SIAM J. CONTROL OPTIM.
Vol. 0, No. 0, pp. S49–S74

PARAMETER ESTIMATION IN EPIDEMIC SPREAD NETWORKS USING LIMITED MEASUREMENTS*

LINTAO YE†, PHILIP E. PARÉ‡, AND SHREYAS SUNDARAM‡

Abstract. We study the problem of estimating the parameters (i.e., infection rate and recovery rate) governing the spread of epidemics in networks. Such parameters are typically estimated by measuring various characteristics (such as the number of infected and recovered individuals) of the infected populations over time. However, these measurements also incur certain costs, depending on the population being tested and the times at which the tests are administered. We thus formulate the epidemic parameter estimation problem as an optimization problem, where the goal is to either minimize the total cost spent on collecting measurements or to optimize the parameter estimates while remaining within a measurement budget. We show that these problems are NP-hard to solve in general and then propose approximation algorithms with performance guarantees. We validate our algorithms using numerical examples.

Key words. epidemic spread networks, parameter estimation, optimization algorithms

AMS subject classifications. 93C10, 93E10, 68Q25, 68W25

DOI. 10.1137/20M1377801

1. Introduction.

Models of spreading processes over networks have been widely studied by researchers from different fields (see, e.g., [15, 24, 6, 3, 26, 21]). The case of epidemics spreading through networked populations has received a particularly significant amount of attention, especially in light of the ongoing COVID-19 pandemic (see, e.g., [21, 22]). A canonical example is the networked SIR model, where each node in the network represents a subpopulation or an individual and can be in one of three states: susceptible (S), infected (I), or recovered (R) [19]. There are two key parameters that govern such models: the infection rate of a given node and the recovery rate of that node. In the case of a novel virus, these parameters may not be known a priori, and must be identified or estimated from gathered data, including, for instance, the number of infected and recovered individuals in the network at certain points of time. For instance, in the COVID-19 pandemic, when collecting the data on the number of infected individuals or the number of recovered individuals in the network, one possibility is to perform virus or antibody tests on the individuals, with each test incurring a cost. Therefore, in the problem of parameter estimation in epidemic spread networks, it is important and of practical interest to take the costs of collecting the data (i.e., measurements) into account, which leads to the problem formulations considered in this paper. The goal is to exactly identify (when possible) or estimate the parameters in the networked SIR model using a limited number of measurements. Specifically, we divide our analysis into two scenarios: (1) when the measurements (e.g., the number of infected individuals) can be collected exactly without error and (2) when only stochastic measurements can be obtained.

*Received by the editors November 2, 2020; accepted for publication (in revised form) August 23, 2021; published electronically November 30, 2021.
https://doi.org/10.1137/20M1377801
Funding: This work was supported in part by the National Science Foundation under grants NSF-CMMI 1635014 and NSF-ECCS 2032258.
†Department of Electrical Engineering, The University of Notre Dame, Notre Dame, IN 46556 USA (lye2@nd.edu).
‡School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (philpare@purdue.edu, sundara2@purdue.edu).

In settings where exact measurements of the infected and recovered proportions of the population at certain nodes in the network can be obtained, we formulate the parameter identification measurement selection (PIMS) problem as minimizing the cost spent on collecting the measurements, while ensuring that the parameters of the SIR model can be uniquely identified (within a certain time interval in the epidemic dynamics). In settings where the measurements are stochastic (thereby precluding exact identification of the parameters), we formulate the parameter estimation measurement selection (PEMS) problem. The goal is to optimize certain estimation metrics, while satisfying the budget on collecting the measurements.

Related work. The authors in [23, 35] studied the parameter estimation problem in epidemic spread networks using a "susceptible-infected-susceptible (SIS)" model of epidemics. When exact measurements of the infected proportion of the population at each node of the network can be obtained, the authors proposed a sufficient and necessary condition on the set of the collected measurements such that the parameters of the SIS model (i.e., the infection rate and the recovery rate) can be uniquely identified. However, this condition does not pose any constraint on the number of measurements that can be collected.

In [25], the authors considered a measurement selection problem in the SIR model. Their problem is to perform a limited number of virus tests among the population such that the probability of undetected asymptotic cases is minimized. The transmission of the disease in the SIR model considered in [25] is characterized by a Bernoulli random variable which leads to a hidden Markov model for the SIR dynamics.

Finally, our work is also closely related to the sensor placement problem that has been studied for control systems (see, e.g., [20, 39, 37]), signal processing (see, e.g., [7, 38]), and machine learning (see, e.g., [18]). The goal of these problems is to optimize certain (problem-specific) performance metrics of the estimate based on the measurements of the placed sensors, while satisfying the sensor placement budget constraints.

Contributions. First, we show that the PIMS problem is NP-hard, which precludes polynomial-time algorithms for the PIMS problem (if P $\neq$ NP). By exploring structural properties of the PIMS problem, we provide a polynomial-time *approximation* algorithm which returns a solution that is within a certain approximation ratio of the optimal. The approximation ratio depends on the cost structure of the measurements and on the graph structure of the epidemic spread network. Next, we show that the PEMS problem is also NP-hard. In order to provide a polynomial-time approximation algorithm that solves the PEMS problem with performance guarantees, we first show that the PEMS problem can be transformed into the problem of maximizing a set function subject to a knapsack constraint. We then apply a greedy algorithm to the (transformed) PEMS problem and provide approximation guarantees for the greedy algorithm. Our analysis for the greedy algorithm also generalizes the results from the literature for maximizing a submodular set function under a knapsack constraint to nonsubmodular settings. We use numerical examples to validate the obtained performance bounds of the greedy algorithm and show that the greedy algorithm performs well in practice.

Notation and terminology. The sets of integers and real numbers are denoted as $\mathbb{Z}$ and $\mathbb{R}$, respectively. For a set $\mathcal{S}$, let $|\mathcal{S}|$ be its cardinality. For any $n \in \mathbb{Z}_{\geq 1}$, let $[n] \triangleq \{1, 2, \ldots, n\}$. Let $\mathbf{0}_{m\times n}$ be a zero matrix of dimension $m \times n$; the subscript will be dropped if the dimension can be inferred from the context. For a matrix $P \in \mathbb{R}^{n\times n}$,

let P^T, $\mathrm{tr}(P)$, and $\det(P)$ be its transpose, trace, and determinant, respectively. The eigenvalues of P are ordered such that $|\lambda_1(P)| \geq \cdots \geq |\lambda_n(P)|$. Let P_{ij} (or $(P)_{ij}$) denote the element in the ith row and jth column of P, and let $(P)_i$ denote the ith row of P. A positive semidefinite matrix $P \in \mathbb{R}^{n\times n}$ is denoted by $P \succeq \mathbf{0}$.

2. Model of epidemic spread network. Suppose a disease (or virus) is spreading over a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} \triangleq [n]$ is the set of n nodes, and $\mathcal{E}$ is the set of directed edges (and self-loops) that captures the interactions among the nodes in $\mathcal{V}$. Here, each node $i \in \mathcal{V}$ is considered to be a group (or population) of individuals (e.g., a city or a country). A directed edge from node i to node j, where $i \neq j$, is denoted by (i, j). For all $i \in \mathcal{V}$, denote $\mathcal{N}_i \triangleq \{j : (j, i) \in \mathcal{E}\}$ and $\bar{\mathcal{N}}_i \triangleq \{j : (j, i) \in \mathcal{E}\} \cup \{i\}$. For all $i \in \mathcal{V}$ and for all $k \in \mathbb{Z}_{\geq 0}$, we let $s_i[k]$, $x_i[k]$, and $r_i[k]$ represent the proportions of the population of node $i \in \mathcal{V}$ that are susceptible, infected, and recovered at time k, respectively. To describe the dynamics of the spread of the disease in $\mathcal{G}$, we will use the following discrete-time SIR model (see, e.g., [12]):

$$s_i[k+1] = s_i[k] - hs_i[k]\beta \sum_{j\in\bar{\mathcal{N}}_i} a_{ij}x_j[k], \tag{2.1a}$$

$$x_i[k+1] = (1-h\delta)x_i[k] + hs_i[k]\beta \sum_{j\in\bar{\mathcal{N}}_i} a_{ij}x_j[k], \tag{2.1b}$$

$$r_i[k+1] = r_i[k] + h\delta x_i[k], \tag{2.1c}$$

where $\beta \in \mathbb{R}_{\geq 0}$ is the infection rate of the disease, $\delta \in \mathbb{R}_{\geq 0}$ is the recovery rate of the disease, $h \in \mathbb{R}_{\geq 0}$ is the sampling parameter, and $a_{ij} \in \mathbb{R}_{\geq 0}$ is the weight associated with edge (j, i). Let $A \in \mathbb{R}^{n\times n}$ be a weight matrix, where $A_{ij} = a_{ij}$ for all $i, j \in \mathcal{V}$ such that $j \in \bar{\mathcal{N}}_i$, and $A_{ij} = 0$ otherwise. Denote $s[k] \triangleq \begin{bmatrix} s_1[k] & \cdots & s_n[k] \end{bmatrix}^T \in \mathbb{R}^n$, $x[k] \triangleq \begin{bmatrix} x_1[k] & \cdots & x_n[k] \end{bmatrix}^T \in \mathbb{R}^n$, and $r[k] \triangleq \begin{bmatrix} r_1[k] & \cdots & r_n[k] \end{bmatrix}^T \in \mathbb{R}^n$ for all $k \in \mathbb{Z}_{\geq 0}$. Throughout this paper, we assume that the weight matrix $A \in \mathbb{R}^{n\times n}$ and the sampling parameter $h \in \mathbb{R}_{\geq 0}$ are given.

3. Preliminaries. We make the following assumptions on the initial conditions $s[0]$, $x[0]$, and $r[0]$, and the parameters of the SIR model in (2.1) (see, e.g., [23, 12]).

Assumption 3.1. For all $i \in \mathcal{V}$, $s_i[0] \in (0, 1]$, $x_i[0] \in [0, 1)$, $r_i[0] = 0$, and $s_i[0] + x_i[0] = 1$.

Assumption 3.2. Assume that $h, \beta, \delta \in \mathbb{R}_{>0}$ with $h\delta < 1$. For all $i, j \in \mathcal{V}$ with $(j, i) \in \mathcal{E}$ and $i \neq j$, assume that $a_{ij} \in \mathbb{R}_{>0}$. For all $i \in \mathcal{V}$, $h\beta \sum_{j\in\mathcal{N}_i} a_{ij} < 1$.

