to extract conditions of percolation thresholds in actively monitored dynamic Bayesian networks where the updated posterior distribution is the sufficient statistic of interest rather than observation independent predictive distributions.

II. PROBLEM FORMULATION

The objective of actively monitoring the n node network is to recursively update the posterior distribution of each hidden node state given various measurements. Specifically, the next set of m measurement actions (nodes to sample), $m \ll n$, at next discrete time are chosen such that they yield the highest quality of *information* about the n hidden states. The condition on $m \ll n$ simulates the reality of fixed resource constraints, where typically only a small subset of nodes in a large network can be observed at any one time.

Here, the hidden states are discrete random variables that correspond to the states encoded by the percolation process on the graph. Here, the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, with \mathcal{V} representing the set of nodes and \mathcal{E} corresponding to the set of edges. Formally, we will assume a state-space representation of a discrete time, finite state, partially observed Markov decision process (POMDP). Here,

$$\mathbf{Z}_k = \{Z_k^1, \dots, Z_k^n\} \quad (1)$$

represents the joint hidden states, e.g., healthy or infected

$$\mathbf{Y}_k = \{\mathbf{Y}_k^{(1)}, \dots, \mathbf{Y}_k^{(m)}\} \quad (2)$$

represents the m observed measurements obtained at time k , e.g., biological assays or PINGing an IP address, and

$$\mathbf{a}_k = \{a_k^1, \dots, a_k^m\} \quad (3)$$

represents the m actions taken at time k , i.e., which nodes to sample. Here, $\mathbf{Y}_k^{(j)}$, continuous/categorical valued vector of measurements, which is induced by action a_k^j , $a_k^j \in \mathcal{A}$, with $\mathcal{A} = \{1, \dots, n\}$ confined to be the set of all n individuals in the graph, and $Z_k^i \in \{0, 1, \dots, r\}$. Since the topology of \mathcal{G} encodes the direction of "flow" for the process, the state equations may be modeled as a decomposable partially observed Markov process:

$$\mathbf{Y}_k^i = \mathbf{f}(Z_k^i) + \mathbf{w}_k^i \quad (4)$$

$$Z_k^i = h(\{Z_{k-1}^j\}_{j \in \eta(i) \cup i}). \quad (5)$$

Here, $\eta(i) = \{j : \mathcal{E}(\mathcal{V}_i, \mathcal{V}_j) \notin \emptyset\}$ is the neighborhood of i , $\mathbf{f}(Z_k^i)$ is a non-random vector-valued function, \mathbf{w}_k^i is measure-

ment noise, and $h(\{Z_{k-1}^j\}_{j \in \eta(i) \cup i})$ is a stochastic equation encoding the transition dynamics of the Markov process (see Figure 1 for a two node graphical model representation).

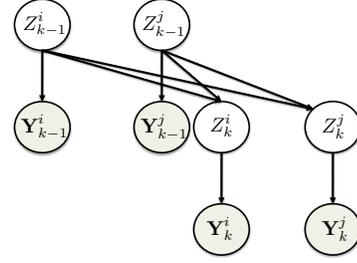


Fig. 1. Partially Observed Markov Structure for i and j for $\mathcal{E}(\mathcal{V}_i, \mathcal{V}_j) \notin \emptyset$

A. Bayesian Filtering

In our proposed framework for actively monitoring the hidden node states in the network, the posterior distribution is the sufficient statistic for inferring these states. The general recursion for updating the joint posterior probability given all past and present observations is given by the standard Bayes update formula:

$$p(\mathbf{Z}_k | \mathbf{Y}_{0:k}) = \frac{f(\mathbf{Y}_k | \mathbf{Z}_k)}{g(\mathbf{Y}_k | \mathbf{Y}_{0:k-1})} p(\mathbf{Z}_k | \mathbf{Y}_{0:k-1}) \quad (6)$$

with

$$p(\mathbf{Z}_k | \mathbf{Y}_{0:k-1}) = \sum_{\mathbf{z} \in \{0,1,\dots,r\}^n} p(\mathbf{Z}_k | \mathbf{Z}_{k-1} = \mathbf{z}) p(\mathbf{Z}_{k-1} = \mathbf{z} | \mathbf{Y}_{0:k-1}). \quad (7)$$

and

$$g(\mathbf{Y}_k | \mathbf{Y}_{0:k-1}) = \sum_{\mathbf{z} \in \{0,1,\dots,r\}^n} f(\mathbf{Y}_k | \mathbf{Z}_k = \mathbf{z}) p(\mathbf{Z}_k = \mathbf{z} | \mathbf{Y}_{0:k-1}). \quad (8)$$

The Chapman-Kolmogorov equations provide the connection between the posterior update (7) and the distribution resulting from the standard percolation equations. In the former, the updates are conditional probabilities that are conditional on past observations, while in the latter, the updates are not dependent on observations.

The local interactions in the graph \mathcal{G} imply the following conditional independence assumptions:

$$f(\mathbf{Y}_k | \mathbf{Z}_k) = \prod_{i=1}^n f(\mathbf{Y}_k^i | Z_k^i). \quad (9)$$

$$p(\mathbf{Z}_k | \mathbf{Z}_{k-1}) = \prod_{i=1}^n p(Z_k^i | \{Z_{k-1}^j\}_{j \in \eta(i) \cup i}) \quad (10)$$

where the likelihood term is defined in (4) and the transition dynamics are defined in (5). This decomposable structure allows the belief state (posterior excluding time k observations)

update, for the i^{th} node in \mathcal{G} , to be written as:

$$p(Z_k | \mathbf{Y}_{0:k-1}) = \sum_{\mathbf{z} \in \{0,1,\dots,r\}^{\|\mathbf{pa}\|}} p(Z_k | \mathbf{Z}_{k-1}^{\text{pa}} = \mathbf{z}) p(\mathbf{Z}_{k-1}^{\text{pa}} = \mathbf{z} | \mathbf{Y}_{0:k-1}) \quad (11)$$

with the parent set, $\text{pa} = \{\eta(i), i\}$. Unfortunately, for highly connected nodes in \mathcal{G} , this marginal update becomes intractable. It thus must be approximated [5], [12], [16].

