Index-Based Sampling Policies for Tracking Dynamic Networks under Sampling Constraints

Ting He*, Animashree Anandkumar†, and Dakshi Agrawal*.
*IBM T.J. Watson Research Center, Hawthorne, NY, USA. Email: {the,agrawal}@us.ibm.com
†Dept. of EECS, University of California Irvine, Irvine, CA, USA. Email: {a.anandkumar}@uci.edu

***Abstract*—We consider the problem of tracking the topology of a large-scale dynamic network with limited monitoring resources. By modeling the dynamics of links as independent ON-OFF Markov chains, we formulate the problem as that of maximizing the overall accuracy of tracking link states when only a limited number of network elements can be monitored at each time step. We consider two forms of sampling policies: link sampling, where we directly observe the selected links, and node sampling, where we observe states of the links adjacent to the selected nodes. We reduce the link sampling problem to a Restless Multi-armed Bandit (RMB) and prove its indexability under certain conditions. By applying the Whittle's index policy, we develop an efficient link sampling policy with methods to compute the Whittle's index explicitly. Under node sampling, we use a linear programming (LP) formulation to derive an extended policy that can be reduced to determining the nodes with maximum coverage of the Whittle's indices. We also derive performance upper bounds in both sampling scenarios. Simulations show the efficacy of the proposed policies. Compared with the myopic policy, our solution achieves significantly better tracking performance for heterogeneous links.**

***Index Terms*—Network sampling, network topology tracking, restless multi-armed bandits, Whittle's index policy.**

I. Introduction

Large-scale dynamic networks are widely used to model relationships and interactions among entities in domains as diverse as IT systems, sociology, and economics. A fundamental challenge in managing large dynamic networks is designing sampling policies so that a network can be efficiently monitored. A concrete example comes from the cloud computing paradigm where large data centers nimbly shuffle load between tens of thousands of servers to optimize server utilization. The monitoring infrastructure in such data centers may gather hundreds of metrics from each server which are then sent to a central location for further analysis. Such analysis often involves determining whether certain relationship (e.g., time-lagged causality of crossing certain thresholds) exists between the metrics. The entire set of relationships can be modeled as a logical network of the metrics where links denote the presence of the pairwise relationships. Due to the large number of metrics involved, it may not be feasible to assess all the relationships at all times, and we need to reduce the problem size by sampling the network. This raises several questions: How do we estimate the unobserved link states when tracking a sampled dynamic network? Given sampling constraints, how do we sample the network to maximize tracking accuracy? Although our research was initially motivated by this scenario, these questions are fundamental in many other application domains as well.

Decisions on which network nodes or links (henceforth referred to as network elements) should be selected for monitoring critically depend on the nature of their temporal evolution. Intuitively, slowly-evolving elements can be sampled at a lower rate, and hence the monitoring resources originally dedicated to them can be deployed elsewhere to better keep track of rapidly-changing elements. In this work, we address the following issues: Are there provably efficient sampling policies that explicitly take into account the monitoring constraints and the temporal evolution of the network? Which networks are amenable for tracking and how does the network topology affect the tracking accuracy? The answers to these questions will provide important insights towards engineering real-world solutions.

A. Our Approach & Contributions

We model the network as a graph $G = (V, E)$ in which a vertex $v \in V$ represents an entity of interest and a link $e \in E$ represents a pairwise interaction between two entities that we would like to monitor. In our formulation, the set of vertices V is fixed and each link in E has a binary state ("ON" or "OFF") which evolves according to a discrete-time Markov chain, independent of other links. We consider two scenarios, viz., *link sampling* and *node sampling*. Under link sampling, we assume that we can sample at most K links at a time to determine their states. Under node sampling, we assume that we can sample at most K nodes at a time to determine the states of all their adjacent links. In both cases, the states of unobserved links need to be estimated from the past observations. Our goal is to track the states of all the links in E with maximum accuracy under given sampling constraints. In this setting, our contributions are as follows:

- *Theoretical framework:* We develop a systematic framework for optimally tracking the link states of dynamic networks through either link or node sampling. We apply *Bayesian filtering* and *maximum a posteriori (MAP) estimation* to optimally estimate the states of unsampled links and design sampling policies to maximize the estimation accuracy under sampling constraints.

The first and the third authors are supported in part by the U.S. National Institute of Standards and Technology under Agreement Number 60NANB10D003. The second author is supported in part by a MURI funded through ARO Grant W911NF-06-1-0076.

• *Efficient solutions:* Under link sampling, we reduce the problem to the well-known *Restless Multi-armed Bandit (RMB)* [1]. The RMB is known to be PSPACE-hard in general [2] unless it is *indexable*, in which case the *Whittle's index policy* [1] is known to be an efficient solution. We establish indexability of our problem and develop an efficient algorithm to compute the Whittle's index. Under node sampling, we show that our problem cannot be reduced to RMB. Instead, we formulate a *linear program (LP)* under relaxed sampling constraints, a technique inspired by Whittle's original formulation [1]. This yields an extended Whittle's index policy which selects nodes that maximize the sum of the Whittle's indices at covered links. A byproduct of our computation is a procedure to compute upper bounds on the optimal performance in both sampling scenarios.

• *Numerical studies:* Our simulations demonstrate the efficacy of the proposed solutions. We show that the Whittle's index policy is near-optimal for link sampling, but has a gap (up to 10%) when compared with the upper bound for node sampling. Compared with the myopic policy, the (extended) Whittle's index policy performs similarly for i.i.d. links but increasingly better (up to more than 20%) as link activities diverge.

We note a limiting assumption in our model—that each link evolves independently according to a Markov chain. This may not be true in certain scenarios, e.g., in a social network, transitivity of relationships is likely to cause correlated link evolution, in which case this assumption precludes the application of our techniques. Nonetheless, our model captures a fundamental aspect of network evolution and provides a sound foundation for further development of network sampling techniques in more general scenarios. To the best of our knowledge, this is the first systematic study of tracking dynamic networks under monitoring constraints that explicitly takes into account the temporal evolution of the networks.

B. Related Work

Learning static network topology has received considerable attention, with extensive studies including those on sampling social networks [3], mapping the Internet [4], and network tomography in a variety of settings. Diverse measurements have been used to infer network structures, such as passive measurements [5], node queries [6], edge counts [7], and subgraph counts [8]. However, these works all assume static network topologies, and learning dynamically-changing networks is a relatively unexplored area. A related problem of tracking dynamic bipartite graphs is considered in [9] but only to track the proximity measures between nodes rather than the graph itself, without any sampling constraints.

Our solution is based on Restless Multi-armed Bandits (RMBs), which have been extensively studied since the seminal work of Whittle [1], where an index-based policy is proposed for a special class of RMBs known as indexable RMBs. It was later proved that finding an optimal policy for a general RMB, where indexability may not hold, is PSPACE-hard [2]. In [10], sufficient conditions for indexability are derived. In [11], conditions for asymptotic optimality of the Whittle's index policy are established. A special class of RMBs based on ON-OFF Markov chains has been extensively studied in the context of UAV routing [12] and dynamic channel access [13], where the goal is to access arms when they are in "ON" states. Our work uses a similar model, but is fundamentally different in that we want to track the states of all the arms, leading to very different solutions and conclusions as shown later.

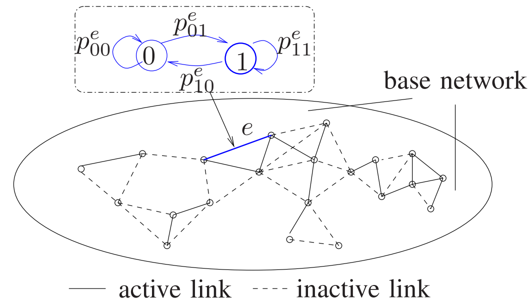


Fig. 1. A dynamic network: each dynamic link changes states over time by a Markov chain; the true topology at any time only contains the active links.