DEFINITION 3.3. *Consider a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with $\mathcal{V} = [n]$. A directed path of length t from node i_0 to node i_t in $\mathcal{G}$ is a sequence of t directed edges $(i_0, i_1), \ldots, (i_{t-1}, i_t)$. For any distinct pair of nodes $i, j \in \mathcal{V}$ such that there exists a directed path from i to j, the distance from node i to node j, denoted as d_{ij}, is defined as the shortest length over all such paths.*

DEFINITION 3.4. *Define $\mathcal{S}_I \triangleq \{i : x_i[0] > 0, i \in \mathcal{V}\}$ and $\mathcal{S}_H \triangleq \{i : x_i[0] = 0, i \in \mathcal{V}\}$. For all $i \in \mathcal{S}_H$, define $d_i \triangleq \min_{j\in\mathcal{S}_I} d_{ji}$, where $d_i \geq 1$, and define $d_i \triangleq +\infty$ if there is no path from j to i for any $j \in \mathcal{S}_I$. For all $i \in \mathcal{S}_I$, define $d_i \triangleq 0$.*

Using arguments similar to those in [12], one can show that $s_i[k], x_i[k], r_i[k] \in [0, 1]$ with $s_i[k] + x_i[k] + r_i[k] = 1$ for all $i \in \mathcal{V}$ and for all $k \in \mathbb{Z}_{\geq 0}$ under Assumptions 3.1–3.2. Thus, given $x_i[k]$ and $r_i[k]$, we can obtain $s_i[k] = 1 - x_i[k] - r_i[k]$ for all $i \in \mathcal{V}$

and for all $k \in \mathbb{Z}_{\geq 0}$. We also have the following result that characterizes properties of $x_i[k]$ and $r_i[k]$ in the SIR model over $\mathcal{G}$ given by (2.1). The proof is omitted here in the interest of space and can be found in the extended version of this paper [36].

LEMMA 3.5. *Consider a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ with $\mathcal{V} = [n]$ and the SIR dynamics given by* (2.1)*. Suppose Assumptions* 3.1–3.2 *hold. Then, the following results hold for all $i \in \mathcal{V}$, where $k \in \mathbb{Z}_{\geq 0}$, and $\mathcal{S}_H$ and d_i are defined in Definition* 3.4*:*

(a) $s_i[k] > 0$ *for all* $k \geq 0$.
(b) *If* $d_i \neq +\infty$, *then* $x_i[k] = 0$ *for all* $k < d_i$, *and* $x_i[k] \in (0, 1)$ *for all* $k \geq d_i$.[1]
(c) *If* $d_i \neq +\infty$, *then* $r_i[k] = 0$ *for all* $k \leq d_i$, *and* $r_i[k] \in (0, 1)$ *for all* $k > d_i$.
(d) *If* $i \in \mathcal{S}_H$ *with* $d_i = +\infty$, *then* $x_i[k] = 0$ *and* $r_i[k] = 0$ *for all* $k \geq 0$.

4. Measurement selection problem in exact measurement setting. In this section, we assume that $\mathcal{S}_I, \mathcal{S}_H \subseteq \mathcal{V}$ defined in Definition 3.4 are known.

4.1. Problem formulation. Given exact measurements of $x_i[k]$ and $r_i[k]$ for a subset of nodes, our goal is to estimate (or uniquely identify, if possible) the unknown parameters β and δ. Here, we consider the scenario where collecting the measurement of $x_i[k]$ (resp., $r_i[k]$) at any node $i \in \mathcal{V}$ and at any time step $k \in \mathbb{Z}_{\geq 0}$ incurs a cost, denoted as $c_{k,i} \in \mathbb{R}_{\geq 0}$ (resp., $b_{k,i} \in \mathbb{R}_{\geq 0}$). Moreover, we can only collect the measurements of $x_i[k]$ and $r_i[k]$ for $k \in \{t_1, t_1+1, \ldots, t_2\}$, where $t_1, t_2 \in \mathbb{Z}_{\geq 0}$ are given with $t_2 > t_1$. Noting that Lemma 3.5 provides a (sufficient and necessary) condition under which $x_i[k] = 0$ (resp., $r_i[k] = 0$) holds, we see that one does not need to collect a measurement of $x_i[k]$ (resp., $r_i[k]$) if $x_i[k] = 0$ (resp., $r_i[k] = 0$) from Lemma 3.5. Given time steps $t_1, t_2 \in \mathbb{Z}_{\geq 0}$ with $t_2 > t_1$, we define a set

$$\mathcal{I}_{t_1:t_2} \triangleq \{\lambda_i[k] : k \in \{t_1, \ldots, t_2\}, i \in \mathcal{V}, \lambda_i[k] > 0, \lambda \in \{x, r\}\}, \tag{4.1}$$

which represents the set of all candidate measurements from time step t_1 to time step t_2. To proceed, we first use (2.1b)–(2.1c) to obtain

$$\begin{bmatrix} x[t_1+1] - x[t_1] \\ \vdots \\ x[t_2] - x[t_2-1] \\ r[t_1+1] - r[t_1] \\ \vdots \\ r[t_2] - r[t_2-1] \end{bmatrix} = h \begin{bmatrix} \Phi^x_{t_1:t_2-1} \\ \Phi^r_{t_1:t_2-1} \end{bmatrix} \begin{bmatrix} \beta \\ \delta \end{bmatrix}, \tag{4.2}$$

where $\Phi^x_{t_1:t_2-1} \triangleq \begin{bmatrix} (\Phi^x_{t_1})^T & \cdots & (\Phi^x_{t_2-1})^T \end{bmatrix}^T$ with

$$\Phi^x_k \triangleq \begin{bmatrix} s_1[k] \sum_{j \in \bar{\mathcal{N}}_1} a_{1j} x_j[k] & -x_1[k] \\ \vdots & \vdots \\ s_n[k] \sum_{j \in \bar{\mathcal{N}}_n} a_{nj} x_j[k] & -x_n[k] \end{bmatrix} \quad \forall k \in \{t_1, \ldots, t_2 - 1\}, \tag{4.3}$$

and $\Phi^r_{t_1:t_2-1} \triangleq \begin{bmatrix} (\Phi^r_{t_1})^T & \cdots & (\Phi^r_{t_2-1})^T \end{bmatrix}^T$ with

$$\Phi^r_k \triangleq \begin{bmatrix} 0 & x_1[k] \\ \vdots & \vdots \\ 0 & x_n[k] \end{bmatrix} \quad \forall k \in \{t_1, \ldots, t_2 - 1\}. \tag{4.4}$$

[1] Note that for the case when $d_i = 0$, i.e., $i \in \mathcal{S}_I$, part (b) implies $x_i[k] > 0$ for all $k \geq 0$.

We can then view (4.2) as a set of $2(t_2-t_1)n$ equations in β and δ. Noting that $s_i[k]$ for all $i \in \mathcal{V}$ can be obtained from $s_i[k] = 1-x_i[k]-r_i[k]$ as argued in section 3, we see that the coefficients in the set of equations in β and δ given by (4.2), i.e., the terms in (4.2) other than β and δ, can be determined given that $x[k]$ and $r[k]$ are known for all $k \in \{t_1, \dots, t_2\}$. Also note that given $x[k]$ and $r[k]$ for all $k \in \{t_1, \dots, t_2\}$, we can uniquely identify β and δ using (4.2) if and only if $\text{rank}(\begin{bmatrix}(\Phi^x_{t_1:t_2-1})^T & (\Phi^r_{t_1:t_2-1})^T\end{bmatrix}) = 2$.

Next, let $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ denote a measurement selection strategy, where $\mathcal{I}_{t_1:t_2}$ is given by (4.1). We will then consider identifying β and δ using measurements contained in $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$. To illustrate our analysis, given any $i \in \mathcal{V}$ and any $k \in \{t_1, \dots, t_2-1\}$, we first consider the following equation from (4.2):

$$x_i[k+1] - x_i[k] = h\begin{bmatrix} s_i[k]\sum_{w\in\bar{\mathcal{N}}_i} a_{iw}x_w[k] & -x_i[k]\end{bmatrix}\begin{bmatrix}\beta\\ \delta\end{bmatrix}, \tag{4.5}$$

where $s_i[k] = 1 - x_i[k] - r_i[k]$, and we index the equation in (4.2) corresponding to (4.5) as (k, i, x). Note that in order to use (4.5) in identifying β and δ, one needs to determine the coefficients (i.e., the terms other than β and δ) in the equation. Also note that in order to determine the coefficients in equation (k, i, x), one can use the measurements contained in $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ and use Lemma 3.5 to determine whether $x_i[k] = 0$ (resp., $r_i[k] = 0$) holds. Supposing $x_i[k+1] = 0$, we see from Lemma 3.5 and (2.1b) that $x_i[k] = 0$ and $s_i[k]\sum_{w\in\bar{\mathcal{N}}_i} a_{iw}x_w[k] = 0$, which makes equation (k, i, x) useless in identifying β and δ. Thus, in order to use equation (k, i, x) in identifying β and δ, we need $x_i[k+1] \in \mathcal{I}$ with $x_i[k+1] > 0$. Similarly, given any $i \in \mathcal{V}$ and any $k \in \{t_1, \dots, t_2-1\}$, we consider the following equation from (4.2):

$$r_i[k+1] - r_i[k] = h\begin{bmatrix}0 & x_i[k]\end{bmatrix}\begin{bmatrix}\beta\\ \delta\end{bmatrix}, \tag{4.6}$$

where we index the above equation as (k, i, r). Supposing $r_i[k+1] = 0$, we see from Lemma 3.5 and (2.1c) that $r_i[k] = x_i[k] = 0$, which makes equation (k, i, r) useless in identifying β and δ. Hence, in order to use equation (k, i, r) in identifying β and δ, we need to have $\{x_i[k], r_i[k+1]\} \subseteq \mathcal{I}$ with $x_i[k] > 0$ and $r_i[k+1] > 0$. More precisely, we observe that equation (k, i, r) can be used in identifying β and δ if and only if $\{x_i[k], r_i[k+1]\} \subseteq \mathcal{I}$, and $r_i[k] \in \mathcal{I}$ or $r_i[k] = 0$ (from Lemma 3.5).