B. Information Theoretic Adaptive Sampling

In most real world situations, acquiring measurements from all n nodes at any time k is unrealistic, and thus, a sampling policy must be exploited for measuring a subset of nodes [1], [3], [8], [14]. Since we are concerned with monitoring the states of the nodes in the network, an appropriate reward is the expected information gain between the *updated* posterior, $p_k = p(\mathbf{Z}_k | \{\mathbf{Y}_k^i\}_{i \in \mathbf{a}_k}, \mathbf{Y}_{0:k-1})$, and the belief state, $p_{k|k-1} = p(\mathbf{Z}_k | \mathbf{Y}_{0:k-1})$:

$$\mathbf{a}_k = \operatorname{argmax}_{\mathbf{a} \subset \mathcal{A}} \mathbb{E} [\mathcal{D}_\alpha (\{\mathbf{Y}_k^i\}_{i \in \mathbf{a}}) | \mathbf{Y}_{0:k-1}] \quad (12)$$

$$\mathcal{D}_\alpha (\{\mathbf{Y}_k^i\}_{i \in \mathbf{a}}) = \mathcal{D}_\alpha (p_k || p_{k|k-1}), \quad 0 < \alpha < 1 \quad (13)$$

with α -Divergence defined by

$$\mathcal{D}_\alpha (p || q) = \frac{1}{\alpha - 1} \log (\mathbb{E}_q [(p/q)^\alpha]) \quad (14)$$

for any distributions p and q with identical support.

The reward in (12) has been widely applied to multi-target, multi-sensor tracking for many problems including, sensor management and surveillance [8], [13]. Note that $\lim_{\alpha \rightarrow 1} \mathcal{D}_\alpha (p || q) \rightarrow \mathcal{D}_{KL} (p || q)$, where $\mathcal{D}_{KL} (p || q)$ is the Kullback-Leibler divergence between p and q . The expectation in (12) is taken with respect to the conditional distribution $g(\mathbf{Y}_k | \mathbf{Y}_{0:k-1})$ given the previous measurements $\mathbf{Y}_{0:k-1}$ and actions \mathbf{a}_k . In practice, the expected information divergence in (12) must be evaluated via Monte-Carlo methods. Also, the maximization in (12) requires enumeration over all $\binom{n}{m}$ actions (for subsets of size m), and therefore, we must resort to approximations. We propose incrementally constructing the set of actions at time k , \mathbf{a}_k , for $j = 1, \dots, m$, according to:

$$a_k^j = \operatorname{argmax}_{i \in \mathcal{A} \setminus \mathbf{a}_k} \mathbb{E} [\mathcal{D}_\alpha (\mathbf{Y}_k^i, \{\mathbf{Y}_k^j\}_{j \in \mathbf{a}_k}) | \mathbf{Y}_{0:k-1}]. \quad (15)$$

Both (12) and (15) are selecting the nodes to sample which yield maximal divergence between the percolation prediction distribution (belief state) and the updated posterior distribution, averaged over all possible observations. Thus (12) provides a metric to assess whether to trust the predictor and defer actions until a future time or choose to take action, sample a node, and update the posterior.

1) *Lower Bound on Expected α -Divergence:* Since the expected α -Divergence in (12) is not closed form, we could resort to numerical methods for estimating this quantity. Alternatively, one could specify an analytical lower-bound that could be used in-lieu of numerically computing the expected information gain in (12) or (15).

We begin by noting that the expected divergence between the updated posterior and the predictive distribution (conditioned on previous observations) differ only through the measurement update factor, $f_k/g_{k|k-1}$ ((12) re-written):

$$\begin{aligned} & \mathbb{E}_{g_{k|k-1}} [\mathcal{D}_\alpha (p_k || p_{k|k-1})] \\ &= \mathbb{E}_{g_{k|k-1}} \left[\frac{1}{\alpha - 1} \log \mathbb{E}_{p_{k|k-1}} \left[\left(\frac{f_k}{g_{k|k-1}} \right)^\alpha \right] \right] \end{aligned} \quad (16)$$

where $f_k = f(\mathbf{Y}_k | \mathbf{Z}_k)$ and $g_{k|k-1} = g(\mathbf{Y}_k | \mathbf{Y}_{0:k-1})$. So, if there is significant overlap between the likelihood distributions of the observations, the expected divergence will tend to zero, implying that there is not much value-added in taking measurements, and thus, it is sufficient to use the predictive distribution for inferring the states.

It would be convenient to interchange the order of the conditional expectations in (16). It is easily seen that Jensen's inequality yields the following lower bound for the expected information gain

$$\begin{aligned} & \mathbb{E}_{g_{k|k-1}} [\mathcal{D}_\alpha (p_k || p_{k|k-1})] \\ & \geq \frac{1}{\alpha - 1} \log \mathbb{E}_{p_{k|k-1}} \left[\mathbb{E}_{g_{k|k-1}} \left[\left(\frac{f_k}{g_{k|k-1}} \right)^\alpha \right] \right]. \end{aligned} \quad (17)$$

Here, the inner conditional expectation can be obtained from $\mathcal{D}_\alpha (f_k || g_{k|k-1})$, which has a closed form for common distributions (e.g., multivariate Gaussians) [8].