The rest of the paper is organized as follows. Section II formulates the problem, Section III presents some preliminaries, Section IV and Section V deal with link and node sampling, respectively, Section VI gives simulation results, and Section VII concludes the paper.

II. Problem Formulation

Notations: For set A, let $|A|$ denote its cardinality. We use the convention that X denote the random variable, x the realization, and $\mathbf{X}$ (or $\mathbf{x}$) the vector versions. Let I denote the indicator function. We will use "arm" and "link" interchangably.

A. Dynamic Networks and Temporal Process

As illustrated in Fig. 1, let $G = (V, E)$ denote a *base network* over a node set V, in which the links in E are active intermittently. Let $N := |V|$ and $M := |E|$ be the numbers of nodes and links. Given a node $v \in V$, we use $\mathcal{N}(v; G)$ ($\mathcal{N}(v)$ for short) to denote the set of its adjacent links in G. The network evolves in discrete time[1] and let $G_t = (V_t, E_t)$ be the network at time t. We assume that the set of nodes is fixed ($V_t \equiv V$), but the set of active links changes dynamically. Using $S_e(t) \in \{0, 1\}$ to denote the state of link $e \in E$ at time t, we have $S_e(t) = 1$ ("active") for $e \in E_t$ and $S_e(t) = 0$ ("inactive") otherwise. To model the temporal dynamics of the network, we assume that the link state $S_e(t)$ evolves according to a discrete-time Markov chain, independent of the other links. Denote the transition probabilities by

$$p^e_{01} := \mathbb{P}[S_e(t) = 1 | S_e(t-1) = 0], \tag{1}$$
$$p^e_{11} := \mathbb{P}[S_e(t) = 1 | S_e(t-1) = 1]. \tag{2}$$

Assume p^e_{01} and p^e_{11} are known for all the links $e \in E$.

[1]The problem can also be formulated in continuous time.

B. Sampling Constraints

For large networks, it may not be feasible to check the states of all links simultaneously, e.g., this requires correlating the sequences of measurements for each pair of metrics in a cloud dependency network. This imposes constraints on concurrent sampling. We consider two sampling scenarios, viz., *link sampling* and *node sampling*. Under link sampling, we assume that at most K links can be sampled at any time and their states can be directly observed. Under node sampling, at most K nodes can be sampled at a time and the states of all the links adjacent to these nodes can be directly observed. In both cases, the states of the unobserved links need to be predicted from past observations.

Let $\pi = (\pi_t)_{t\in\mathbb{N}}$ denote a policy which samples the network under the given constraints. Let $E^{\pi_t} \subset E$ denote the subset of links which are directly observed. Under link sampling, it is required that $|E^{\pi_t}| \leq K$ for all $t \in \mathbb{N}$, while under node sampling, it is required that E^{π_t} can be covered by at most K nodes. Let $\Pi_{E,K}$ and $\Pi_{V,K}$ be the set of valid link and node sampling policies, respectively.

Let $Z_e^\pi(t)$ be the observed state of link e at time t, i.e., $Z_e^\pi(t) = S_e(t)$ if $e \in E^{\pi_t}$ and $Z_e^\pi(t) = \emptyset$ otherwise. Let $\mathbf{Z}_e^\pi(t) := (Z_e^\pi(s))_{s=1}^t$ be the observation history of link e until time t. For the links that are not observed at a given time, we can predict their current state based on the past observations. Let $\widehat{S}_e(t) : \mathbf{Z}_e^\pi(t) \mapsto \{0, 1\}$ be the estimated state.

C. Optimal Tracking of Dynamic Networks

Our goal is to track the link states with the maximum accuracy under given sampling constraints. To this end, we measure the reward of a sampling policy π at time t by the expected number of correctly tracked links:

$$R(t;\pi) := \sum_{e\in E} \mathbb{P}[S_e(t) = \widehat{S}_e(t)]. \tag{3}$$

For the overall reward, we consider the discounted case with *discount factor* β, where the long-term reward is given by[2]

$$R_\infty(\pi) := \mathbb{E}_\pi[\sum_{t=1}^{\infty} \beta^{t-1} R(t;\pi)], \quad \beta \in [0, 1). \tag{4}$$

We want to find policies which maximize the tracking accuracy measured by (4) under link or node sampling constraints:

$$\pi^*_{E,K} := \arg\max_{\pi\in\Pi_{E,K}} R_\infty(\pi), \quad \pi^*_{V,K} := \arg\max_{\pi\in\Pi_{V,K}} R_\infty(\pi), \tag{5}$$

III. Preliminaries

A. MAP Estimation

Under a fixed sampling policy π, we now derive the optimal estimator for unobserved links. For any link $e \in E$, given the past and current observations $\mathbf{Z}_e^\pi(t)$, the optimal estimator which minimizes the error probability is the maximum a posteriori (MAP) estimator [14]

$$\widehat{S}_e(t) = \begin{cases} 1, & \text{if } \mathbb{P}[S_e(t) = 1 | \mathbf{Z}_e^\pi(t)] \geq 0.5, \\ 0 & \text{o.w.} \end{cases} \tag{6}$$

It is clear that when a link is observed at time t, i.e., $e \in E^{\pi_t}$, then $\widehat{S}_e(t) = S_e(t)$ and there is no tracking error.

Under the Markovian link model in (1–2), the MAP estimator can be efficiently implemented using a Bayesian filter. Let $X_e^\pi(t) := \mathbb{P}[S_e(t) = 1 | \mathbf{Z}_e^\pi(t-1)]$ denote the *belief* that the link is in the active state given all the past observations. The update rule for the belief is given by

$$X_e^\pi(t) = \begin{cases} p_{11}^e & \text{if } Z_e^\pi(t-1) = 1, \\ p_{01}^e & \text{if } Z_e^\pi(t-1) = 0, \\ \mathcal{T}_e(X_e^\pi(t-1)) & \text{if } Z_e^\pi(t-1) = \emptyset, \end{cases} \tag{7}$$

where the mapping $\mathcal{T}_e(x) := (p_{11}^e - p_{01}^e)x + p_{01}^e$ denotes a passive update under the Markovian rule. Hence, the tracking accuracy under MAP estimation is given by

$$\mathbb{P}[\widehat{S}_e(t) = S_e(t)] = \begin{cases} 1, & \text{if } e \in E^{\pi_t}, \\ \max(X_e^\pi(t), 1 - X_e^\pi(t)), & \text{o.w.} \end{cases} \tag{8}$$

It thus suffices to evaluate $R(t;\pi)$ in (3) by (8).

B. Recap of Restless Multiarmed Bandits

Since our problem is closely related to the Restless Multi-armed Bandits (RMB), we first provide a brief recap. In a RMB [1], M stochastic arms are given, where each arm e is an independently evolving Markov chain with different transition matrices $\mathbf{P}_e^U = [P_e^U(i,j)]$ under active ($U = 1$) and passive ($U = 0$) actions. Given rewards $R_e^U(x)$ for taking action U when e is in state x, the goal is to select actions such that the total reward as in (4) is maximized, under the constraint that only K arms can be activated at each time.

Whittle [1] proposed an efficient index-based policy for a limited class of RMBs which are called *indexable*. The policy assigns an index to each arm at every time step depending on its current state, and activates the arms with the K largest indices. It is a simple policy with only $O(M)$ computational complexity at each step. To define indexability and Whittle's index, we introduce a notion of *subsidy for passivity*: under subsidy m, an arm earns an extra reward m for being passive, i.e., the passive reward becomes $m + R_e^0(x)$. Define a *passive set* $\mathcal{P}(m)$ as the set of states x where the passive action is optimal under subsidy m.