In general, let us denote the following two coefficient matrices corresponding to equations (k, i, x) and (k, i, r) in (4.2), respectively:

$$\Phi^x_{k,i} \triangleq \begin{bmatrix} s_i[k]\sum_{j\in\mathcal{N}_i} a_{ij}x_j[k] & -x_i[k]\end{bmatrix}, \tag{4.7a}$$

$$\Phi^r_{k,i} \triangleq \begin{bmatrix}0 & x_i[k]\end{bmatrix} \tag{4.7b}$$

for all $k \in \{t_1, \dots, t_2-1\}$ and for all $i \in \mathcal{V}$. Moreover, given any measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$, we let

$$\begin{aligned}\bar{\mathcal{I}} \triangleq\ & \{(k,i,x) : x_i[k+1] \in \mathcal{I}, x_i[k] = 0\} \cup \{(k,i,x) : \{x_i[k+1], x_i[k]\} \subseteq \mathcal{I}\} \\ & \cup \{(k,i,r) : \{r_i[k+1], x_i[k]\} \subseteq \mathcal{I}, r_i[k] = 0\} \cup \{(k,i,r) : \{r_i[k+1], r_i[k], x_i[k]\} \subseteq \mathcal{I}\}\end{aligned} \tag{4.8}$$

be the set that contains indices of the equations from (4.2) that can be *potentially* used in identifying β and δ, based on the measurements contained in $\mathcal{I}$. In other words, the coefficients on the left-hand side of equation (k, i, x) (resp., (k, i, r)) can be determined using the measurements from $\mathcal{I}$ and using Lemma 3.5 for all $(k, i, x) \in \bar{\mathcal{I}}$

(resp., $(k,i,r) \in \bar{\mathcal{I}}$). Let us now consider the coefficient matrix $\Phi^x_{k,i}$ (resp., $\Phi^r_{k,i}$) corresponding to $(k,i,x) \in \bar{\mathcal{I}}$ (resp., $(k,i,r) \in \bar{\mathcal{I}}$). One can then show that it is possible that there exist equations in $\bar{\mathcal{I}}$ whose coefficients cannot be (directly) determined using the measurements contained in $\mathcal{I}$ or using Lemma 3.5, where the undetermined coefficients come from the first element in $\Phi^x_{k,i}$ given by (4.7a). Nevertheless, it is also possible that one can perform algebraic operations among the equations in $\bar{\mathcal{I}}$ such that the undetermined coefficients get cancelled. Formally, we define the following.

DEFINITION 4.1. *Consider a measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$, where $\mathcal{I}_{t_1:t_2}$ is given by* (4.1)*. Stack coefficient matrices $\Phi^x_{k,i} \in \mathbb{R}^{1\times 2}$ for all $(k,i,x) \in \bar{\mathcal{I}}$ and $\Phi^r_{k,i} \in \mathbb{R}^{1\times 2}$ for all $(k,i,r) \in \bar{\mathcal{I}}$ into a single matrix, where $\Phi^x_{k,i}$ and $\Phi^r_{k,i}$ are given by* (4.7) *and $\bar{\mathcal{I}}$ is given by* (4.8)*. The resulting matrix is denoted as $\Phi(\mathcal{I}) \in \mathbb{R}^{|\bar{\mathcal{I}}|\times 2}$. Moreover, define $\tilde{\Phi}(\mathcal{I})$ to be the set that contains all the matrices $\Phi \in \mathbb{R}^{2\times 2}$ such that $(\Phi)_1$ and $(\Phi)_2$ can be obtained via algebraic operations among the rows in $\Phi(\mathcal{I})$, and the elements in $(\Phi)_1$ and $(\Phi)_2$ can be fully determined using the measurements from $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ and using Lemma* 3.5*.*

In other words, $\Phi \in \tilde{\Phi}(\mathcal{I})$ corresponds to two equations (in β and δ) obtained from (4.2) such that the coefficients on the right-hand side of the two equations can be determined using the measurements contained in $\mathcal{I}$ and using Lemma 3.5 (if the coefficients contain $x_i[k] = 0$ or $r_i[k] = 0$). Moreover, one can show that the coefficients on the left-hand side of the two equations obtained from (4.2) corresponding to Φ can also be determined using measurements from $\mathcal{I}$ and using Lemma 3.5. Putting the above arguments together, we see that given a measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$, β and δ can be uniquely identified if and only if there exists $\Phi \in \tilde{\Phi}(\mathcal{I})$ such that $\operatorname{rank}(\Phi) = 2$. Equivalently, denoting

$$r_{\max}(\mathcal{I}) \triangleq \max_{\Phi \in \tilde{\Phi}(\mathcal{I})} \operatorname{rank}(\Phi), \tag{4.9}$$

where $r_{\max}(\mathcal{I}) \triangleq 0$ if $\tilde{\Phi}(\mathcal{I}) = \emptyset$, we see that β and δ can be uniquely identified using the measurements from $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ if and only if $r_{\max}(\mathcal{I}) = 2$.

Remark 4.2. Note that if a measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ satisfies $r_{\max}(\mathcal{I}) = 2$, it follows from the above arguments that $|\bar{\mathcal{I}}| \geq 2$, i.e., $\Phi(\mathcal{I}) \in \mathbb{R}^{|\bar{\mathcal{I}}|\times 2}$ has at least two rows, where $\bar{\mathcal{I}}$ is defined in (4.8).

Recall that collecting the measurement of $x_i[k]$ (resp., $r_i[k]$) at any node $i \in \mathcal{V}$ and at any time step $k \in \mathbb{Z}_{\geq 0}$ incurs cost $c_{k,i} \in \mathbb{R}_{\geq 0}$ (resp., $b_{k,i} \in \mathbb{R}_{\geq 0}$). Given any measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$, we denote the cost associated with $\mathcal{I}$ as

$$c(\mathcal{I}) \triangleq \sum_{x_i[k] \in \mathcal{I}} c_{k,i} + \sum_{r_i[k] \in \mathcal{I}} b_{k,i}. \tag{4.10}$$

We then define the parameter identification measurement selection (PIMS) problem in the perfect measurement setting as follows.

PROBLEM 4.3. *Consider a discrete-time SIR model given by* (2.1) *with a directed graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, a weight matrix $A \in \mathbb{R}^{n\times n}$, a sampling parameter $h \in \mathbb{R}_{\geq 0}$, and sets $\mathcal{S}_I, \mathcal{S}_H \subseteq \mathcal{V}$ defined in Definition* 3.4*. Moreover, consider time steps $t_1, t_2 \in \mathbb{Z}_{\geq 0}$ with $t_1 < t_2$, and a cost $c_{k,i} \in \mathbb{R}_{\geq 0}$ of measuring $x_i[k]$ and a cost $b_{k,i} \in \mathbb{R}_{\geq 0}$ of measuring $r_i[k]$, for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$. The PIMS problem is to*

find $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$ *that solves*

$$\begin{aligned} &\min_{\mathcal{I}\subseteq\mathcal{I}_{t_1:t_2}} c(\mathcal{I}) \\ &s.t.\ r_{\max}(\mathcal{I}) = 2, \end{aligned} \tag{4.11}$$

where $\mathcal{I}_{t_1:t_2}$ *is defined in* (4.1), $c(\mathcal{I})$ *is defined in* (4.10), *and* $r_{\max}(\mathcal{I})$ *is defined in* (4.9).

We show that the PIMS problem is NP-hard via a polynomial-time reduction from the exact cover by 3-sets problem which is known to be NP-complete [10]. The proof is omitted here in the interest of space and can be found in [36].

THEOREM 4.4. *The PIMS problem is NP-hard.*

Theorem 4.4 indicates that there is no polynomial-time algorithm that solves all instances of the PIMS problem optimally (if P $\neq$ NP). Moreover, we note from the formulation of the PIMS problem given by Problem 4.3 that for a measurement selection strategy $\mathcal{I} \subseteq \mathcal{I}_{t_1:t_2}$, one needs to check whether $\max_{\Phi\in\bar{\Phi}(\mathcal{I})} \operatorname{rank}(\Phi) = 2$ holds, *before* the corresponding measurements are collected. However, in general, it is not possible to calculate $\operatorname{rank}(\Phi)$ when no measurements are collected. In order to bypass these issues, we will explore additional properties of the PIMS problem in the following.

4.2. Solving the PIMS problem. We start with the following result.

LEMMA 4.5. *Consider a discrete time SIR model given by* (2.1). *Suppose Assumptions* 3.1–3.2 *hold. Then, the following results hold, where* $\Phi^x_{k_1,i_1} \in \mathbb{R}^{1\times 2}$ *and* $\Phi^r_{k_2,i_2} \in \mathbb{R}^{1\times 2}$ *are defined in* (4.7), $\mathcal{S}'_I \triangleq \{i \in \mathcal{S}_I : a_{ii} > 0\}$, $\mathcal{S}' \triangleq \{i \in \mathcal{V}\setminus\mathcal{S}'_I : \mathcal{N}_i \neq \emptyset, \min\{d_j : j \in \mathcal{N}_i\} \neq \infty\}$, *and* $\mathcal{S}_I$ *and* d_i *are defined in Definition* 3.4 *for all* $i \in \mathcal{V}$.

(a) *For any* $i_1 \in \mathcal{S}'_I$ *and any* $i_2 \in \mathcal{V}$ *with* $d_{i_2} \neq \infty$, $\operatorname{rank}\big(\begin{bmatrix}(\Phi^x_{k_1,i_1})^T & (\Phi^r_{k_2,i_2})^T\end{bmatrix}\big) = 2$ *for all* $k_1 \geq 0$ *and for all* $k_2 \geq d_{i_2}$, *where* $k_1, k_2 \in \mathbb{Z}_{\geq 0}$.

(b) *For any* $i_1 \in \mathcal{S}'$ *and any* $i_2 \in \mathcal{V}$ *with* $d_{i_2} \neq \infty$, $\operatorname{rank}\big(\begin{bmatrix}(\Phi^x_{k_1,i_1})^T & (\Phi^r_{k_2,i_2})^T\end{bmatrix}\big) = 2$ *for all* $k_1 \geq \min\{d_j : j \in \mathcal{N}_{i_1}\}$, *and for all* $k_2 \geq d_{i_2}$, *where* $k_1, k_2 \in \mathbb{Z}_{\geq 0}$.

Proof. Noting from (4.7), we have

$$\begin{bmatrix}\Phi^x_{k_1,i_1}\\ \Phi^r_{k_2,i_2}\end{bmatrix} = \begin{bmatrix} s_{i_1}[k_1]\sum_{j\in\bar{\mathcal{N}}_{i_1}} a_{i_1 j}x_j[k_1] & -x_{i_1}[k_1]\\ 0 & x_{i_2}[k_2]\end{bmatrix}. \tag{4.12}$$

To prove part (a), consider any $i_1 \in \mathcal{S}'_I$ and any $i_2 \in \mathcal{V}$ with $d_{i_2} \neq \infty$, where we note $x_{i_1}[0] > 0$ and $a_{i_1 i_1} > 0$ from the definition of $\mathcal{S}'_I$. We then see from Lemma 3.5(a)–(b) that $s_{i_1}[k_1] > 0$ and $x_{i_1}[k_1] > 0$ for all $k_1 \geq 0$. It follows that $s_{i_1}[k_1]\sum_{j\in\bar{\mathcal{N}}_{i_1}} a_{i_1 j}x_j[k_1] > 0$ for all $k_1 \geq 0$. Also, we obtain from Lemma 3.5(b) $x_{i_2}[k_2] > 0$ for all $k_2 \geq d_{i_2}$, which proves part (a).