III. ASYMPTOTIC ANALYSIS OF MARGINAL POSTERIOR

For tracking the percolation process across \mathcal{G} , we have discussed recursive updating of the posterior. However, computing these updates is generally intractable. For the remainder of the paper, we will use (4) and (5) to directly update the marginal posterior distribution using the following matrix representation:

$$\mathbf{p}_k(z) = \mathbf{D}_k(z) \mathbf{p}_{k|k-1}(z) \quad (18)$$

with updated marginal posterior $\mathbf{p}_k(z) = [p_{1,k}(z), \dots, p_{n,k}(z)]^T$ with $p_{i,k}(z) = p(Z_k^i = z | \mathbf{Y}_k^i, \mathbf{Y}_{0:k-1})$, $\mathbf{D}_k(z) = \operatorname{diag} (f_{i,k}^{(z)} / g_{i,k|k-1})$, and marginal belief state $\mathbf{p}_{k|k-1}(z) = [p_{1,k|k-1}(z), \dots, p_{n,k|k-1}(z)]^T$ with $p_{i,k|k-1}(z) = p(Z_k^i = z | \mathbf{Y}_{0:k-1})$.

Note that for $i \notin \mathbf{a}_k$, $(\mathbf{D}_k(z))_{i,i} = 1$, and $p_{i,k}(z) = p_{i,k|k-1}(z)$. Given that we can find an efficient way of updating $\mathbf{p}_{k|k-1}(z)$, according to the transition dynamics (5), we can solve a modified version of (15), for $j = 1, \dots, m$:

$$a_k^j = \operatorname{argmax}_{i \in \mathcal{A} \setminus \mathbf{a}_k} \mathbb{E} [\mathcal{D}_\alpha (\mathbf{Y}_k^i) | \mathbf{Y}_{0:k-1}] \quad (19)$$

$$\mathcal{D}_\alpha (\mathbf{Y}_k^i) = \mathcal{D}_\alpha (p_{i,k}(z) || p_{i,k|k-1}(z)), \quad 0 < \alpha < 1. \quad (20)$$

A. Pearson χ^2 Divergence of Updated Marginal Posterior

One interesting property of the Bayesian filtering equations is that the updated posterior can be written as a perturbation of the predictive percolation distribution through the following

relationship (z omitted for clarity):

$$\mathbf{p}_k = \mathbf{D}_k \mathbf{p}_{k|k-1} = \mathbf{p}_{k|k-1} + (\mathbf{D}_k - \mathbf{I}) \mathbf{p}_{k|k-1}. \quad (21)$$

Hence, when the sensors do a poor job in discriminating the observations, $\mathbf{D}_k \approx \mathbf{I}$, we have $\mathbf{p}_k \approx \mathbf{p}_{k|k-1}$. It is of interest to determine when there is significant difference between the posterior update and the prior update specified by the standard percolation equations. Recall that the updated posterior is, in the mean, equal to the predictive distribution, $\mathbb{E}[\mathbf{p}_k | \mathbf{Y}_{0:k-1}] = \mathbf{p}_{k|k-1}$. The total deviation of the updated posterior from the percolation distribution can be summarized by computing the trace of the following conditional covariance:

$$\begin{aligned} \text{tr}(\text{Cov}[\mathbf{p}_k | \mathbf{Y}_{0:k-1}]) &= \quad (22) \\ \text{tr} \left(\mathbb{E} \left[(\mathbf{p}_k - \mathbb{E}[\mathbf{p}_k | \mathbf{Y}_{0:k-1}]) (\mathbf{p}_k - \mathbb{E}[\mathbf{p}_k | \mathbf{Y}_{0:k-1}])^T | \mathbf{Y}_{0:k-1} \right] \right). \end{aligned}$$

Using (21) and properties of the trace operator, we obtain the following measure of total deviation of the updated posterior from the predictive distribution in terms of f_k and $g_{k|k-1}$:

$$\text{tr}(\text{Cov}[\mathbf{p}_k | \mathbf{Y}_{0:k-1}]) = \text{tr} \left(\mathbb{E} \left[(\mathbf{D}_k - \mathbf{I})^2 | \mathbf{Y}_{0:k-1} \right] \mathbf{p}_{k|k-1} \right) \quad (23)$$

with $\mathbf{p}_{k|k-1} = \mathbf{p}_{k|k-1} \mathbf{p}_{k|k-1}^T$. The conditional expectation in (23) is the Pearson χ^2 divergence between distributions $f_{i,k}$ and $g_{i,k|k-1}$, for all i . This joint measure of deviation is analytical for particular families of distributions and thus can be used as an alternative measure of divergence for activation of sensors [8].

B. Decay Dynamics of Updated Posterior Distribution

There has recently been significant interest in deriving the conditions of a percolation/epidemic threshold in terms of transition parameters and the graph adjacency matrix spectra for two state causal Markov processes [2], [6], [11]. Such thresholds yield conditions necessary for phase transition in the probability of local infections becoming epidemics. Knowledge of these conditions are particularly useful for designing “robust” networks, where the probability of epidemics is minimized.

Epidemic thresholds are typically obtained by extracting the sufficient conditions of the network and model parameters for the node states to be driven to their stationary point, with high probability. The probability of these events are computed using the observation independent distribution encoding the stochastic dynamics of the process [2], [6], [11].