Definition 1 (Indexability of RMB): A stochastic arm is indexable if the passive set $\mathcal{P}(m)$ monotonically increases from the empty set to the full state space as m increases from $-\infty$ to ∞. An RMB is indexable if every arm is indexable.

Definition 2 (Whittle's Index): For an indexable arm in state x, the Whittle's index $W(x)$ is defined as

$$W(x) := \inf_{m\in\mathbb{R}} \{m : x \in \mathcal{P}(m)\}. \tag{9}$$

Hence, the Whittle's index for a state is the minimum subsidy needed to make it worthwhile to be passive (assuming optimal actions in the future). Under regularity conditions, the

[2]Strictly, (4) should also be a function of the initial state $\mathbf{x}(1)$, but we suppress this in the sequel for convenience.

Whittle's index policy is asymptotically optimal for i.i.d. arms as $M \to \infty$ and $K/M \to c$ (a constant) [11]. Empirical studies have also demonstrated its superior performance in various applications (see [13]). The Whittle's index policy is thus considered a simple and effective solution for indexable RMBs. On the other hand, a general RMB is PSPACE-hard [2]. Hence, establishing indexability and finding the Whittle's index are crucial for solving RMBs efficiently.

IV. Network Tracking under Link Sampling

We now cast the link sampling problem as a RMB. The node sampling problem will be considered later in Section V.

A. Restless Bandit Formulation

Each arm corresponds to a link $e \in E$, and the belief $X_e(t) \in [0,1]$ defined as in (7) (π has been dropped) denotes the state of arm e at time t. An active action is taken ($U_e(t) = 1$) if e is sampled; otherwise, it is passive ($U_e(t) = 0$). We can thus rewrite the update of $X_e(t)$ based on the sampling action[3]

$$X_e(t+1) = \begin{cases} \left.\begin{array}{ll} p_{11}^e & \text{w.p. } X_e(t) \\ p_{01}^e & \text{w.p. } 1 - X_e(t) \end{array}\right\} & \text{if } U_e(t) = 1, \\ \mathcal{T}_e(X_e(t)) & \text{if } U_e(t) = 0. \end{cases} \quad (10)$$

The transition probability from belief y to belief x is thus

$$P_e^U(y,x) = \begin{cases} I_{x=\mathcal{T}_e(y)}, & \text{if } U = 0, \\ y I_{x=p_{11}^e} + (1-y) I_{x=p_{01}^e}, & \text{o.w.,} \end{cases} \quad (11)$$

and the one-step reward is given by

$$R_e^1(x) = 1, \; R_e^0(x) = \max(x, 1-x). \quad (12)$$

Under subsidy m, the passive reward will be replaced by $R_e^0(x) = m + \max(x,\, 1-x)$.

A subtle difference between our problem and a classic RMB is that our reward in (12) does not depend on the actual link state $S_e(t)$, but on the belief $X_e(t)$. Thus, our problem is actually an RMB on uncountable state space even though the underlying link state is binary.

B. Myopic Sampling Policy

Before we proceed to computing the Whittle's index policy for the above RMB, we first present a simple index policy called the *myopic policy*. It is a greedy policy which maximizes the immediate reward in the current time step. For the RMB for link sampling, it assigns an index

$$Y_e(x) := \min(x, 1-x) \quad (13)$$

for each link in state x and picks the K links with the largest indices. This policy is always applicable (i.e., the RMB need not be indexable) and easy to compute. The drawback is that it is suboptimal in general. We have the following result comparing it to the Whittle's index.

Proposition 1: For the link sampling problem, the Whittle's index (if exists) is no smaller than the myopic index, i.e., $W(x) \geq Y(x)\ \forall x \in [0,\, 1]$. Moreover, $W(x) \to Y(x)$ as $|p_{11} - p_{01}| \to 0$.

Proof: See [15]. □

For the special case of $p_{11}^e = p_{01}^e$, the link state $S_e(t)$ becomes i.i.d. over time, and it can be shown that the Whittle's index policy and the myopic policy are equivalent and both optimal. The two will be further compared in Section VI.

C. Indexability of the RMB

Since the Whittle's index policy is only valid for indexable arms, it is crucial to establish indexability of the above RMB. We will focus on one arm and drop the subscript e. Define the *value function* $V_{\beta,m}(x)$ as the maximum total discounted reward for an arm initially at state x under discount β and subsidy m. By definition, it satisfies

$$V_{\beta,m}(x) = \max(V_{\beta,m}(x; U=0), V_{\beta,m}(x; U=1)), \quad (14)$$

where $V_{\beta,m}(x;U)$ is the optimal discounted reward if taking action U in the first time step. By substituting for the reward functions in (12), we have

$$V_{\beta,m}(x; U{=}0) = m + \max(x,\, 1-x) + \beta V_{\beta,m}(\mathcal{T}(x)), \quad (15)$$

$$V_{\beta,m}(x; U{=}1) = 1 + \beta\left(x V_{\beta,m}(p_{11}) + (1-x) V_{\beta,m}(p_{01})\right). \quad (16)$$

Equations (14–16) jointly give an equation of the (unknown) value function $V_{\beta,m}(x)$, called the *Bellman equation*. Theoretically, once we solve the Bellman equation for $V_{\beta,m}(x)$, we can obtain the optimal policy for the single-armed bandit under subsidy m and the associated passive set as

$$\pi_m^*(x) = \arg\max_{U=0,1} V_{\beta,m}(x;\, U), \quad (17)$$

$$\mathcal{P}(m) = \{x :\, V_{\beta,m}(x;\, U=0) \geq V_{\beta,m}(x;\, U=1)\}. \quad (18)$$

We now test the indexability of the link sampling problem by comparing the passive set (18) with the conditions in Definition 1. We first show the following property of $\mathcal{P}(m)$.

Lemma 1: The optimal policy for the single-armed bandit with subsidy m is a threshold policy: $\pi_m^*(x) = 1$ if and only if $\tau^-(m) < x < \tau^+(m)$, i.e., $\mathcal{P}(m) = [0,\, \tau^-(m)] \cup [\tau^+(m),\, 1]$.

Proof: See [15]. ■

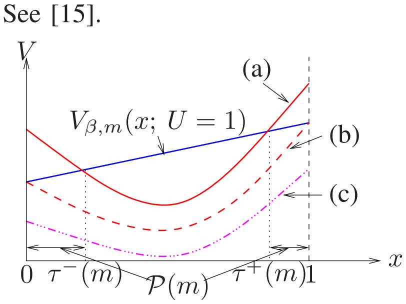


Fig. 2. Threshold structure of the optimal policy: value function $V_{\beta,m}(x;\, U=0)$ for (a) $m > 0$, (b) $m = 0$, and (c) $m < 0$.

The thresholds $\tau^-(m)$, $\tau^+(m)$ are intersecting points between $V_{\beta,m}(x;\, U=0)$ and $V_{\beta,m}(x;\, U=1)$ (see [15]), as illustrated in Fig. 2. This structure leads us to analyzing the properties of $\tau^-(m)$ and $\tau^+(m)$, with the following result.

[3]"w.p." means "with probability".

Lemma 2: The thresholds $\tau^-(m)$, $\tau^+(m)$ are monotone increasing and decreasing, respectively, with m for $\beta \leq 0.5$.

Proof: See [15]. ■

We now prove the indexability of our problem.

Theorem 1: The RMB for link sampling is indexable if the discount factor $\beta \leq 0.5$.