We now prove part (b). Considering any $i_1 \in \mathcal{S}'$ and any $i_2 \in \mathcal{V}$ with $d_2 \neq \infty$, we see from the definition of $\mathcal{S}'$ that $\mathcal{N}_{i_1} \neq \emptyset$ and there exists $j \in \mathcal{N}_{i_1}$ such that $d_j \neq \infty$. Letting j_1 be a node in $\mathcal{N}_{i_1}$ such that $d_{j_1} = \min\{d_j : j \in \mathcal{N}_{i_1}\} \neq \infty$, we note from Lemma 3.5(a) that $x_{j_1}[k_1] > 0$ for all $k_1 \geq \min\{d_j : j \in \mathcal{N}_{i_1}\}$. Also note that $a_{i_1 j_1} > 0$ from Assumption 3.2. The rest of the proof of part (b) is then identical to that of part (a). ∎

Recalling the way we index the equations in (4.2) (see (4.5)–(4.6) for examples), we define the set that contains all the indices of the equations in (4.2):

$$\mathcal{Q} \triangleq \{(k, i, \lambda) : k \in \{t_1, \ldots, t_2 - 1\}, i \in \mathcal{V}, \lambda \in \{x, r\}\}. \tag{4.13}$$

Following the arguments in Lemma 4.5, we denote

$$\mathcal{Q}_1 \triangleq \{(k,i,x) \in \mathcal{Q} : i \in \mathcal{S}_I'\} \cup \{(k,i,x) \in \mathcal{Q} : k \geq \min\{d_j : j \in \mathcal{N}_i\}, i \in \mathcal{S}'\}, \tag{4.14}$$

$$\mathcal{Q}_2 \triangleq \{(k,i,r) \in \mathcal{Q} : k \geq d_i, i \in \mathcal{V}, d_i \neq \infty\}, \tag{4.15}$$

where $\mathcal{S}_I'$ and $\mathcal{S}'$ are defined in Lemma 4.5, and d_i is defined in Definition 3.4. Next, for all $(k,i,x) \in \mathcal{Q}$, we define the set of measurements that are needed to determine the coefficients in equation (k,i,x) (when no other equations are used) to be

$$\mathcal{I}(k,i,x) \triangleq \big(\{x_i[k+1], r_i[k]\} \cup \{x_j[k] : j \in \bar{\mathcal{N}}_i\}\big) \cap \mathcal{I}_{t_1:t_2},$$

where $\mathcal{I}_{t_1:t_2}$ is defined in (4.1). Similarly, for all $(k,i,r) \in \mathcal{Q}$, we define

$$\mathcal{I}(k,i,r) \triangleq \big(\{r_i[k+1], r_i[k], x_i[k]\}\big) \cap \mathcal{I}_{t_1:t_2}.$$

Moreover, let us denote

$$\mathcal{I}((k_1,i_1,\lambda_1),(k_2,i_2,\lambda_2)) \triangleq \mathcal{I}(k_1,i_1,\lambda_1) \cup \mathcal{I}(k_2,i_2,\lambda_2) \tag{4.16}$$

for all $(k_1,i_1,\lambda_1),(k_2,i_2,\lambda_2) \in \mathcal{Q}$. Similarly to (4.10), denote the sum of the costs of the measurements from $\mathcal{I}((k_1,i_1,\lambda_1),(k_2,i_2,\lambda_2))$ as $c(\mathcal{I}((k_1,i_1,\lambda_1),(k_2,i_2,\lambda_2)))$.

Algorithm 4.1. Algorithm for PIMS.

1: **Input**: An instance of PIMS
2: Find $(k_1,i_1,x) \in \mathcal{Q}_1$, $(k_2,i_2,r) \in \mathcal{Q}_2$ s.t. $c(\mathcal{I}((k_1,i_1,x),(k_2,i_2,r)))$ is minimized
3: **return** $\mathcal{I}((k_1,i_1,x),(k_2,i_2,r))$

Based on the above arguments, we propose an algorithm defined in Algorithm 4.1 for the PIMS problem. Note that Algorithm 4.1 finds an equation from $\mathcal{Q}_1$ and an equation from $\mathcal{Q}_2$ such that the sum of the costs of the two equations is minimized, where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are defined in (4.14) and (4.15), respectively.

PROPOSITION 4.6. *Consider an instance of the PIMS problem under Assumptions* 3.1–3.2. *Algorithm* 4.1 *returns a solution* $\mathcal{I}((k_1,i_1,x),(k_2,i_2,r))$ *to the PIMS problem that satisfies the constraint in* (4.11) *and the following:*

$$\frac{c(\mathcal{I}((k_1,i_1,x),(k_2,i_2,r)))}{c(\mathcal{I}^\star)} \leq \frac{\min_{(k,i,x)\in\mathcal{Q}_1}(b_{k+1,i} + b_{k,i} + c_{k+1,i} + \sum_{j\in\bar{\mathcal{N}}_i} c_{k,j})}{3c_{\min}}, \tag{4.17}$$

where $\mathcal{I}^\star$ *is an optimal solution to the PIMS problem,* $\mathcal{Q}_1$ *is defined in* (4.14)*, and* $c_{\min} \triangleq \min\{\min_{x_i[k]\in\mathcal{I}_{t_1:t_2}} c_{k,i}, \min_{r_i[k]\in\mathcal{I}_{t_1:t_2}} b_{k,i}\} > 0$ *with* $\mathcal{I}_{t_1:t_2}$ *given by* (4.1).

Proof. The feasibility of $\mathcal{I}((k_1,i_1,x),(k_2,i_2,r))$ follows directly from the definition of Algorithm 4.1 and Lemma 4.5. We now prove (4.17). Consider any equations $(k,i,x) \in \mathcal{Q}_1$ and $(k,i,r) \in \mathcal{Q}_2$. We have from (4.16) the following:

$$\begin{aligned}&\mathcal{I}((k,i,x),(k,i,r))\\ &\quad= \big(\{x_i[k+1], r_i[k]\} \cup \{x_j[k] : j \in \bar{\mathcal{N}}_i\} \cup \{r_i[k+1], r_i[k], x_i[k]\}\big) \cap \mathcal{I}_{t_1:t_2},\end{aligned}$$

which implies

$$c(\mathcal{I}((k_1,i_1,x),(k_2,i_2,r))) \leq \min_{(k,i,x)\in\mathcal{Q}_1} \left(b_{k+1,i} + b_{k,i} + c_{k+1,i} + \sum_{j\in\bar{\mathcal{N}}_i} c_{k,j}\right).$$

Next, since $\mathcal{I}^\star$ satisfies $r_{\max}(\mathcal{I}^\star) = 2$, we recall from Remark 4.2 that $|\bar{\mathcal{I}}^\star| \geq 2$, where

$$\bar{\mathcal{I}}^\star = \{(k,i,x) : x_i[k+1] \in \mathcal{I}^\star, x_i[k] = 0\} \cup \{(k,i,x) : \{x_i[k+1], x_i[k]\} \subseteq \mathcal{I}^\star\}$$
$$\cup\{(k,i,r) : \{r_i[k{+}1], x_i[k]\} \subseteq \mathcal{I}^\star, r_i[k] = 0\} \cup \{(k,i,r) : \{r_i[k{+}1], r_i[k], x_i[k]\} \subseteq \mathcal{I}^\star\},$$

which implies $|\mathcal{I}^\star| \geq 2$. In fact, suppose $\{x_i[k+1], x_j[k+1]\} \subseteq \mathcal{I}^\star$, where $i, j \in \mathcal{V}$ and $k \in \{t_1 - 1, \ldots, t_2 - 1\}$. Since the elements in $\Phi^x_{k,i}$ and $\Phi^x_{k,j}$ (defined in (4.7)) cannot all be zero, we see that there exists $x_w[k] \in \mathcal{I}^\star$ (with $x_w[k] > 0$), where $w \in \mathcal{V}$. This further implies $|\mathcal{I}^\star| \geq 3$. Using similar arguments, one can show that $|\mathcal{I}^\star| \geq 3$ holds in general, which implies $c(\mathcal{I}^\star) \geq 3c_{\min}$. Combining the above arguments leads to (4.17). ∎

Finally, note that $\mathcal{Q}_1$, $\mathcal{Q}_2$, and $\mathcal{I}_{t_1:t_2}$ can be obtained by calling the breadth-first-search algorithm (see, e.g., [8]) $|\mathcal{S}_I|$ times with $O(|\mathcal{S}_I|(n + |\mathcal{E}|))$ total time complexity. Also note that the time complexity of line 2 in Algorithm 4.1 is $O(n^2(t_2 - t_1 + 1)^2)$. Thus, the overall time complexity of Algorithm 4.1 is $O(|\mathcal{Q}|^2 + |\mathcal{S}_I||\mathcal{E}|)$.

5. Measurement selection problem in random measurement setting. In this section, we assume that the initial condition $l \triangleq [(s[0])^T \ (x[0])^T \ (r[0])^T]^T$ is known. Nevertheless, our analysis can potentially be extended to cases where the initial condition l is given by a probability distribution.

5.1. Problem formulation. We consider the scenario where the measurement of $x_i[k]$ (resp., $r_i[k]$), denoted as $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$), is given by a pmf $p(\hat{x}_i[k]|x_i[k])$ (resp., $p(\hat{r}_i[k]|r_i[k])$). Note that one can express $x_i[k]$ in terms of l and $\theta \triangleq [\beta \ \delta]^T$ using (2.1b). Hence, given l and θ, we can alternatively write $p(\hat{x}_i[k]|x_i[k])$ as $p(\hat{x}_i[k]|l, \theta)$ for all $i \in \mathcal{V}$ and for all $k \in \mathbb{Z}_{\geq 1}$. Since the initial conditions are assumed to be known, we drop the dependency of $p(\hat{x}_i[k]|l, \theta)$ on l and denote the pmf of $\hat{x}_i[k]$ as $p(\hat{x}_i[k]|\theta)$ for all $i \in \mathcal{V}$ and for all $k \in \mathbb{Z}_{\geq 1}$. Similarly, given l and θ, we denote the pmf of $\hat{r}_i[k]$ as $p(\hat{r}_i[k]|\theta)$ for all $i \in \mathcal{V}$ and for all $k \in \mathbb{Z}_{\geq 1}$. Note that when collecting measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$) under a limited budget, one possibility is to give virus (resp., antibody) tests to a group of randomly and uniformly sampled individuals of the population at node $i \in \mathcal{V}$ and at time $k \in \mathbb{Z}_{\geq 1}$ (see, e.g., [2]), where a positive testing result indicates that the tested individual is infected (resp., recovered) at time k (see, e.g., [1]). Thus, the obtained random measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$ and the corresponding pmfs $p(\hat{x}_i[k]|\theta)$ and $p(\hat{r}_i[k]|\theta)$ depend on the total number of conducted virus tests and antibody tests at node i and at time k, respectively. Consider any node $i \in \mathcal{V}$ and any time step $k \in \mathbb{Z}_{\geq 1}$, where the total population of i is denoted by $N_i \in \mathbb{Z}_{\geq 1}$ and is assumed to be fixed over time. Suppose we are allowed to choose the number of virus (resp., antibody) tests that will be performed on the (randomly sampled) individuals at node i and at time k. Assume that the cost of performing the virus (resp., antibody) tests is proportional to the number of the tests. For any $i \in \mathcal{V}$ and for any $k \in \{t_1, \ldots, t_2\}$, let