We use the results in [2], [6] to derive a percolation threshold based upon the *updated* posterior distribution (6) assuming a restricted class of two-state Markov processes. The dominant mode of decay characterizing the conditions of a threshold should more accurately model the *current* dynamic response of the posterior distribution since the updated posterior tracks a particular “disease” trajectory better than the observation independent predictive distributions.

Formally, $Z_k^i \in \{0, 1\}$, $f_{i,k}^{(z)} = f(\mathbf{Y}_k^i | Z_k^i = z)$ is the conditional likelihood for node i , $p_{i,k} = p(Z_k^i = 1 | \mathbf{Y}_{0:k-1}^i)$, and $p_{k|k-1} = p(Z_k^i = 1 | \mathbf{Y}_{0:k-1})$. Here, we will assume that $\mathbf{Z}_k = \mathbf{0}$ is the unique absorbing state of the system.

The Bayes update for $p_{i,k}$ can be written as (i subscript omitted for clarity):

$$\begin{aligned} p_k &= \frac{f_k^{(1)}}{f_k^{(1)} p_{k|k-1} + f_k^{(0)} (1 - p_{k|k-1})} p_{k|k-1} \\ &= \frac{f_k^{(1)} / f_k^{(0)}}{1 + \frac{f_k^{(1)} - f_k^{(0)}}{f_k^{(0)}} p_{k|k-1}} p_{k|k-1} \\ &= \frac{f_k^{(1)} / f_k^{(0)}}{1 + \frac{\Delta f_k}{f_k^{(0)}} p_{k|k-1}} p_{k|k-1}. \quad (24) \end{aligned}$$

There are three different sampling/observation dependent possibilities for each individual at time k : case (1), i is not sampled and therefore, $p_k = p_{k|k-1}$, case (2), $\Delta f_k > 0$, and case (3), $\Delta f_k < 0$. We first derive a tight-upper bound for cases (2) and (3) of the form $p_k \leq c_k p_{k|k-1}$. For the remainder of the analysis we will assume that $|\frac{\Delta f_k}{f_k^{(0)}} p_{k|k-1}| < 1$ for cases (2) and (3).

In case (2), when $\Delta f_k > 0$, we can re-write (24) in terms of an *alternating geometric series*:

$$\begin{aligned} p_k &= \frac{f_k^{(1)}}{f_k^{(0)}} \left[\sum_{l=0}^{\infty} (-1)^l \left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} \right)^l \right] p_{k|k-1} \\ &\leq \frac{f_k^{(1)}}{f_k^{(0)}} \left[1 + \frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} \right] p_{k|k-1} \quad (25) \end{aligned}$$

where we have used the fact that $1/(1+|a|) \leq 1+|a|$. Recalling that $p \geq p^2$ for $0 \leq p \leq 1$, we have

$$p_k \leq \frac{f_k^{(1)}}{f_k^{(0)}} \left[1 + \frac{|\Delta f_k|}{f_k^{(0)}} \right] p_{k|k-1}. \quad (26)$$

In case (3), when $\Delta f_k < 0$, (24) can be expanded as a *geometric series*:

$$\begin{aligned} p_k &= \frac{f_k^{(1)}}{f_k^{(0)}} \left[\sum_{l=0}^{\infty} \left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} \right)^l \right] p_{k|k-1} \\ &= \frac{f_k^{(1)}}{f_k^{(0)}} \left[1 + \frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} + \sum_{l=2}^{\infty} \left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} \right)^l \right] p_{k|k-1}. \quad (27) \end{aligned}$$

Exploiting the fact that $p \geq p^2$ for $0 \leq p \leq 1$, we obtain:

$$p_k \leq \frac{f_k^{(1)}}{f_k^{(0)}} \left[1 + \frac{|\Delta f_k|}{f_k^{(0)}} \right] p_{k|k-1} + \mathcal{O} \left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1} \right) \quad (28)$$

where the higher order terms of $\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1}$ are captured in

$$\mathcal{O}\left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1}\right) = \sum_{l=2}^{\infty} \left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1}\right)^l. \quad (29)$$

Now that an upper-bound has been established for conditions when a node is sampled (under both scenarios on the signed difference of the sensor likelihoods), we can state the general equality/inequality (equality for case (1)) of $p_k \leq c_k p_{k|k-1}$ with

$$b_k = \begin{cases} 1 & , i \notin \mathbf{a}_k \\ \frac{f_k^{(1)}}{f_k^{(0)}} \left[1 + \frac{|\Delta f_k|}{f_k^{(0)}}\right] & , |\Delta f_k| > 0 \end{cases}$$

with $c_k = b_k$ for cases (1) and (2) and $c_k = b_k + \mathcal{O}\left(\frac{|\Delta f_k|}{f_k^{(0)}} p_{k|k-1}\right)$ for case (3).

After gathering all n nodes into vector notation, we have the following element-wise upper-bound on the updated belief state:

$$\mathbf{p}_k \leq \mathbf{C}_k \mathbf{p}_{k|k-1} = (\mathbf{B}_k + \mathcal{O}_k) \mathbf{p}_{k|k-1}. \quad (30)$$

with

$$\mathbf{B}_k = \text{diag}(b_{i,k}) \quad (31)$$

and

$$\mathcal{O}_k = \text{diag}\left(\mathbb{I}_{\{\Delta f_{i,k} < 0\}} \mathcal{O}\left(\frac{|\Delta f_{i,k}|}{f_{i,k}^{(0)}} p_{i,k|k-1}\right)\right) \quad (32)$$

where $\mathbb{I}_{\{\Delta f_{i,k} < 0\}}$ is the indicator function for the event $\Delta f_{i,k} < 0$.