Proof: By the proof of Lemma 1, we see that $\mathcal{P}(m) = \emptyset$ for $m < 0$ and $\mathcal{P}(m) = [0, 1]$ for $m > m_{\max}$. For $m \in [0, m_{\max}]$, Lemma 2 implies that $\mathcal{P}(m)$ is monotone increasing with m. Hence, Definition 1 is satisfied. ■

In fact, numerical calculation (Fig. 4) shows that the monotonicity of $\tau^-(m)$, $\tau^+(m)$ holds for $\beta > 0.5$ as well. We conjecture that the RMB is indexable for all $\beta \in [0, 1)$.

D. Computation of Whittle's Index

By definition (9), the Whittle's index of state x is the subsidy m such that $\tau^-(m) = x$ or $\tau^+(m) = x$. To evaluate $\tau^-(m)$, $\tau^+(m)$, it is necessary to compute $V_{\beta,m}(x; U)$ and thus the value function. The problem is that there are uncountably many beliefs, making it impossible to directly solve the Bellman equation. In this section, we will provide an alternative procedure to compute the Whittle's index.

1) Preliminaries: We first prepare some tools. Define $\mathcal{T}^l(x) := \Pr\{S(1 + l) = 1 | X(1) = x\}$ as the l-step transition of a passive arm. It is easy to show that

$$\mathcal{T}^l(x) = x_0 - (x_0 - x)(p_{11} - p_{01})^l, \tag{19}$$

where $x_0 := p_{01}/(1 + p_{01} - p_{11})$ is the equilibrium distribution of the arm. As shown in [13], when $l \to \infty$, $\mathcal{T}^l(x)$ converges to x_0 monotonically if $p_{11} > p_{01}$ (called a *positively-correlated arm*), or alternatively from above and below if $p_{11} < p_{01}$ (called a *negatively-correlated arm*). Now define

$$\mathcal{L}(x;\, c_1,\, c_2) := \min\{l :\, \mathcal{T}^l(x) \in (c_1,\, c_2)\} \tag{20}$$

to be the hitting time of the interval (c_1, c_2) from initial state x under transition $\mathcal{T}^l$. We can calculate the hitting time in closed form as in [15].

2) Computing Value Function: Next, we derive a method to compute the value function directly. Due to the threshold structure of the optimal single-armed policy (Lemma 1), the belief evolution has the following pattern: starting from initial state x, it passively transits $\mathcal{L}(x;\, \tau^-(m),\, \tau^+(m)) =: L$ steps to $\mathcal{T}(x),\, \mathcal{T}^2(x), \ldots$ until hitting the active set $\mathcal{T}^L(x) \in (\tau^-(m),\, \tau^+(m))$, at which time the arm will be activated (sampled) and the state reset to p_{01} or p_{11}. Accordingly, we can decompose the total value into two parts: (i) passive rewards and subsidies for steps $0, \ldots, L-1$, and (ii) total future rewards from step L onward. Part (i) can be easily computed as $\frac{(1-\beta^L)m}{1-\beta} + f(x;\, L)$, where $f(x;\, L) := \sum_{i=0}^{L-1} \beta^i \max(\mathcal{T}^i(x),\, 1 - \mathcal{T}^i(x))$ can be computed in closed form as in [15]. Part (ii) is simply $\beta^L V_{\beta,m}(\mathcal{T}^L(x); U = 1)$. Therefore, given m, $\tau^-(m)$, and $\tau^+(m)$, we can compute the value function by

$$\begin{aligned} V_{\beta,m}(x) = &\frac{(1 - \beta^L)m}{1 - \beta} + f(x;\, L) + \beta^L \\ &+ \beta^{L+1}\mathcal{T}^L(x) V_{\beta,m}(p_{11}) + \beta^{L+1}(1 - \mathcal{T}^L(x)) V_{\beta,m}(p_{01}). \end{aligned} \tag{21}$$

The only unknowns left are $V_{\beta,m}(p_{11})$, $V_{\beta,m}(p_{01})$. Note that $x = p_{11}$ or p_{01} should also satisfy (21), giving us two equations with two unknowns. Solving these equations yields $V_{\beta,m}(p_{11})$, $V_{\beta,m}(p_{01})$; see [15] (equations (29–30)).

3) Computing Whittle's Index: We are now ready to compute the Whittle's index. First, we note that although seemingly complicated, the expression of value function given by (21) has a simple linear form in m for fixed thresholds.

Lemma 3: Given thresholds $\tau^-(m)$ and $\tau^+(m)$, for coefficients a_i, b_i $(i = 1,\, 2)$ as in [15] (equations (3–6)), define functions

$$\begin{aligned} a(x) \;&:=\; \frac{1 - \beta^{L(x)}}{1 - \beta} + \beta^{L(x)+1}\mathcal{T}^{L(x)}(x) a_2 \\ &\quad + \beta^{L(x)+1}(1 - \mathcal{T}^{L(x)}(x)) a_1, \end{aligned} \tag{22}$$

$$\begin{aligned} b(x) \;&:=\; f(x;\, L(x)) + \beta^{L(x)} + \beta^{L(x)+1}\mathcal{T}^{L(x)}(x) b_2 \\ &\quad + \beta^{L(x)+1}(1 - \mathcal{T}^{L(x)}(x)) b_1, \end{aligned} \tag{23}$$

where $L(x) := \mathcal{L}(x;\, \tau^-(m),\, \tau^+(m))$. Then the value function is equal to $V_{\beta,m}(x) = a(x)m + b(x)$, with end values $V_{\beta,m}(p_{01}) = a_1 m + b_1$, and $V_{\beta,m}(p_{11}) = a_2 m + b_2$.

Proof: See [15]. ■

Remarks: The above result reveals the following meanings of $a(x)$ and $b(x)$: $a(x)$ is the total discounted time for the arm to be passive (under the optimal single-armed policy), as any change in subsidy m results in an $a(x)$-fold change in the total value; $b(x)$ is the actual reward (excluding subsidies), i.e., the discounted total number of times the arm is tracked correctly.

Based on Lemma 3, we can easily compute the subsidy m for given $\tau^-(m)$ and $\tau^+(m)$. By definition, for $x = \tau^-(m)$ or $\tau^+(m)$, $V_{\beta,m}(x;\, u = 0) = V_{\beta,m}(x;\, u = 1)$, i.e.,

$$\begin{aligned} &1 + \beta x V_{\beta,m}(p_{11}) + \beta(1 - x) V_{\beta,m}(p_{01}) \\ &\qquad = m + \max(x,\, 1 - x) + \beta V_{\beta,m}(\mathcal{T}(x)). \end{aligned}$$

Substituting the linear forms of $V_{\beta,m}(\cdot)$ into the above gives

$$\begin{aligned} m &= \frac{1 + \beta x b_2 + \beta(1 - x) b_1 - \max(x, 1 - x) - \beta b(\mathcal{T}(x))}{1 + \beta a(\mathcal{T}(x)) - \beta x a_2 - \beta(1 - x) a_1} \\ &=: \mathcal{M}(x;\, \tau^-(m),\, \tau^+(m)), \quad x = \tau^-(m),\, \tau^+(m). \end{aligned} \tag{24}$$

Given state x, let x' be the other threshold under subsidy $m = W(x)$ (x is also a threshold). We then have $W(x) = \mathcal{M}(x;\, \min(x,\, x'),\, \max(x,\, x')) = \mathcal{M}(x';\, \min(x,\, x'),\, \max(x,\, x'))$, where the only unknown is x'. We can solve this equation (numerically) to compute x' and hence $W(x)$. The above gives a procedure to compute the Whittle's index as in lines (4–5) of Algorithm 1, which also gives the corresponding index policy.