$$\mathcal{C}_{k,i} \triangleq \{\zeta c_{k,i} : \zeta \in (\{0\} \cup [\zeta_i])\} \tag{5.1}$$

be the set of all possible costs that we can spend on collecting the measurement $\hat{x}_i[k]$, where $c_{k,i} \in \mathbb{R}_{\geq 0}$ and $\zeta_i \in \mathbb{Z}_{\geq 1}$. Similarly, for any $i \in \mathcal{V}$ and any $k \in \{t_1, \ldots, t_2\}$, let

$$\mathcal{B}_{k,i} \triangleq \{\eta b_{k,i} : \eta \in (\{0\} \cup [\eta_i])\} \tag{5.2}$$

denote the set of all possible costs that we can spend on collecting the measurement $\hat{r}_i[k]$, where $b_{k,i} \in \mathbb{R}_{\geq 0}$ and $\eta_i \in \mathbb{Z}_{\geq 1}$. For instance, $\zeta c_{k,i}$ can be viewed as the cost

of performing virus tests on ζN_i^x (randomly sampled) individuals in the population at node i, where $N_i^x \in \mathbb{Z}_{\geq 1}$ and $\zeta N_i^x \leq N_i$. To reflect the dependency of the pmf $p(\hat{x}_i[k]|\theta)$ (resp., $p(\hat{r}_i[k]|\theta)$) of measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$) on the cost spent on collecting the measurement of $x_i[k]$ (resp., $r_i[k]$), we further denote the pmf of $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$) as $p(\hat{x}_i[k]|\theta, \varphi_{k,i})$ (resp., $p(\hat{r}_i[k]|\theta, \omega_{k,i})$), where $\varphi_{k,i} \in \mathcal{C}_{k,i}$ (resp., $\omega_{k,i} \in \mathcal{B}_{k,i}$) with $\mathcal{C}_{k.i}$ (resp., $\mathcal{B}_{k,i}$) given by (5.1) (resp., (5.2)). Note that $\varphi_{k,i}$ (resp., $\omega_{k,i}$) is the cost spent on collecting measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$), and $\varphi_{k,i} = 0$ (resp., $\omega_{k,i} = 0$) indicates that measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$) is not collected.

In contrast with the exact measurement case studied in section 4, it is not possible to uniquely identify β and δ using measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$ which are now random variables. Thus, we will consider estimators of β and δ based on the measurements indicated by a measurement selection strategy. Similarly to section 4, given time steps $t_1, t_2 \in \mathbb{Z}_{\geq 1}$ with $t_2 \geq t_1$, define the set of all candidate measurements as

$$\mathcal{U}_{t_1:t_2} \triangleq \{\hat{x}_i[k] : i \in \mathcal{V}, k \in \{t_1, \ldots, t_2\}\} \cup \{\hat{r}_i[k] : i \in \mathcal{V}, k \in \{t_1, \ldots, t_2\}\}. \tag{5.3}$$

Recalling $\mathcal{C}_{k,i}$ and $\mathcal{B}_{k,i}$ defined in (5.1) and (5.2), respectively, we let $\mu \in \mathbb{Z}_{\geq 0}^{\mathcal{U}_{t_1:t_2}}$ be a measurement selection that specifies the costs spent on collecting measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$. Moreover, we define the set of all candidate measurement selections as

$$\mathcal{M} \triangleq \{\mu \in \mathbb{Z}_{\geq 0}^{\mathcal{U}_{t_1:t_2}} : \mu(\hat{x}_i[k]) \in (\{0\} \cup [\zeta_i]), \mu(\hat{r}_i[k]) \in (\{0\} \cup [\eta_i])\}, \tag{5.4}$$

where $\zeta_i, \eta_i \in \mathbb{Z}_{\geq 1}$ for all $i \in \mathcal{V}$. In other words, a measurement selection μ is defined over the integer lattice $\mathbb{Z}_{\geq 0}^{\mathcal{U}_{t_1:t_2}}$ so that μ is a vector of dimension $|\mathcal{U}_{t_1:t_2}|$, where each element of μ corresponds to an element in $\mathcal{U}_{t_1:t_2}$ and is denoted as $\mu(\hat{x}_i[k])$ (or $\mu(\hat{r}_i[k])$). The set $\mathcal{M}$ contains all $\mu \in \mathbb{Z}_{\geq 0}^{\mathcal{U}_{t_1:t_2}}$ such that $\mu(\hat{x}_i[k]) \in (\{0\} \cup [\zeta_i])$ and $\mu(\hat{r}_i[k]) \in (\{0\} \cup [\eta_i])$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$. Thus, for any $\varphi_{k,i} \in \mathcal{C}_{k,i}$ and for any $\omega_{k,i} \in \mathcal{B}_{k,i}$, there exists $\mu \in \mathcal{M}_{\geq 0}^{\mathcal{U}_{t_1:t_2}}$ such that $\mu(\hat{x}_i[k])c_{k,i} = \varphi_{k,i}$ and $\mu(\hat{r}_i[k])b_{k,i} = \omega_{k,i}$. In other words, $\mu(\hat{x}_i[k])c_{k,i}$ (resp., $\mu(\hat{r}_i[k])b_{k,i}$) indicates the cost spent on collecting the measurement of $x_i[k]$ (resp., $r_i[k]$). Given a measurement selection $\mu \in \mathbb{Z}_{\geq 0}^{t_1:t_2}$, we can also denote the pmfs of $\hat{x}_i[k]$ and $\hat{r}_i[k]$ as $p(\hat{x}_i[k]|\theta, \mu(\hat{x}_i[k]))$ and $p(\hat{r}_i[k]|\theta, \mu(\hat{r}_i[k]))$, respectively, where we drop the dependencies of the pmfs on $c_{k,i}$ and $b_{k,i}$ for notational simplicity.

To proceed, we consider the scenario where measurements can only be collected under a budget constraint given by $B \in \mathbb{R}_{\geq 0}$. Using the above notations, the budget constraint can be expressed as

$$\sum_{\hat{x}_i[k] \in \mathcal{U}_{t_1:t_2}} c_{k,i}\mu(\hat{x}_i[k]) + \sum_{\hat{r}_i[k] \in \mathcal{U}_{t_1:t_2}} b_{k,i}\mu(\hat{r}_i[k]) \leq B. \tag{5.5}$$

We then consider estimators of $\theta = [\beta\ \delta]^T$ based on any given measurement selection $\mu \in \mathcal{M}$. Considering any $\mu \in \mathcal{M}$, we denote

$$\mathcal{U}_i^\lambda \triangleq \{k : \mu(\hat{\lambda}_i[k]) > 0, k \in \{t_1, \ldots, t_2\}\} \tag{5.6}$$

for all $i \in \mathcal{V}$ and for all $\lambda \in \{x, r\}$. For all $i \in \mathcal{V}$ and for all $\lambda \in \{x, r\}$ with $\mathcal{U}_i^\lambda \neq \emptyset$, denote $y(\mathcal{U}_i^\lambda) \triangleq \begin{bmatrix} \hat{\lambda}_i[k_1] & \cdots \hat{\lambda}_i[k_{|\mathcal{U}_i^\lambda|}] \end{bmatrix}^T$, where $\mathcal{U}_i^\lambda = \{k_1, \ldots, k_{|\mathcal{U}_i^\lambda|}\}$. Letting

$$\mathcal{U}_\lambda \triangleq \{i : \mathcal{U}_i^\lambda \neq \emptyset, i \in \mathcal{V}\} \quad \forall \lambda \in \{x, r\},$$

we denote the measurement vector indicated by $\mu \in \mathcal{M}$ as

$$y(\mu) \triangleq \left[y(\mathcal{U}_{i_1}^x)^T \quad \cdots \quad y(\mathcal{U}_{i_{|\mathcal{U}_x|}}^x)^T \quad y(\mathcal{U}_{j_1}^r)^T \quad \cdots \quad y(\mathcal{U}_{j_{|\mathcal{U}_r|}}^r)^T \right]^T, \tag{5.7}$$

where $\mathcal{U}_x = \{i_1, \ldots, i_{|\mathcal{U}_x|}\}$ and $\mathcal{U}_r = \{j_1, \ldots, j_{|\mathcal{U}_r|}\}$. Note that $\hat{x}_i[k]$ and $\hat{r}_i[k]$ are (discrete) random variables with pmfs $p(\hat{x}_i[k]|\theta, \mu(\hat{x}_i[k]))$ and $p(\hat{r}_i[k]|\theta, \mu(\hat{r}_i[k]))$, respectively. We then see from (5.7) that $y(\mu)$ is a random vector whose pmf is denoted as $p(y(\mu)|\theta, \mu)$. Similarly, the pmf of $y(\mathcal{U}_i^x)$ (resp., $y(\mathcal{U}_i^r)$) is denoted as $p(y(\mathcal{U}_i^x)|\theta, \mu)$ (resp., $p(y(\mathcal{U}_i^r)|\theta, \mu)$). Given $t_1, t_2 \in \mathbb{Z}_{\geq 1}$ with $t_2 \geq t_1$, we make the following assumption on measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$.

Assumption 5.1. For any $i \in \mathcal{V}$ and for any $k_1, k_2 \in \{t_1, \ldots, t_2\}$ $(k_1 \neq k_2)$, $\hat{x}_i[k_1]$, $\hat{x}_i[k_2]$, $\hat{r}_i[k_1]$, and $\hat{r}_i[k_2]$ are independent of each other. Moreover, for any $i, j \in \mathcal{V}$ $(i \neq j)$ and for any $k_1, k_2 \in \{t_1, \ldots, t_2\}$, $\hat{x}_i[k_1]$ and $\hat{x}_j[k_2]$ are independent and $\hat{x}_i[k_1]$ and $\hat{r}_j[k_2]$ are independent.

The above assumption ensures that measurements from different nodes or from different time steps are independent and the measurements of $x_i[k]$ and $r_i[k]$ are also independent. It then follows from (5.7) that the pmf of $y(\mu)$ can be written as

$$p(y(\mu)|\theta, \mu) = \prod_{i \in \mathcal{U}_x} p(y(\mathcal{U}_i^x)|\theta, \mu) \cdot \prod_{j \in \mathcal{U}_r} p(y(\mathcal{U}_j^r)|\theta, \mu), \tag{5.8}$$

where we can further write $p(y(\mathcal{U}_i^x)|\theta, \mu) = \prod_{k \in \mathcal{U}_i^x} p(\hat{x}_i[k]|\theta, \mu(\hat{x}_i[k]))$ for all $i \in \mathcal{U}_x$, and $p(y(\mathcal{U}_j^r)|\theta, \mu) = \prod_{k \in \mathcal{U}_j^r} p(\hat{r}_j[k]|\theta, \mu(\hat{r}_j[k]))$ for all $j \in \mathcal{U}_r$.