Thus far, we have established, under the assumptions of $\left|\frac{\Delta f_k}{f_k^{(0)}} p_{k|k-1}\right| < 1$, an upper-bound for the updated posterior in terms of observation likelihoods and the belief state (30).

Next, consider the restricted class of two-state Markov processes on \mathcal{G} , for which we can produce a bound of the form

$$\mathbf{p}_{k|k-1} \leq \mathbf{S} \mathbf{p}_{k-1} \quad (33)$$

where \mathbf{S} contains information about the transition parameters and the topology of the network.

A class of models where such a bound exist is the *Susceptible-Infected-Susceptible (SIS)* model of mathematical epidemiology [2]. The *SIS* model on a graph \mathcal{G} , assumes that each of the n individuals are in states 0 or 1, where 0 corresponds to *susceptible* and 1 corresponds to *infected*. At any time k , an individual can receive the infection from their neighbors, $\eta(i)$, based upon their states at $k-1$.

Under this *SIS* model, the matrix \mathbf{S} is given by

$$\mathbf{S} = (1 - \gamma) \mathbf{I} + \beta \mathbf{A} \quad (34)$$

where the Markov transition parameters γ is the probability of

i transitioning from 1 to 0, β is the probability of transmission between neighbors i and j , and \mathbf{A} is the graph adjacency matrix (see Figure 2).

Returning to the derivation, using the bound on updating the belief state (33) and updating the posterior (30), we have by induction, the following recursion:

$$\begin{aligned} \mathbf{p}_k &\leq \mathbf{C}_k \mathbf{p}_{k|k-1} \leq \mathbf{C}_k \mathbf{S} \mathbf{p}_{k-1} \leq (\mathbf{C}_k \mathbf{S} \cdots \mathbf{C}_1 \mathbf{S}) \mathbf{p}_0 \\ &= (\mathbf{B}_k \mathbf{S} \cdots \mathbf{B}_1 \mathbf{S}) \mathbf{p}_0 + \mathcal{O}_{\mathbf{C}_k \mathbf{S}} \end{aligned} \quad (35)$$

where we have lumped the higher order modes and higher order cross-terms into $\mathcal{O}_{\mathbf{C}_k \mathbf{S}}$.

The *dominant mode of decay* of the updated posterior may be found by investigating the following eigen-decomposition:

$$\mathbf{B}_k \mathbf{S} = \left(\sum_{j=1}^n b_{j,k} \mathbf{e}_j \mathbf{e}_j^T\right) \left(\sum_{j=1}^n \lambda_j \mathbf{u}_j \mathbf{u}_j^T\right) \quad (36)$$

with $\mathbf{e}_j = [0, \dots, 0, 1, 0, \dots, 0]^T$ (1 at j^{th} element). Without loss of generality, we can assume the eigenvalues of \mathbf{S} are listed in decreasing order, $|\lambda_1| \geq \dots \geq |\lambda_n|$. Now rewriting (36), we have

$$\begin{aligned} \mathbf{B}_k \mathbf{S} &= (b_{j_k} \mathbf{e}_{j_k} \mathbf{e}_{j_k}^T + \mathcal{O}_B) (\lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \mathcal{O}_S) \\ &= (\lambda_1 b_{j_k} \mathbf{e}_{j_k} \mathbf{e}_{j_k}^T \mathbf{u}_1 \mathbf{u}_1^T + \mathcal{O}_{BS}) \end{aligned} \quad (37)$$

where $b_{j_k} = \max_{j \in \{1, \dots, n\}} b_{j,k}$ and the $\mathcal{O}_B, \mathcal{O}_S, \mathcal{O}_{BS}$ variables corresponds to the higher order terms. Inserting (37) into (35), and matching the largest eigenvalues of \mathbf{B}_k with λ_1 we obtain

$$\begin{aligned} \mathbf{p}_k &\leq (\mathbf{B}_k \mathbf{S} \cdots \mathbf{B}_1 \mathbf{S}) \mathbf{p}_0 + \mathcal{O}_{\mathbf{C}_k \mathbf{S}} \\ &= \lambda_1^k \prod_{l=1}^k b_{j_l} \left(\prod_{l=1}^k (\mathbf{e}_{j_l} \mathbf{e}_{j_l}^T \mathbf{u}_1 \mathbf{u}_1^T)\right) \mathbf{p}_0 + \mathcal{O}(\varphi^k). \end{aligned} \quad (38)$$

Thus, at large k , the dominant mode of the posterior goes as $\lambda_1^k \prod_{l=1}^k b_{j_l}$ (the modes in $\mathcal{O}(\varphi^k)$ decay faster than the dominant mode presented above).

We can see that if the spectral radius of \mathbf{S} is less than one, $|\lambda_1| < 1$, then for large k , $\mathbf{p}_k \rightarrow \mathbf{0}$, which is the unique absorbing state of the system.

This epidemic threshold condition on λ_1 has been previously established for unforced (observation independent) *SIS*-percolation processes [2]. However, in the tracking framework, the rate at which the posterior decays to the *susceptible* state is damped by an additional measurement dependent factor, $\prod_{l=1}^k b_{j_l}$, resulting from using the *updated* posterior distribution.

This measurement-dependent dominant mode of the posterior should more accurately model the true dynamic response of the node states better than that in [2] since the posterior better tracks the truth than the unforced predictive distribution.

Additionally, this dominant mode of the updated posterior distribution allows one to simulate the response of the percolation threshold to intervention and control actions which are designed to increase the threshold, such that the probability of epidemics is minimized.