E. Upper Bound under Relaxed Link Sampling Constraint

A nice property of the Whittle's index is that it is closely related to a performance upper bound. Specifically, consider a relaxed sampling constraint that only requires the discounted average number of sampled links to be bounded by K. The set of valid policies under the relaxed constraint is given by

$$\widetilde{\Pi}_{E,K} := \{\pi : \mathbb{E}_\pi[(1-\beta) \sum_{t=1}^{\infty} \beta^{t-1} \sum_{e \in E} I_{U_e(t)=1}] \leq K\}. \tag{25}$$

Algorithm 1 Whittle's Index Policy for Link Sampling

Require: Link parameters p_{01}^e, p_{11}^e for $e \in E$, current beliefs $\mathbf{x}(t) = \{x_e(t)\}_{e\in E}$, discount factor β, link sample size K.

Ensure: Return a link set E^{π_t} to sample at time step t.

1: $\mathbf{W}(t) \leftarrow (W_e(x_e(t)))_{e\in E}$ {compute Whittle's indices}
2: $E^{\pi_t} \leftarrow \{e : W_e(t)$is among the K largest in $\mathbf{W}(t)\}$
3: Compute $\mathbf{x}(t+1)$ from $\mathbf{x}(t)$ and $\mathbf{z}(t)$ by (7)

Compute Whittle's index:

4: $\tau \leftarrow$ the other solution to $\mathcal{M}(x; \min(x,\tau), \max(x,\tau)) = \mathcal{M}(\tau; \min(x, \tau), \max(x, \tau))$ {x is also a solution}
5: $W(x) \leftarrow \mathcal{M}(x; \min(x, \tau), \max(x, \tau))$

Let $\widetilde{\pi}^*_{E,K}$ be the optimal policy in $\widetilde{\Pi}_{E,K}$. Recall that $\pi^*_{E,K}$ is the optimal policy under the strict constraint. Let $\pi^W_{E,K}$ and $\pi^Y_{E,K}$ denote the Whittle's index policy and the myopic policy under the strict constraint. The following relationship holds.

Proposition 2 (Upper Bound on Reward [1]): For the discounted total reward defined in (4), we have

$$R_\infty(\pi^Y_{E,K}),\ R_\infty(\pi^W_{E,K}) \le R_\infty(\pi^*_{E,K}) \le R_\infty(\widetilde{\pi}^*_{E,K}). \quad (26)$$

It turns out ([1]) that the optimal policy $\widetilde{\pi}^*_{E,K}$ is also based on the Whittle's index, but instead of sampling exactly K links with the largest indices, it samples all the links whose Whittle's indices exceed a certain threshold $m^*(K)$ such that the discounted average sample size is equal to K. By LP duality, it is known ([13]) that $m^*(K)$ is the value of m achieving the following minimization:

$$R_\infty(\widetilde{\pi}^*_{E,K}) = \inf_m \{\sum_{e\in E} V^{(e)}_{\beta,m}(X_e(1)) - \frac{m(M-K)}{1-\beta}\}, \quad (27)$$

and the minimum value gives the upper bound. Here $\mathbf{X}(1)$ is the initial state and $V^{(e)}_{\beta,m}(x)$ the value function given by (21) for link e. Solving (27) numerically will give the upper bound (the solution is well defined since $V^{(e)}_{\beta,m}(x)$ is convex in m, as shown in Fig. 4 of [15]).

V. Network Tracking under Node Sampling

We have so far considered the scenario where we can choose the sampled links independently. We now consider an alternative scenario where the sampling is based on nodes. This scenario is more challenging because the sampling of adjacent links is now coupled, where the coupling is governed by the topology of the base network. Under the new constraint, the problem is no longer an RMB, and the original Whittle's index policy does not apply. We will, however, show that there is a natural way to extend Whittle's solution via a linear programming (LP) relaxation analogous to [1].

A. LP Formulation under Relaxed Constraint

We first solve a relaxed version of the node sampling problem by formulating an LP. We introduce the following notations. Let $\widetilde{\Pi}_{V,K}$ denote the set of valid node sampling policies under relaxed constraint similar to (25) and $\widetilde{\pi}^*_{V,K}$ the associated optimal policy. For a given policy $\pi \in \widetilde{\Pi}_{V,K}$, let $T^U_e(x)$ denote the total discounted time that link e spends under state x and action U ($U=0$ if unsampled and $U=1$ if sampled), i.e., (note everything implicitly depends on the initial state $\mathbf{X}(1)$)

$$T^1_e(x) := \mathbb{E}_\pi[\sum_{t=1}^{\infty} \beta^{t-1} I_{U_e(t)=1\cap X_e(t)=x} | X_e(1)],$$

and similarly for $T^0_e(x)$. Let $\mathcal{X}_e$ denote its state space[4]. Similarly, $T^U_v(x')$ denotes the total discounted time that node v spends under state x' and action U, where the node state $X'_v(t){:=}(X_e(t)\ :\ e\in\mathcal{N}(v))$ is the tuple of states at its adjacent links. Let $\mathcal{X}'_v$ denote the associated state space.

We now formulate the LP for the node sampling problem:

$$\max \quad \sum_{e\in E}\sum_{x\in\mathcal{X}_e} (R^0_e(x)T^0_e(x) + R^1_e(x)T^1_e(x)) \quad (28)$$

$$\text{s.t.} \sum_{x\in\mathcal{X}_e} T^1_e(x) \le \sum_{x'\in\mathcal{X}'_u} T^1_u(x') + \sum_{x'\in\mathcal{X}'_v} T^1_v(x'),\ \forall e=(u,v)\in E, (29)$$

$$\sum_{v\in V}\sum_{x'\in\mathcal{X}'_v} T^1_v(x') = \frac{K}{1-\beta},\ \forall v\in V, \quad (30)$$

$$(T^0_e(x), T^1_e(x))_{x\in\mathcal{X}_e} \in \mathcal{F}_e, \quad \forall e\in E, \quad (31)$$

$$(T^0_v(x'), T^1_v(x'))_{x'\in\mathcal{X}'_v} \in \mathcal{F}_v, \quad \forall v\in V. \quad (32)$$

Here (28) is the expected reward under discounting (4) (recall that $R^U_e(x)$ is the one-step reward defined in (12)); (29) is the graph constraint, saying that the time a link is sampled cannot be larger than the total time that either of its endpoints is sampled; (30) is the relaxed sampling constraint saying that only K nodes are sampled on the (discounted) average; (31–32) are regularity constraints that ensure the values of variables are valid and satisfy the Markov chains. Specifically, analogous to [10], we have

$$\mathcal{F}_e := \{(T^0_e(x),\ T^1_e(x))_x :\ \text{(i)}\ T^U_e(x) \ge 0, \quad (33)$$
$$\text{(ii)} \sum_{U,\,x} T^U_e(x) = \frac{1}{1-\beta},$$
$$\text{(iii)} \sum_{U=0,\,1} \big(T^U_e(x) - \beta \sum_{y\in\mathcal{X}_e} P^U_e(y,x)T^U_e(y)\big) = I_{X_e(1)=x}\},$$

$$\mathcal{F}_v := \{(T^0_v(x'),\ T^1_v(x'))_{x'} : \text{(i)}\ T^U_v(x') \ge 0, \quad (34)$$
$$\text{(ii)} \sum_{U,\,x'} T^U_v(x') = \frac{1}{1-\beta}\},$$

where $P^U_e(y,x)$ is the belief transition probability as in (11). We can write the Lagrangian of (28) as

$$L(\mathbf{m},\ \gamma) = \max_{\substack{\mathbf{T}_e\in\mathcal{F}_e,\\ e\in E}} \sum_{e,\,x} \big((R^0_e(x)+m_e)T^0_e(x) + R^1_e(x)T^1_e(x)\big)$$
$$+ \max_{\substack{\mathbf{T}_v\in\mathcal{F}_v,\\ v\in V}} \sum_{v,\,x'} (\gamma - \sum_{e\in\mathcal{N}(v)} m_e)T^0_v(x') + \frac{\sum_e m_e - (N-K)\gamma}{1-\beta}. \ (35)$$

[4]Although the entire belief space is uncountable, conditioned on an initial state $X_e(1)$, only a countable subset is reachable. Denote this subset as $\mathcal{X}_e$.