In order to quantify the performance (e.g., precision) of estimators of θ based on μ, we use the Bayesian Cramer–Rao lower bound (BCRLB) (see, e.g., [34]) associated with μ. In the following, we introduce the BCRLB and explain why we choose it as a performance metric. First, given any measurement $\mu \in \mathcal{M}$, let $F_\theta(\mu)$ be the corresponding Fisher information matrix defined as

$$F_\theta(\mu) \triangleq -\mathbb{E}\begin{bmatrix} \frac{\partial^2 \ln p(y(\mu)|\theta,\mu)}{\partial \beta^2} & \frac{\partial^2 \ln p(y(\mu)|\theta,\mu)}{\partial \beta \partial \delta} \\ \frac{\partial^2 \ln p(y(\mu)|\theta,\mu)}{\partial \delta \partial \beta} & \frac{\partial^2 \ln p(y(\mu)|\theta,\mu)}{\partial \delta^2} \end{bmatrix} \tag{5.9}$$

with the expectation $\mathbb{E}[\cdot]$ taken with respect to $p(y(\mu)|\theta, \mu)$. Under Assumption 5.1 and some regularity conditions on the pmfs of $\hat{x}_i[k]$ and $\hat{r}_i[k]$, (5.9) can be written as (see, e.g., [14])

$$F_\theta(\mu) = \sum_{\lambda \in \{x,r\}} \sum_{i \in \mathcal{U}_\lambda} \sum_{k \in \mathcal{U}_i^\lambda} \mathbb{E}\left[\frac{\partial \ln p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))}{\partial \theta} \left(\frac{\partial \ln p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))}{\partial \theta} \right)^T \right]. \tag{5.10}$$

Consider any estimator $\hat{\theta}(\mu)$ of θ based on a measurement selection $\mu \in \mathcal{M}$, and assume that we have a prior pdf of $\theta = [\beta \ \delta]^T$, denoted as $p(\theta)$. Under some regularity conditions on the pmfs of $\hat{x}_i[k]$ and $\hat{r}_i[k]$, and $p(\theta)$, we have (see, e.g., [33, 34])

$$R_{\hat{\theta}(\mu)} = \mathbb{E}[(\hat{\theta}(\mu) - \theta)(\hat{\theta}(\mu) - \theta)^T] \succeq \bar{C}(\mu), \tag{5.11}$$

where $R_{\hat{\theta}(\mu)} \in \mathbb{R}^{2\times 2}$ is the error covariance of the estimator $\hat{\theta}(\mu)$, the expectation $\mathbb{E}[\cdot]$ is taken with respect to $p(y(\mu)|\theta, \mu)p(\theta)$, and $\bar{C}(\mu) \in \mathbb{R}^{2\times 2}$ is the BCRLB associated with the measurement selection μ. The BCRLB is defined as (see, e.g., [33, 34])

$$\bar{C}(\mu) \triangleq (\mathbb{E}_\theta[F_\theta(\mu)] + F_p)^{-1}, \tag{5.12}$$

where $\mathbb{E}_\theta[\cdot]$ denotes the expectation taken with respect to $p(\theta)$, $F_\theta(\mu)$ is given by (5.9), and $F_p \in \mathbb{R}^{2\times 2}$ encodes the prior knowledge of θ as

$$F_p = -\mathbb{E}_\theta \begin{bmatrix} \frac{\partial^2 \ln p(\theta)}{\partial \beta^2} & \frac{\partial^2 \ln p(\theta)}{\partial \beta \partial \delta} \\ \frac{\partial^2 \ln p(\theta)}{\partial \delta \partial \beta} & \frac{\partial^2 \ln p(\theta)}{\partial \delta^2} \end{bmatrix} = \mathbb{E}_\theta \left[\frac{\partial \ln p(\theta)}{\partial \theta} \left(\frac{\partial \ln p(\theta)}{\partial \theta} \right)^T \right] \succeq \mathbf{0}, \tag{5.13}$$

where the second equality holds under some regularity conditions on $p(\theta)$ (see, e.g., [33]).

Thus, the above arguments motivate us to consider (functions of) $\bar{C}(\cdot)$ as optimization metrics in the measurement selection problem studied in this section, in order to characterize the estimation performance corresponding to a measurement selection $\mu \in \mathcal{M}$. In particular, we will consider $\mathrm{tr}(\bar{C}(\cdot))$ and $\ln\det(\bar{C}(\cdot))$, which are widely used criteria in parameter estimation (see, e.g., [13]) and are also known as the Bayesian A-optimality and D-optimality criteria, respectively, in the context of experimental design (see, e.g., [28]). First, considering the optimization metric $\mathrm{tr}(\bar{C}(\cdot))$, we see from the above arguments that (5.11) directly implies $\mathrm{tr}(R_{\hat{\theta}(\mu)}) \geq \mathrm{tr}(\bar{C}(\mu))$ for all estimators $\hat{\theta}(\mu)$ of θ and for all $\mu \in \mathcal{M}$. Therefore, a measurement selection $\mu^\star$ that minimizes $\mathrm{tr}(\bar{C}(\mu))$ potentially yields a lower value of $\mathrm{tr}(R_{\hat{\theta}(\mu)})$ for an estimator $\hat{\theta}(\mu)$ of θ. Furthermore, there may exist an estimator $\hat{\theta}(\mu)$ that achieves the BCRLB (see, e.g., [33]); i.e., $\mathrm{tr}(\bar{C}(\mu))$ provides the minimum value of $\mathrm{tr}(R_{\hat{\theta}(\mu)})$ that can possibly be achieved by any estimator $\hat{\theta}(\mu)$ of θ, given a measurement selection μ. Similar arguments hold for $\ln\det(\bar{C}(\cdot))$. To proceed, denoting

$$f_a(\mu) \triangleq \mathrm{tr}(\bar{C}(\mu)) \text{ and } f_d(\mu) \triangleq \ln\det(\bar{C}(\mu)) \quad \forall \mu \in \mathcal{M}, \tag{5.14}$$

we define the parameter estimation measurement selection (PEMS) problem.

PROBLEM 5.2. *Consider a discrete-time SIR model given by* (2.1) *with a directed graph* $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$*, a weight matrix* $A \in \mathbb{R}^{n\times n}$*, a sampling parameter* $h \in \mathbb{R}_{\geq 0}$*, and an initial condition* $l = [((s[0])^T\ (x[0])^T\ (r[0])^T]^T$*. Moreover, consider time steps* $t_1, t_2 \in \mathbb{Z}_{\geq 1}$ *with* $t_2 \geq t_1$*; a set* $\mathcal{C}_{k,i} = \{\zeta c_{k,i} : \zeta \in (\{0\} \cup [\zeta_i])\}$ *with* $c_{k,i} \in \mathbb{R}_{\geq 0}$ *and* $\zeta_i \in \mathbb{Z}_{\geq 1}$ *for all* $i \in \mathcal{V}$ *and for all* $k \in \{t_1, \ldots, t_2\}$*; a set* $\mathcal{B}_{k,i} = \{\eta b_{k,i} : \eta \in (\{0\} \cup [\eta_i])\}$ *with* $b_{k,i} \in \mathbb{R}_{\geq 0}$ *and* $\eta_i \in \mathbb{Z}_{\geq 1}$ *for all* $i \in \mathcal{V}$ *and for all* $k \in \{t_1, \ldots, t_2\}$*; a budget* $B \in \mathbb{R}_{\geq 0}$*; and a prior pdf* $p(\theta)$*. Suppose* $\hat{x}_i[k]$ *(resp.,* $\hat{r}_i[k]$*) is given by a pmf* $p(\hat{x}_i[k]|\theta, \varphi_{k,i})$ *(resp.,* $p(\hat{r}_i[k]|\theta, \omega_{k,i})$*), where* $\varphi_{k,i} \in \mathcal{C}_{k,i}$ *(resp.,* $\omega_{k,i} \in \mathcal{B}_{k,i}$*). The PEMS problem is to find a measurement selection* μ *that solves*

$$\begin{aligned} &\min_{\mu\in\mathcal{M}} f(\mu) \\ &s.t. \sum_{\hat{x}_i[k]\in\mathcal{U}_{t_1:t_2}} c_{k,i}\mu(\hat{x}_i[k]) + \sum_{\hat{r}_i[k]\in\mathcal{U}_{t_1:t_2}} b_{k,i}\mu(\hat{r}_i[k]) \leq B, \end{aligned} \tag{5.15}$$

where $\mathcal{M}$ *is defined in* (5.4)*,* $f(\cdot)$ *can be either of* $f_a(\cdot)$ *or* $f_d(\cdot)$ *with* $f_a(\cdot)$ *and* $f_d(\cdot)$ *defined in* (5.14)*,* $\mathcal{U}_{t_1:t_2}$ *is defined in* (5.3)*, and* $\bar{C}(\mu)$ *is given by* (5.12)*.*

Note that $F_p \succeq \mathbf{0}$ from (5.13), and $f_a(\mathbf{0}) = \mathrm{tr}(\bar{C}(\mathbf{0})) = \mathrm{tr}((F_p)^{-1})$ and $f_d(\mathbf{0}) = \ln\det(\bar{C}(\mathbf{0})) = \ln\det((F_p)^{-1})$ from (5.12). We further assume that $F_p \succ \mathbf{0}$ in what follows, which implies $f(\mu) > 0$ for all $\mu \in \mathcal{M}$.

5.2. Solving the PEMS problem. In this section, we consider a measurement model with specific pmfs of $\hat{x}_i[k]$ and $\hat{r}_i[k]$ (see, e.g., [4, 12]). Nonetheless, our analysis can potentially be extended to other measurement models.