IV. NUMERICAL EXAMPLE

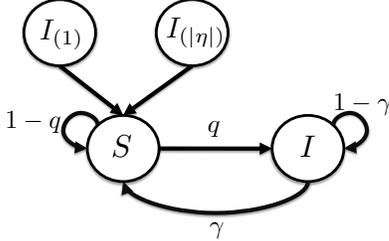


Fig. 2. SIS Markov Chain for Node i Interacting with the Infected States of its Neighbors

Here, we present results of simulations of our adaptive sampling for the active tracking of a causal Markov ground truth process across a random 200 node, scale-free network (Figure 3(a)). Since most modern networks in which this method is most applicable, e.g., social networks, tend to be scale-free in their degree distribution (see Figure 3(b)), the proposed network shall suffice for extracting various statistics under the proposed adaptive tracking method. Since the goal in tracking is to accurately classify the states of each node, we are interested in exploring the detection performance as the likelihood of an epidemic increases through the percolation threshold for this network.

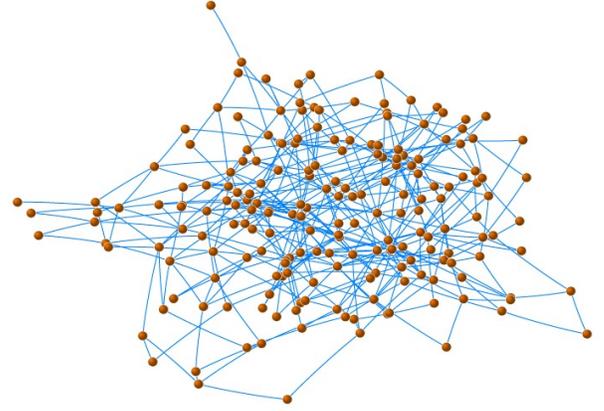
One would expect different phase transitions (thresholds) in detection performance for various sampling strategies, ranging from the lowest threshold for unforced predictive distributions to highest for a continuous monitoring of all n nodes. We will present a few of these detection surfaces that depict these phase transitions for the unforced percolation distribution, random $m = 40$ node sampling, and our proposed information theoretic adaptive sampling of $m = 40$.

Here, we will restrict our simulations to the two-state SIS model of mathematical epidemiology described above.

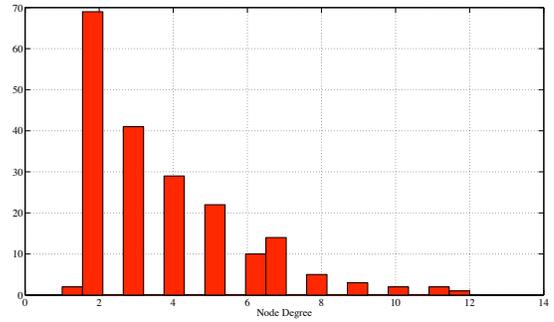
The sensor models (4), are of the form of two-dimensional multivariate Gaussians with common covariance and shifted mean vector. The transition dynamics of the i^{th} individual (5), for the SIS model is given by:

$$Z_k^i | \mathbf{Z}_{k-1}^{\{i, \eta(i)\}} \sim (1 - \gamma) Z_{k-1}^i + (1 - Z_{k-1}^i) \left[1 - \prod_{j \in \eta(i)} (1 - \beta Z_{k-1}^j) \right]. \quad (39)$$

where $Z_{k-1}^i \in \{0, 1\}$ is the indicator function of i being infected at time $k - 1$. The transmission term between i and $\eta(i)$ is known the Reed-Frost model [2], [6], [10]. Since the tail



(a) 200 Node Scale Free Synthetic Network



(b) Degree Distribution of the 200 Node Scale Free Network

Fig. 3. Structure of the 200 Node Synthetic Network Used in Simulations

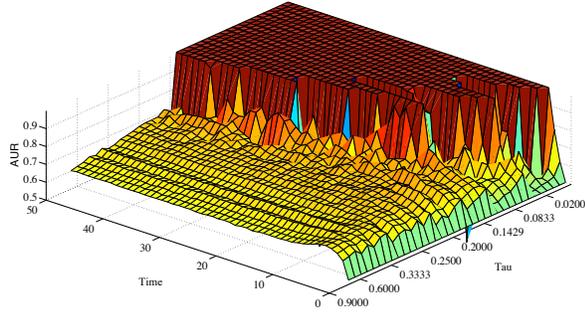
of the degree distribution of our synthetic scale-free graph contains nodes with degree greater than 10, updating (11) exactly is unrealistic and we must resort to approximate algorithms. Here, we will assume the mean field approximation used by [2] for this SIS model, resulting in the following marginal belief state update for the i^{th} node of infected ($Z_k^i = 1$):

$$p_{i,k|k-1} = (1 - \gamma)p_{i,k-1} + (1 - p_{i,k-1}) \left[1 - \prod_{j \in \eta} (1 - \beta p_{j,k-1}) \right]. \quad (40)$$

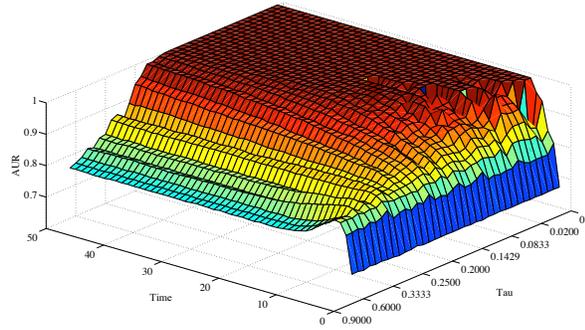
Equation (40) allows us to efficiently update the marginal belief state directly for all n nodes which are then used for estimating the best m measurements using (19).