Algorithm 2 Extended Whittle's Policy for Node Sampling

Require: Base network $G = (V, E)$, parameters p_{01}^e, p_{11}^e for $e \in E$, current belief $\mathbf{x}(t) = \{x_e(t)\}_{e \in E}$, discount factor β, node sample size K.

Ensure: Return a node set V^{π_t} to sample at timestep t.

1: $\mathbf{W}(t) \leftarrow (W_e(x_e(t)))_{e \in E}$ {compute Whittle's indices as in Algorithm 1}

2: $V^{\pi_t} \leftarrow \arg\max_{V' \subseteq V,\ |V'|=K} \sum_{e \in \bigcup_{v \in V'} \mathcal{N}(v)} W_e(t)$ {compute max. coverage}

3: Compute $\mathbf{x}(t+1)$ from $\mathbf{x}(t)$ and $\mathbf{z}(t)$ by (7)

Compute greedy coverage (alternative to line 2):

4: Initialize $G' \leftarrow G$, $V^{\pi_t} \leftarrow \emptyset$

5: **for** $k = 1, 2, \ldots, K$ **do**

6: $\quad v_*(G') := \arg\max_{v \in V'} \sum_{e \in \mathcal{N}(v; G')} W_e(t)$

7: $\quad V^{\pi_t} \leftarrow V^{\pi_t} \cup v_*(G')$ {select the node with the largest sum index in G'}

8: $\quad G' \leftarrow G' \setminus \{v_*(G')\}$ {Remove selected node from G'}

Remarks: The dual variables m_e, γ have intuitive explanations. From (35), we can view m_e as the subsidy link e receives for being passive, and γ is the subsidy any node gets for being passive, i.e., keeping all its links passive by paying link e a subsidy m_e. Each link (node) wants to activate/not activate itself to maximize its income. Unlike Whittle's original solution, here the link subsidy m_e can vary for different links, whereas the node subsidy γ is the same for all the nodes.

B. Extension of Whittle's Index Policy

The above derivation provides a natural heuristic for node sampling. The LP relaxation (35) suggests that we should sample a node whenever the sum of its link subsidies exceeds its subsidy. As the link optimization (over $\mathbf{T}_e$) is the same as that in link sampling, we can use the Whittle's indices on links to represent link subsidies. Under strict sampling constraint, this translates to sampling the K nodes with the largest sum of the Whittle's indices on the covered links, which leads to a *maximum coverage problem* [16]. The extended policy is summarized in Algorithm 2. Finding the maximum coverage in a general graph is NP-hard. Fortunately, there is a greedy approximation, which selects nodes iteratively such that in each iteration, the node maximizing the sum index of newly covered links is chosen, as shown in lines 4–8 of Algorithm 2. This greedy algorithm has an approximation ratio of $\frac{\mathrm{e}}{\mathrm{e}-1}$ (e is the base of natural logarithm) [16].

Remarks: The above policy is also applicable to arbitrary link indices. In particular, we can replace the Whittle's index $W_e(x_e(t))$ by the myopic index $Y_e(x_e(t))$ defined in (13), the result being a myopic policy for node sampling.

C. Upper Bound under Relaxed Node Sampling Constraint

We now provide an upper bound for node sampling by solving for the optimal reward under relaxed constraint $R_\infty(\widetilde{\pi}_{V,K}^*)$ using (35). By weak duality, we have $L(\mathbf{m}, \gamma) \geq R_\infty(\widetilde{\pi}_{V,K}^*)$; by strong duality, there must exist $\mathbf{m}^*, \gamma^*$ such that

$$L(\mathbf{m}^*, \gamma^*) = \min_{\mathbf{m},\, \gamma} L(\mathbf{m}, \gamma) = R_\infty(\widetilde{\pi}_{V,K}^*). \tag{36}$$

It is easy to see that the optimization over $\mathbf{T}_v$ in (35) is achieved at $\sum_{x'} T_v^0(x') = 1/(1-\beta)$ if $\gamma > \sum_{e \in \mathcal{N}(v)} m_e$ and $\sum_{x'} T_v^0(x') = 0$ otherwise. Moreover, the optimization over $\mathbf{T}_e$ is the same as that in the single-armed bandit, yielding $\max_{\mathbf{T}_e \in \mathcal{F}_e} \sum_x \left((R_e^0(x) + m_e) T_e^0(x) + R_e^1(x) T_e^1(x) \right) = V_{\beta, m_e}^{(e)}(X_e(1))$. Substituting these into (35, 36) gives

$$R_\infty(\widetilde{\pi}_{V,K}^*) = \min_{\mathbf{m} \geq 0} \Bigg[\sum_{e \in E} V_{\beta, m_e}^{(e)}(X_e(1)) + \frac{\sum_{e \in E} m_e}{1-\beta} + \min_{\gamma} \Big(\sum_{v \in V} \frac{\max(\gamma - \sum_{e \in \mathcal{N}(v)} m_e,\ 0)}{1-\beta} - \frac{(N-K)\gamma}{1-\beta} \Big) \Bigg].$$

Given $\mathbf{m}$, it can be shown that the optimal γ is $\gamma^*(\mathbf{m}) := (\sum_{e \in \mathcal{N}(v)} m_e)^{(N-K)}$, where $(\sum_{e \in \mathcal{N}(v)} m_e)^{(i)}$ is the ith smallest sum subsidy per node. Thus,

$$R_\infty(\widetilde{\pi}_{V,K}^*) = \min_{\mathbf{m} \geq 0} \Bigg[\sum_{e \in E} V_{\beta, m_e}^{(e)}(X_e(1)) + \frac{\sum_{e \in E} m_e}{1-\beta} - \frac{\sum_{i=1}^{N-K} (\sum_{e \in \mathcal{N}(v)} m_e)^{(i)}}{1-\beta} \Bigg]. \tag{37}$$

Since $R_\infty(\pi_{V,K}^*) \leq R_\infty(\widetilde{\pi}_{V,K}^*)$, solving (37) gives us a performance upper bound under node sampling. Unfortunately, unlike the bound for link sampling (27), the above bound is difficult to evaluate as we now face a multi-variate optimization over m_e's for all the links. Nevertheless, for each given $\mathbf{m} \geq 0$, the right-hand side of (37) gives an upper bound on $R_\infty(\widetilde{\pi}_{V,K}^*)$ and hence on $R_\infty(\pi_{V,K}^*)$.

VI. Numerical Studies

We now evaluate the proposed policies through simulations.

A. Verification of Analysis

We first verify our calculation of the Whittle's index and the value function. Fig. 3 plots the single-armed value functions $V_{\beta,m}(x;\ u=0)$, $V_{\beta,m}(x;\ u=1)$ in (15–16) under two subsidies: (i) $m = m_{\max} := \max_{x \in [0,1]} W(x)$, achieved at τ^*, under which $V_{\beta,m}(x;\ u=0)$ and $V_{\beta,m}(x;\ u=1)$ should be tangent at τ^*, and (ii) $m = W(x')$, under which the value functions should intersect at x' (and one other point). The plot has verified the above. It also verifies the threshold structure of the optimal single-armed policy as in Lemma 1. Note that the thresholds $\tau^-(m)$, $\tau^+(m)$ may not be symmetric around 0.5.