5.2.1. Pmfs of measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$. Consider any $i \in \mathcal{V}$ and any $k \in \{t_1, \dots, t_2\}$. Assume that the total population of node i is fixed over time and is denoted as $N_i \in \mathbb{Z}_{\geq 1}$. Given any measurement selection $\mu \in \mathcal{M}$ with $\mathcal{M}$ defined in (5.4), we recall from section 5.1 that $\mu(\hat{x}_i[k])c_{k,i}$ can be viewed as the cost of performing virus tests on $\mu(\hat{x}_i[k])N_i^x$ randomly and uniformly sampled individuals in the population of node $i \in \mathcal{V}$, where $\mu(\hat{x}_i[k]) \in (\{0\} \cup [\zeta_i])$ (with $\zeta_i \in \mathbb{Z}_{\geq 1}$), $c_{k,i} \in \mathbb{R}_{\geq 0}$, and $N_i^x \in \mathbb{Z}_{\geq 1}$ with $\zeta_i N_i^x \leq N_i$. Note that $x_i[k]$ is the proportion of the population at node i and at time k that is infected, and $x_i[k] \in [0, 1)$ under Assumptions 3.1–3.2 as shown by Lemma 3.5. Thus, a randomly and uniformly sampled individual in the population at node i and at time k will be an infected individual (at time k) with probability $x_i[k]$ and will be a noninfected (i.e., susceptible or recovered) individual with probability $1 - x_i[k]$. Supposing the tests are accurate,[2] we see from the above arguments that the obtained number of individuals that test positive, i.e., $N_i\hat{x}_i[k]$, is a binomial random variable with parameters $N_i^x\mu(\hat{x}_i[k]) \in \mathbb{Z}_{\geq 1}$ and $x_i[k] \in [0, 1)$. Thus, for any $i \in \mathcal{V}$ and for any $k \in \{t_1, \dots, t_2\}$, the pmf of $\hat{x}_i[k]$ is

$$p(\hat{x}_i[k] = x|\theta, \mu(\hat{x}_i[k])) = \binom{N_i^x\mu(\hat{x}_i[k])}{N_i x}(x_i[k])^{N_i x}(1 - x_i[k])^{N_i^x\mu(\hat{x}_i[k]) - N_i x}, \tag{5.16}$$

where $x \in \{0, \frac{1}{N_i}, \frac{2}{N_i}, \dots, \frac{N_i^x\mu(\hat{x}_i[k])}{N_i}\}$ with $x \in [0, 1]$ since $N_i^x\zeta_i \leq N_i$. Note that we do not define the pmf of measurement $\hat{x}_i[k]$ when $N_i^x\mu(\hat{x}_i[k]) = 0$, i.e., when $\mu(\hat{x}_i[k]) = 0$, since $\mu(\hat{x}_i[k]) = 0$ indicates that no measurement is collected for state $x_i[k]$. Also note that when $x_i[k] = 0$, the pmf of $\hat{x}_i[k]$ given in (5.16) reduces to $p(\hat{x}_i[k] = 0|\theta, \mu(\hat{x}_i[k])) = 1$. Moreover, since the weight matrix $A \in \mathbb{R}^{n\times n}$ and the sampling parameter $h \in \mathbb{R}_{\geq 0}$ are assumed to be given, we see that given $\theta = [\beta\ \delta]^T$ and initial condition $l = [(s[0])^T\ (x[0])^T\ (r[0])^T]^T$, $x_i[k]$ can be obtained using (2.1b) for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \dots, t_2\}$, where we can view $x_i[k]$ as a function in the unknown parameter θ. In other words, given l, θ, $\mu(\hat{x}_i[k])$, N_i^x, and N_i, one can obtain the right-hand side of (5.16). Again, we only explicitly express the dependency of the pmf of $\hat{x}_i[k]$ on θ and $\mu(\hat{x}_i[k])$ in (5.16). Following arguments similar to those above, we assume that for any $i \in \mathcal{V}$ and for any $k \in \{t_1, \dots, t_2\}$, measurement $\hat{r}_i[k]$ has the following pmf:

$$p(\hat{r}_i[k] = r|\theta, \mu(\hat{r}_i[k])) = \binom{N_i^r\mu(\hat{r}_i[k])}{N_i r}(r_i[k])^{N_i r}(1 - r_i[k])^{N_i^r\mu(\hat{r}_i[k]) - N_i r}, \tag{5.17}$$

where $r \in \{0, \frac{1}{N_i}, \frac{2}{N_i}, \dots, \frac{N_i^r\mu(\hat{r}_i[k])}{N_i}\}$ with $r \in [0, 1]$, $\mu(\hat{r}_i[k]) \in \{0, \dots, \eta_i\}$, $N_i^r \in \mathbb{Z}_{\geq 1}$, and $N_i^r\mu(\hat{r}_i[k]) \leq N_i$. Similarly, we note that the pmf of $\hat{r}_i[k]$ given in (5.17) reduces to $p(\hat{r}_i[k] = 0|\theta, \mu(\hat{r}_i[k])) = 1$ when $r_i[k] = 0$. Considering any measurement selection $\mu \in \mathcal{M}$ and any measurement $\hat{\lambda}_i[k] \in \mathcal{U}_{t_1:t_2}$, where $\lambda \in \{x, r\}$ and $\mathcal{U}_{t_1:t_2}$ is defined in (5.3), we have the following:

$$\begin{aligned}
&\mathbb{E}\left[\frac{\partial \ln p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))}{\partial\theta}\left(\frac{\partial \ln p(\hat{\lambda}_i[k]|\theta, \mu(\lambda_i[k]))}{\partial\theta}\right)^T\right] \\
&= \mathbb{E}\left[\left(\frac{\partial \ln p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))}{\partial\lambda_i[k]}\right)^2 \cdot \frac{\partial\lambda_i[k]}{\partial\theta}\left(\frac{\partial\lambda_i[k]}{\partial\theta}\right)^T\right] \qquad (5.18) \\
&= \frac{N_i^\lambda\mu(\hat{\lambda}_i[k])}{\lambda_i[k](1 - \lambda_i[k])} \cdot \frac{\partial\lambda_i[k]}{\partial\theta}\left(\frac{\partial\lambda_i[k]}{\partial\theta}\right)^T, \qquad (5.19)
\end{aligned}$$

[2]Here, "accurate" means that an infected individual (at time k) will test positive with probability one, and an individual that is not infected will test negative with probability one.

where the expectation $\mathbb{E}[\cdot]$ is taken with respect to $p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))$, and $\lambda_i[k] \in [0,1)$. To obtain (5.18), we note the form of $\ln p(\hat{\lambda}_i[k]|\theta, \mu(\hat{\lambda}_i[k]))$ in (5.16) (or (5.17)) and use the chain rule. Moreover, one can obtain (5.19) from the fact that $\hat{\lambda}_i[k]$ is a binomial random variable. Noting that the pmf of $\hat{\lambda}_i[k]$ reduces to $p(\hat{\lambda}_i[k] = 0|\theta, \mu(\hat{\lambda}_i[k])) = 1$ if $\lambda_i[k] = 0$ as argued above, we let the right-hand side of (5.19) be zero if $\lambda_i[k] = 0$.

5.2.2. Complexity of the PEMS problem. Under the measurement model described above, we show that the PEMS problem is also NP-hard via a polynomial-time reduction from the knapsack problem which is known to be NP-hard [10]. That is, there exist instances of the PEMS problem that cannot be solved optimally by any polynomial-time algorithm (if P $\neq$ NP). The proof is omitted here in the interest of space and can be found in [36].

THEOREM 5.3. *The PEMS problem is NP-hard.*

5.2.3. An equivalent formulation for the PEMS problem. Theorem 5.3 motivates us to consider approximation algorithms for solving the PEMS problem. To begin with, we note that the objective function in the PEMS problem can be viewed as a function defined over an integer lattice. We then have $f_a : \mathcal{M} \to \mathbb{R}_{\geq 0}$ and $f_d : \mathcal{M} \to \mathbb{R}_{\geq 0}$, where $\mathcal{M}$ is defined in (5.4). First, considering $f_a : \mathcal{M} \to \mathbb{R}_{\geq 0}$, we will define a set function $f_{Pa} : 2^{\bar{M}} \to \mathbb{R}_{\geq 0}$, where $\bar{\mathcal{M}}$ is a set constructed as

$$\bar{\mathcal{M}} \triangleq \{(\hat{x}_i[k], l_1) : i \in \mathcal{V}, k \in \{t_1, \ldots, t_2\}, l_1 \in [\zeta_i]\} \cup \{(\hat{r}_i[k], l_2) : i \in \mathcal{V}, k \in \{t_1, \ldots, t_2\}, l_2 \in [\eta_i]\}. \tag{5.20}$$

In other words, for any $i \in \mathcal{V}$ and for any $k \in \{t_1, \ldots, t_2\}$, we associate elements $(\hat{x}_i[k], 1), \ldots, (\hat{x}_i[k], \zeta_i)$ (resp., $(\hat{r}_i[k], 1), \ldots, (\hat{r}_i[k], \eta_i)$) in set $\bar{\mathcal{M}}$ to measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$). The set function $f_{Pa}(\cdot)$ is then defined as

$$f_{Pa}(\mathcal{Y}) \triangleq f_a(\mathbf{0}) - f_a(\mu_{\mathcal{Y}}) = \mathrm{tr}(\bar{C}(\mathbf{0})) - \mathrm{tr}(\bar{C}(\mu_{\mathcal{Y}})) \quad \forall \mathcal{Y} \subseteq \bar{\mathcal{M}}, \tag{5.21}$$

where for any $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, we define $\mu_{\mathcal{Y}} \in \mathcal{M}$ such that $\mu_{\mathcal{Y}}(\hat{x}_i[k]) = |\{(\hat{x}_i[k], l_1) : (\hat{x}_i[k], l_1) \in \mathcal{Y}\}|$ and $\mu_{\mathcal{Y}}(\hat{r}_i[k]) = |\{(\hat{r}_i[k], l_2) : (\hat{r}_i[k], l_2) \in \mathcal{Y}\}|$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$. In other words, $\mu_{\mathcal{Y}}(\hat{x}_i[k])$ (resp., $\mu_{\mathcal{Y}}(\hat{r}_i[k])$) is set to be the number of elements in $\mathcal{Y}$ that correspond to the measurement $\hat{x}_i[k]$ (resp., $\hat{r}_i[k]$). Also note that $f_{Pa}(\emptyset) = 0$. Following the arguments leading to (5.19), we define

$$H_y \triangleq \begin{cases} \mathbb{E}_\theta\big[\frac{N_i^x}{x_i[k](1-x_i[k])} \frac{\partial x_i[k]}{\partial \theta} \big(\frac{\partial x_i[k]}{\partial \theta}\big)^T\big] \text{ if } y = (\hat{x}_i[k], l_1) \\ \mathbb{E}_\theta\big[\frac{N_i^r}{r_i[k](1-r_i[k])} \frac{\partial r_i[k]}{\partial \theta} \big(\frac{\partial r_i[k]}{\partial \theta}\big)^T\big] \text{ if } y = (\hat{r}_i[k], l_2) \end{cases} \quad \forall y \in \bar{\mathcal{M}}, \tag{5.22}$$

where $x_i[k], r_i[k] \in [0,1)$, $i \in \mathcal{V}$, $k \in \{t_1, \ldots . t_2\}$, $l_1 \in [\zeta_i]$, $l_2 \in [\eta_i]$, and the expectation $\mathbb{E}_\theta[\cdot]$ is taken with respect to the prior pdf $p(\theta)$. Given any $\theta = [\beta\ \delta]^T$, we see from the arguments for (5.19) that $\frac{N_i^x}{x_i[k](1-x_i[k])} \frac{\partial x_i[k]}{\partial \theta} \big(\frac{\partial x_i[k]}{\partial \theta}\big)^T \succeq \mathbf{0}$. Moreover, one can show that $\mathbb{E}_\theta\big[\frac{N_i^x}{x_i[k](1-x_i[k])} \frac{\partial x_i[k]}{\partial \theta} \big(\frac{\partial x_i[k]}{\partial \theta}\big)^T\big] \succeq \mathbf{0}$. Similarly, one can obtain $\mathbb{E}_\theta\big[\frac{N_i^r}{r_i[k](1-r_i[k])} \frac{\partial r_i[k]}{\partial \theta} \big(\frac{\partial r_i[k]}{\partial \theta}\big)^T\big] \succeq \mathbf{0}$, which implies $H_y \succeq \mathbf{0}$ for all $y \in \bar{\mathcal{M}}$. Now, suppose the pmfs of $\hat{x}_i[k]$ and $\hat{r}_i[k]$ are given by (5.16) and (5.17), respectively. Recall from (5.12) that $\mathrm{tr}(\bar{C}(\mu)) = \mathrm{tr}((\mathbb{E}_\theta[F_\theta(\mu)] + F_p)^{-1})$ for all $\mu \in \mathcal{M}$, where F_p and $F_\theta(\mu)$ are given by (5.13) and (5.10), respectively. Supposing Assumption 5.1 holds, for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, one can first express $F_\theta(\mu_{\mathcal{Y}})$ using (5.19) and then use (5.22) to obtain