As we are interested in detection performance, as a function of time and epidemic intensity, the Area Under the ROC Curve (AUR) is a natural statistic to quantify the detection power at each time and each propensity of epidemic (detection of the infected state). The AUR is evaluated at each time k , each SIS percolation intensity parameter

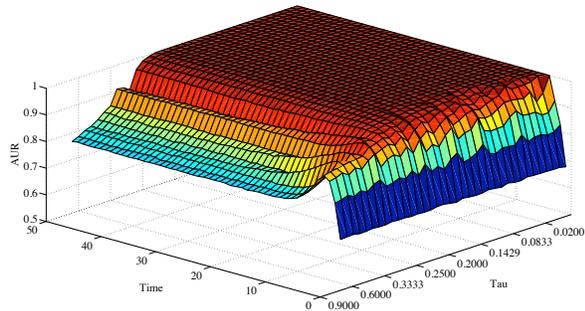
$$\tau = \beta/\gamma \quad (41)$$



(a) AUR Surface for Unforced Prediction Distribution (no evidence acquired throughout the monitoring)



(b) AUR Surface for Updated Posterior Distribution with $m = 40$ Random Measurements at Each Time k



(c) AUR Surface for Updated Posterior Distribution with $m = 40$ Information Theoretic Adaptive Measurements at Each Time k

Fig. 4. Detection Performance Surface: Area Under the ROC (AUR) Curve Surface as a Function of Percolation Parameter $\tau = \beta/\gamma$ and Time

and over 500 random initial states of the network. For the *SIS* model, τ is the single parameter (aside from the topology of the graph) that characterizes the intensity of the percolation/epidemic. It is useful to understand how the detection performance varies as a function of epidemic intensity, as it indicates how well the updated posteriors are playing “catch-up” in tracking the true dynamics on the network.

For this *SIS* model, the percolation threshold is defined as $\tau_c = 1/\lambda_1(\mathbf{A})$ where $\lambda_1(\mathbf{A}) = \max_{i \in \{1, \dots, n\}} |\lambda_i|$ is the

spectral radius of the graph adjacency matrix, \mathbf{A} [2]. Values of τ greater than τ_c imply that any infection tend to become an epidemic, whereas those values less than τ_c imply that small epidemics tend to die out.

For the network under investigation (Figure 3(a)), $\tau_c = 0.1819$. We see from Figure 4(a) that a phase transition in detection power (AUR) for the unforced percolation distribution does indeed coincide with the epidemic threshold τ_c . While the epidemic threshold for the random and adaptive sampling policies is still $\tau_c = 0.1819$, the measurements acquired allow the posterior to better track the truth, but only up to their respective phase transitions in detection power (see Figures 4(b) and 4(c)).

Figure 4(c) confirms that the adaptive sampling better tracks the truth than randomly sampling nodes, while pushing the phase transition in detection performance to higher percolation intensities, τ . We see that the major benefit of the adaptive sampling is apparent when conditions of the network are changing moderately, at medium epidemic conditions. Beyond a certain level of percolation intensity, more resources will need to be allocated to sampling to maintain a high level of detection performance.

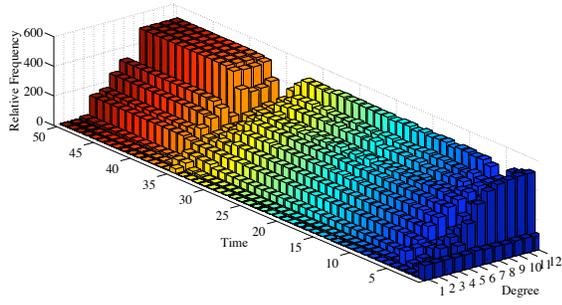
A heuristic sampling strategy based on the topology of \mathcal{G} was also explored (results not shown) by sampling the “hubs” (highly-connected nodes). However, detection performance was only slightly better than random sampling and poorer than our adaptive sampling method.

It is often useful for developing sampling heuristics and offline control/intervention policies to inspect what *type* of nodes, topologically speaking, is the adaptive sampling strategy targeting, under various network conditions (different values of τ). In Figure 5, the relative frequency of nodes sampled with a particular degree is plotted against time (under the $m = 40$ adaptive sampling strategy) for three different values of τ (over 500 random initial conditions of the network).

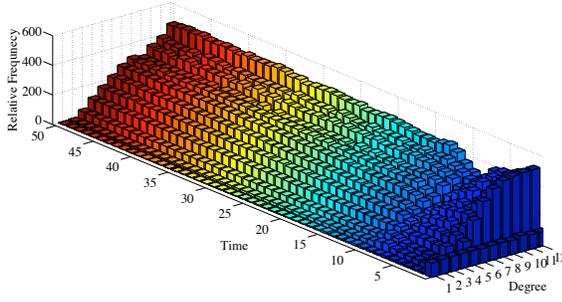
For the larger of the three values explored ($\tau = 0.5 > \tau_c$) we see that the sampling is approximately uniform across the nodes of each degree on the graph (Figure 5(c)). Therefore, under extremely intense epidemic conditions, the adaptive sampling strategy is targeting all nodes of each degree equally, and therefore, it is sufficient to perform random sampling. For the two lower values of τ , Figure 5(a) and Figure 5(b) (near τ_c), we see that adaptive policy targets highly connected nodes more frequently than those of lesser degree and thus, it is more advantageous to exploit such a strategy, as compared to random sampling (see AUR surface in Figure 4(c)).