Next, we verify our conjecture of indexability for $\beta > 0.5$. By Lemma 1, it suffices to verify the monotonicity of $\tau^-(m)$ and $\tau^+(m)$. As shown in Fig. 4, we divide possible scenarios for positively-correlated arms into four cases: (a) $p_{01} < p_{11} \leq 0.5$, (b) $0.5 \leq p_{01} < p_{11}$, (c) $p_{01} < x_0 \leq 0.5 < p_{11}$, and (d) $p_{01} < 0.5 < x_0 < p_{11}$. The plot shows that $\tau^-(m)$ is monotone increasing and $\tau^+(m)$ monotone decreasing with m in all cases. Similar result holds under negative correlation.

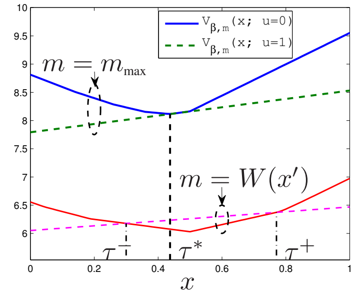


Fig. 3. Value functions $V_{\beta,m}(x;\ u = 0)$, $V_{\beta,m}(x;\ u = 1)$ ($\beta = 0.8$, $p_{01} = 0.15$, $p_{11} = 0.95$, $x' = 0.3$): (i) for $m = m_{\max} \approx 1.02$, $\tau^* \approx 0.44$; (ii) for $m = W(x') \approx 0.51$, $\tau^- = 0.3$, $\tau^+ \approx 0.77$.

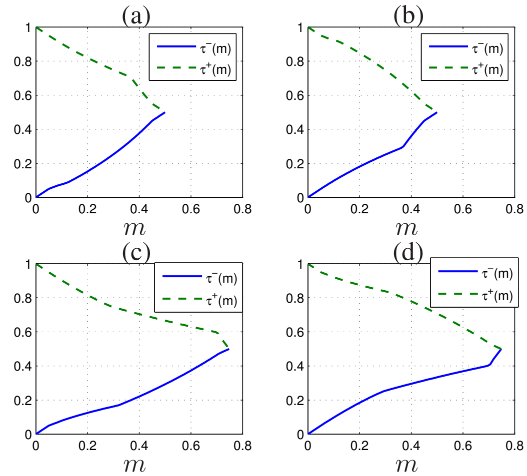


Fig. 4. Thresholds $\tau^-(m)$ and $\tau^+(m)$ vs. m ($\beta = 0.8$): (a) $p_{01} = 0.05$, $p_{11} = 0.45$; (b) $p_{01} = 0.55$, $p_{11} = 0.95$; (c) $p_{01} = 0.05$, $p_{11} = 0.75$; (d) $p_{01} = 0.25$, $p_{11} = 0.95$.

B. Comparison of Myopic & Whittle's Indices

Under a similar model, [13] shows that if the goal is to access arms in a certain state, then the Whittle's index policy coincides with the myopic policy for i.i.d. arms. In sharp contrast, we find that for the goal of tracking the states of all arms (as in our problem), the two policies are generally different. As shown in Fig. 5, the indices used by the two policies differ not only in the absolute value, but also in the relative value. In particular, the myopic index is symmetric around 0.5, whereas the Whittle's index can be asymmetric. A closer look reveals: (i) the Whittle's index is larger than the myopic index, and tends to be relatively larger around the equilibrium state x_0, and (ii) the difference between the two indices tends to increase with $|p_{11} - p_{01}|$, both observations consistent with Proposition 1. In particular, $|p_{01} - p_{11}|$ indicates how different the arm is from an i.i.d. process, and since the Whittle's index takes the temporal correlation into account while myopic index does not, they deviate more when this correlation gets stronger. In the special case that p_{01}, p_{11} are symmetric around 0.5 (i.e., $p_{01} + p_{11} = 1$), the Whittle's index is also symmetric (not shown). Recalling from (10) that all states (except for the initial one) must lie between p_{01} and p_{11}, we can draw the following conclusions: for i.i.d. arms, the Whittle's index policy is equivalent to the myopic policy if (1) p_{01}, p_{11} are on the same side of 0.5, or (2) p_{01}, p_{11} are symmetric about 0.5.

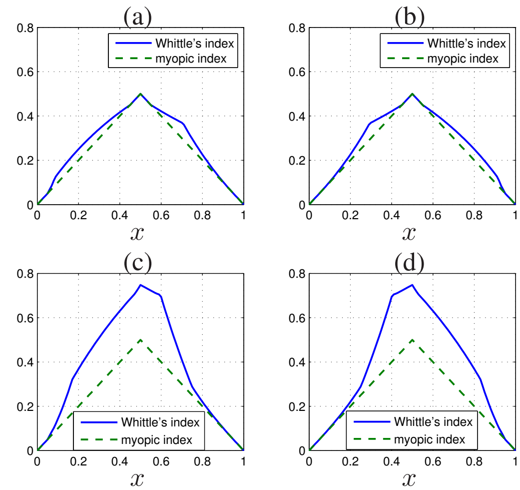
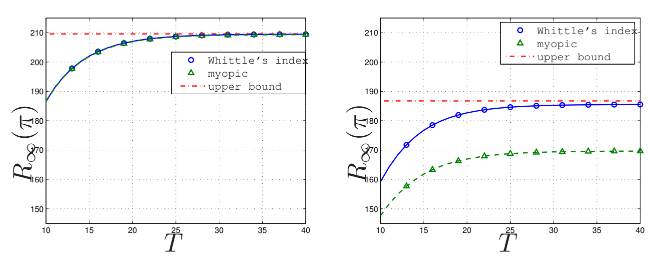


Fig. 5. Indices used by the Whittle's index policy and the myopic policy (same parameters as in Fig. 4).

C. Comparison of Tracking Accuracy

1) Link Sampling: Surprisingly, although the two indices behave differently, the corresponding policies end up with the same performance in terms of (4) if the links are i.i.d., as illustrated in Fig. 6 (a). The Whittle's index policy starts to show improvement as the links diverge, modeled by dividing the links into equal parts of fast-varying and slow-varying links. In this case, as shown in Fig. 6 (b), the gap between the two policies can be significant ($\approx 10\%$). For comparison, we also plot the upper bound given by (27). The Whittle's index policy always approaches the bound closely, whereas the myopic policy falls below for non-i.i.d. links (we have shown a case of extreme heterogeneity, but clear gaps are also observed in other cases).



(a) i.i.d. links (b) non-i.i.d. links

Fig. 6. Link sampling ($\beta = 0.8$, $M = 60$, $K = 3$, $T = 40$, $\mathbf{x}(1) = \mathbf{x}_0$, 100 Monte Carlo runs): (a) $p_{01} = 0.2$, $p_{11} = 0.9$; (b) $p_{01} = 0.999$ for fast links and 0.001 for slow links, $p_{11} = 1 - p_{01}$.