V. DISCUSSION

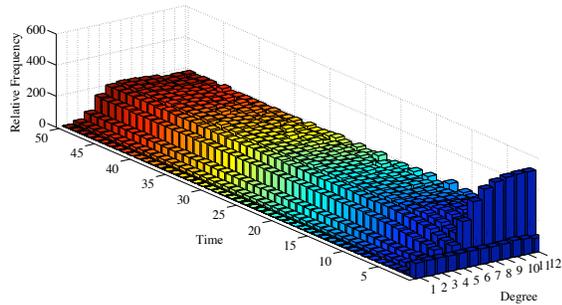
In this paper, we have presented an information theoretic framework for recursively selecting the best subset of nodes



(a) $\tau = 0.125 < \tau_c$



(b) $\tau = 0.2143 \approx \tau_c$



(c) $\tau = 0.5 > \tau_c$

Fig. 5. Relative Frequency of Nodes Sampled of a Given Degree under $m = 40$ Adaptive Sampling Strategy

to sample in a dynamic Bayesian network that yield the largest expected information gain about the hidden state of the network. This framework can be applied to a variety of problem domains, including actively tracking an influenza outbreak across a population or adaptively monitoring the diffusion of information across large networks, such as a terrorism network.

Within the proposed adaptive tracking/sampling framework, we have derived conditions for a network specific percolation threshold using an updated posterior distribution rather than an observation independent predictive distribution. These conditions recover the unforced percolation threshold derived in [2] but with an additional factor involving sensor likelihood terms due to measurements obtained throughout the monitoring. A term of the form $\lambda_1^k \prod_{l=1}^k b_{jl}$ (derived in (38)) was shown to be the dominant mode of the updated posterior dynamic response to active intervention of immunizing the nodes (holding node states constant). The conditions of the

threshold, using the updated posterior, should more accurately model the phase transition in detection performance and thus, enable a better assessment of immunization strategies and any subsequent observations resulting from such actions. The framework and modes of decay of the updated posterior should provide additional insight into active monitoring of large complex networks under resource constraints.

Exploring phase transitions in updated posteriors for other classes of diffusion is the subject of future work. One particularly interesting question is identifying conditions of phase transitions between multi-state processes (> 2) and explore the rates at which a system transitions between various states.

REFERENCES

- [1] E. K. CHONG, C. M. KREUCHER, AND A. O. HERO, III, *Partially observable markov decision process approximations for adaptive sensing*, Discrete Event Dynamic Systems, 19 (2009), pp. 377–422.
- [2] D. CHAKRABARTI, Y. WANG, C. WANG, J. LESKOVEC, AND C. FALOUTSOS, *Epidemic thresholds in real networks*, ACM Trans. Inf. Syst. Secur., 10 (2008), pp. 1094–9224.
- [3] C. KREUCHER, K. KASTELLA, AND A. O. HERO, III, *Sensor management using an active sensing approach*, Signal Process., 85 (2005), pp. 607–624.
- [4] R. COHEN, S. HAVLIN, AND D. BEN AVRAHAM, *Efficient immunization strategies for computer networks and populations*, Physical Review Letters, 91 (2003), p. 247901.
- [5] A. DOUCET, S. GODSILL, AND C. ANDRIEU, *On sequential monte carlo sampling methods for bayesian filtering*, Statistics and Computing, 10 (2000), pp. 197–208.
- [6] M. DRAIEF, A. J. GANESH, AND L. MASSOULIÉ, *Thresholds for virus spread on networks*, Ann. Appl. Probab., 18 (2008), pp. 359–378.
- [7] S. EUBANK, H. GUCLU, ANIL, M. V. MARATHE, A. SRINIVASAN, Z. TOROCZKAI, AND N. WANG, *Modelling disease outbreaks in realistic urban social networks*, Nature, 429 (2004), pp. 180–184.
- [8] A. O. HERO, D. CASTENON, D. COCHRAN, AND K. D. KASTELLA, *Foundations and applications of sensor management*, Springer, NY, 2007.
- [9] D. J. C. MACKAY, *Information Theory, Inference & Learning Algorithms*, Cambridge University Press, 2002.
- [10] M. E. J. NEWMAN, *The spread of epidemic disease on networks*, Physical Review E, 66 (2002), p. 016128.
- [11] M. E. J. NEWMAN, *Threshold effects for two pathogens spreading on a network*, Physical Review Letters, 95 (2005), p. 108701.
- [12] B. NG, L. PESHKIN, AND A. PFEFFER, *Factored particles for scalable monitoring*, in In Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, 2002, pp. 370–377.
- [13] Z. RABINOVICH AND J. S. ROSENSCHEIN, *Extended markov tracking with an application to control*, in The Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004, pp. 95–100.
- [14] A. X. ZHENG, I. RISH, AND A. BEYGELZIMER, *Efficient test selection in active diagnosis via entropy approximation*, in Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence, 2005, p. 675.
- [15] M. KEELING AND K. EAMES, *Networks and epidemic models*, Journal of the Royal Society Interface, 2 (2005), pp. 295–307.
- [16] D. J. C. MACKAY, *Information Theory, Inference & Learning Algorithms*, Cambridge University Press, 2002.
- [17] L. A. MEYERS, *Contact network epidemiology: Bond percolation applied to infectious disease prediction and control*, Bull. Amer. Math. Soc., 44 (2007), pp. 63–86.
- [18] I. RISH, M. BRODIE, S. MA, N. ODINTSOVA, A. BEYGELZIMER, G. GRABARNIK, AND K. HERNANDEZ, *Adaptive diagnosis in distributed systems.*, IEEE Trans Neural Netw, 16 (2005 Sep), pp. 1088–1109.