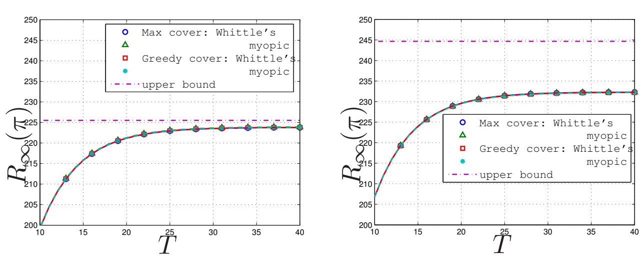
2) Node Sampling: We simulate four policies based on Algorithm 2: the optimal maximum coverage vs. the greedy coverage, based on the Whittle's index vs. the myopic index. For comparison, we also plot upper bounds via a variation of (37). As (37) is difficult to evaluate, we set $m_e \equiv m$ for all $e \in E$ and optimize over m:

$$R'_\infty(\tilde{\pi}^*_{V,K}) := \min_{m \geq 0} \Bigg[\sum_{e \in E} V^{(e)}_{\beta,m}(X_e(1)) + \frac{M}{1-\beta} m - \frac{\sum_{i=1}^{N-K} |\mathcal{N}(v)|^{(i)}}{1-\beta} m \Bigg], \tag{38}$$

where $|\mathcal{N}(v)|^{(i)}$ is the ith smallest node degree in the base network. It is easy to see that $R'_\infty(\tilde{\pi}^*_{V,K}) \geq R_\infty(\tilde{\pi}^*_{V,K}) \geq R_\infty(\pi^*_{V,K})$. Moreover, since node sampling depends on the topology of the base network, we simulate two topologies with different levels of regularity: a lattice graph and a random graph, where the latter has the same size (N and M) as the former but randomly distributed (base) links.

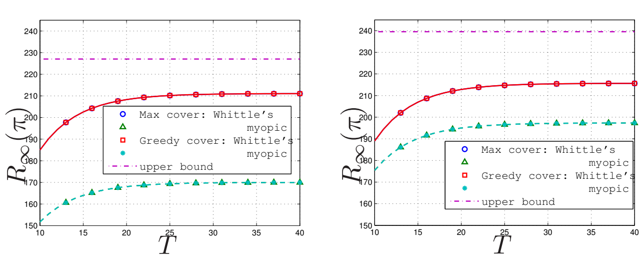
For i.i.d. links (Fig. 7), there is no visible difference between the extended Whittle's index policy and the myopic policy as in the link sampling case (Fig. 6 (a)). Moreover, the greedy algorithm ("Greedy cover") performs almost as well as the optimal algorithm ("Max cover"). Between the two base topologies, we see that the gap with the relaxed upper bound in (38) is much smaller for the lattice graph. We then simulate the non-i.i.d. case by selecting half of the links as fast-varying links and the rest as slow-varying links, keeping links of the same type close to each other[5]. As shown in Fig. 8, we see similar behavior as before, except that the extended Whittle's policy now outperforms the myopic policy significantly (by $\approx 24\%$ for lattice and $\approx 10\%$ for random graph), and there is always a gap ($< 10\%$) with the upper bound.

Combining the observations in Section VI-C1–VI-C2, we conclude that the Whittle's index policy and the myopic policy perform equivalently if the links are i.i.d. For non-i.i.d. links, the Whittle's index policy can be significantly better (by $> 20\%$), with near-optimal performance under link sampling, but possibly suboptimal performance under node sampling. Note that the suboptimality is not certain as the bound in (38) is further relaxed and may not be achievable. Similar observations have also been made under the undiscounted reward [15].



(a) lattice (b) random graph

Fig. 7. Node sampling: i.i.d. links ($\beta = 0.8$, $N = 36$, $M = 60$, $K = 2$, $T = 40$, $p_{01} = 0.2$, $p_{11} = 0.9$, $\mathbf{x}(1) = \mathbf{x}_0$, 100 Monte Carlo runs).



(a) lattice (b) random graph

Fig. 8. Node sampling: non-i.i.d. links ($p_{01} = 0.5$ for fast links and 0.001 for slow links, $p_{11} = 1 - p_{01}$, rest as in Fig. 7).

[5]Otherwise, the heterogeneity will cancel out at node level and the gap between the policies will be smaller (not shown).

VII. Conclusion

We have developed efficient sampling policies based on the Whittle's indices for tracking the topology of dynamic networks under sampling constraints. For heterogeneous links, the proposed policies achieve significantly higher tracking accuracy than the myopic policy that always samples links/nodes with the highest uncertainty, especially under regular base network topology. Although our discussion is limited to link or node sampling, the idea of maximum coverage using the Whittle's indices as link weights is applicable to more general scenarios of subgraph sampling. In this work, we have ignored the correlation among links to focus on their temporal evolution. Although link correlation can be modeled by extending the per-link models to a joint model of all links (e.g., an M-dimensional Markov chain), a straightforward extension (via Partially Observable Markov Decision Process) may not allow for efficient, (near) optimal sampling policies, and it remains open to what extent one can capture link correlation and still have tractable solutions. Our work provides a concrete foundation for tackling these problems.

References

[1] P. Whittle, "Restless Bandits: Activity Allocation in a Changing World," *Journal of Applied Probability*, vol. 25, 1988.

[2] C. Papadimitriou and J. Tsitsiklis, "The Complexity of Optimal Queueing Network Control," *Mathematics of Operations Research*, no. 2, pp. 293–305, May 1999.

[3] M. Granovetter, "Network Sampling: Some First Steps," *American Journal of Sociology*, vol. 81, no. 6, pp. 1287–1303, 1976.

[4] R. Govindan and H. Tangmunarunkit, "Heuristics for Internet Map Discovery," in *IEEE INFOCOM*, Tel-Aviv, Israel, June 2000.

[5] B. Eriksson, P. Barford, R. Nowak, and M. Crovella, "Learning Network Structure from Passive Measurements," in *Proc. of the ACM SIGCOMM conference on Internet measurement*, Kyoto, Japan, Aug. 2007.

[6] Z. Beerliova, F. Eberhard, T. Erlebach, A. Hall, M. Hoffmann, M. Mihal ak, and L. Ram, "Network Discovery and Verification," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 12, p. 2168, 2006.

[7] L. Reyzin and N. Srivastava, "Learning and verifying graphs using queries with a focus on edge counting," *Lecture Notes in Computer Science*, vol. 4754, p. 285, 2007.

[8] S. Choi and J. Kim, "Optimal query complexity bounds for finding graphs," in *Proc. of annual ACM symposium on Theory of computing*, 2008, pp. 749–758.

[9] H. Tong, S. Papadimitriou, P. Yu, and C. Faloutsos, "Proximity tracking on time-evolving bipartite graphs," in *Proc. of the SIAM Intl. Conf. on Data Mining*, Atlanta, USA, 2008.

[10] J. Nino-Mora, "Restless bandits, partial conservation laws and indexability," *Advances in Applied Probability*, vol. 33, no. 1, pp. 76–98, 2001.

[11] R. Weber and G. Weiss, "On an Index Policy for Restless Bandits," *Journal of Applied Probability*, vol. 27, no. 3, pp. 637–648, Sept 1990.

[12] J. L. Ny, M. Dahleh, and E. Feron, ""Multi-UAV Dynamic Routing with Partial Observations Using Restless Bandit Allocation Indices"," in *Proc. American Control Conference*, Seattle, WA, June 2008.

[13] K. Liu and Q. Zhao, "Indexability of Restless Bandit Poblems and Optimality of Whittle's Index for Dynamic Multichannel Access," *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5547–5567, November 2010.

[14] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ 07632: Prentice Hall, 1993.

[15] T. He, A. Anandkumar, and D. Agrawal, "Tracking Dynamic Networks under Sampling Constraints: Supporting Materials," IBM, Tech. Rep. RC25025, July 2010, https://researcher.ibm.com/researcher/files/us-the/rc25025.pdf.

[16] D. Hochbaum, *Approximation algorithms for NP-hard problems*. PWS Publishing Co. Boston, MA, USA, 1996.