$\mathbb{E}_\theta[F_\theta(\mu_\mathcal{Y})] = \sum_{y\in\mathcal{Y}} H_y \triangleq H(\mathcal{Y})$, where $\mu_\mathcal{Y}$ is defined above given $\mathcal{Y} \subseteq \bar{\mathcal{M}}$. Putting the above arguments together, we have from (5.21) the following:

$$f_{Pa}(\mathcal{Y}) = \operatorname{tr}\left((F_p)^{-1}\right) - \operatorname{tr}\left((F_p + H(\mathcal{Y}))^{-1}\right) \quad \forall \mathcal{Y} \subseteq \bar{\mathcal{M}}. \tag{5.23}$$

Next, let the cost of $(\hat{x}_i[k], l_1)$ be $c_{k,i}$, denoted as $c(\hat{x}_i[k], l_1)$, for all $(\hat{x}_i[k], l_1) \in \bar{\mathcal{M}}$, and let the cost of $(\hat{r}_i[k], l_2)$ be $b_{k,i}$, denoted as $c(\hat{r}_i[k], l_2)$, for all $(\hat{r}_i[k], l_2) \in \bar{\mathcal{M}}$, where $c_{k,i} \in \mathbb{R}_{>0}$ and $b_{k,i} \in \mathbb{R}_{>0}$ are given in the instance of the PEMS problem. Setting the cost structure of the elements in $\bar{\mathcal{M}}$ in this way, we establish an equivalence between the cost of a subset $\mathcal{Y} \subseteq \bar{\mathcal{M}}$ and the cost of $\mu_\mathcal{Y} \in \mathcal{M}$, where $\mu_\mathcal{Y}$ is defined above. Similarly, considering the objective function $f_d : \mathcal{M} \to \mathbb{R}_{\geq 0}$ in the PEMS problem, we define a set function $f_{Pd} : 2^{\bar{\mathcal{M}}} \to \mathbb{R}_{\geq 0}$ as

$$f_{Pd}(\mathcal{Y}) \triangleq f_d(\mathbf{0}) - f_d(\mu_\mathcal{Y}) = \ln\det(F_p + H(\mathcal{Y})) - \ln\det(F_p) \quad \forall \mathcal{Y} \subseteq \bar{\mathcal{M}}, \tag{5.24}$$

where we define $\mu_\mathcal{Y} \in \mathcal{M}$ such that $\mu_\mathcal{Y}(\hat{x}_i[k]) = |\{(\hat{x}_i[k], l_1) : (\hat{x}_i[k], l_1) \in \mathcal{Y}\}|$ and $\mu_\mathcal{Y}(\hat{r}_i[k]) = |\{(\hat{r}_i[k], l_2) : (\hat{r}_i[k], l_2) \in \mathcal{Y}\}|$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$. Note that given an instance of the PEMS problem in Problem 5.2, we can construct the set $\bar{\mathcal{M}}$ with the associated costs of the elements in $\bar{\mathcal{M}}$ in $O(n(t_2 - t_1 + 1)(\zeta + \eta))$ time, where n is the number of nodes in graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, and $\zeta, \eta \in \mathbb{Z}_{\geq 1}$ with $\zeta_i \leq \zeta$ and $\eta_i \leq \eta$ for all $i \in \mathcal{V}$. Assuming that ζ and η are (fixed) constants, the construction of the set $\bar{\mathcal{M}}$ with the associated costs takes $O(n(t_2 - t_1 + 1))$ time, which is polynomial in the parameters of the PEMS problem (Problem 5.2). Based on the above arguments, we further consider the following problem:

$$\text{(P)} \qquad \begin{aligned} &\max_{\mathcal{Y}\subseteq\bar{\mathcal{M}}} f_P(\mathcal{Y}) \\ &\text{s.t. } c(\mathcal{Y}) \leq B, \end{aligned}$$

where $f_P(\cdot)$ can be either of $f_{Pa}(\cdot)$ or $f_{Pd}(\cdot)$ with $f_{Pa}(\cdot)$ and $f_{Pd}(\cdot)$ given by (5.23) and (5.24), respectively, and $c(\mathcal{Y}) \triangleq \sum_{y\in\mathcal{Y}} c(y)$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$. By the way we construct $f_P(\cdot)$ and the costs of elements in $\bar{\mathcal{M}}$, one can verify that $\mathcal{Y}_a^\star \subseteq \bar{\mathcal{M}}$ (resp., $\mathcal{Y}_d^\star \subseteq \bar{\mathcal{M}}$) is an optimal solution to problem (P) with $f_P(\cdot) = f_{Pa}(\cdot)$ (resp., $f_P(\cdot) = f_{Pd}(\cdot)$) if and only if $\mu_{\mathcal{Y}_a^\star}$ (resp., $\mu_{\mathcal{Y}_d^\star}$) defined above is an optimal solution to (5.15) in Problem 5.2 with $f(\cdot) = f_a(\cdot)$ (resp., $f(\cdot) = f_d(\cdot)$). Thus, given a PEMS instance we can first construct $\bar{\mathcal{M}}$ with the associated cost for each element in $\bar{\mathcal{M}}$ and then solve problem (P).

5.3. Greedy algorithm for the PEMS problem. Note that problem (P) can be viewed as a problem of maximizing a set function subject to a knapsack constraint, and greedy algorithms have been proposed to solve this problem with performance guarantees when the objective function is monotone nondecreasing and submodular[3] (see, e.g., [17, 30]). Before we formally introduce the greedy algorithm for the PEMS problem, we first note from (5.22)–(5.24) that given a prior pdf of θ and any $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, one has to take the expectation $\mathbb{E}_\theta[\cdot]$ in order to obtain the value of $f_P(\mathcal{Y})$. However, it is in general intractable to explicitly calculate the integration corresponding to $\mathbb{E}_\theta[\cdot]$. Hence, one may alternatively evaluate the value of $f_P(\mathcal{Y})$ using numerical integration with respect to $\theta = [\beta\ \delta]^T$ (see, e.g., [29]). Specifically, a typical numerical integration

[3] A set function $g : 2^\mathcal{V} \to \mathbb{R}$, where $\mathcal{V} = [n]$ is the ground set, is said to be monotone nondecreasing if $g(\mathcal{A}) \leq g(\mathcal{B})$ for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$. $g(\cdot)$ is called submodular if $g(\{y\} \cup \mathcal{A}) - g(\mathcal{A}) \geq g(\{y\} \cup \mathcal{B}) - g(\mathcal{B})$ for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and for all $y \in \mathcal{V} \setminus \mathcal{B}$.

method (e.g., the trapezoid rule) approximates the integral of a function (over an interval) based on a summation of (weighted) function values evaluated at certain points within the integration interval, which incurs an approximation error (see, e.g., [29] for more details). We then see from (5.22)–(5.24) that in order to apply the numerical integration method described above to $f_P(\mathcal{Y})$, one has to obtain the values of $x_i[k]$, $r_i[k]$, $\frac{\partial x_i[k]}{\partial \theta}$, and $\frac{\partial r_i[k]}{\partial \theta}$ for a given θ (within the integration interval), where $i \in \mathcal{V}$ and $t_1 \leq k \leq t_2$ with t_1, t_2 given in an instance of the PEMS problem. Recall that the initial conditions $s[0]$, $x[0]$, and $r[0]$ are assumed to be known. We first observe that for any given θ, the values of $x_i[k]$ and $r_i[k]$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$ can be obtained using the recursions in (2.1) in $O((t_2 - t_1 + 1)n^2)$ time. Next, noting that $\frac{\partial x_i[k]}{\partial \theta} = [\frac{\partial x_i[k]}{\partial \beta} \ \frac{\partial x_i[k]}{\partial \delta}]^T$ and $\frac{\partial r_i[k]}{\partial \theta} = [\frac{\partial r_i[k]}{\partial \beta} \ \frac{\partial r_i[k]}{\partial \delta}]^T$, we take the derivative with respect to β on both sides of the equation in (2.1a) and obtain

(5.25)
$$\frac{\partial s_i[k+1]}{\partial \beta} = \frac{\partial s_i[k]}{\partial \beta} - h\left(\frac{\partial s_i[k]}{\partial \beta}\beta + s_i[k]\right)\left(\sum_{j \in \bar{\mathcal{N}}_i} a_{ij} x_j[k]\right) - h s_i[k]\beta\left(\sum_{j \in \bar{\mathcal{N}}_i} a_{ij} \frac{\partial x_j[k]}{\partial \beta}\right).$$

Similarly, we take the derivative with respect to β on both sides of the equations in (2.1b) and (2.1c). Considering any given β, we can then use the recursion in (2.1) together with the recursion in (5.25) (and those obtained from (2.1b) and (2.1c), as we described above) to obtain the values of $\frac{\partial x_i[k]}{\partial \beta}$ and $\frac{\partial r_i[k]}{\partial \beta}$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$ in $O((t_2 - t_1 + 1)n^2)$ time. Similarly, considering any given δ, one can obtain the values of $\frac{\partial x_i[k]}{\partial \delta}$ and $\frac{\partial r_i[k]}{\partial \delta}$ for all $i \in \mathcal{V}$ and for all $k \in \{t_1, \ldots, t_2\}$ in $O((t_2 - t_1 + 1)n^2)$ time.

Putting the above arguments together and considering the prior pdf of θ, i.e., $p(\theta)$, we see from (5.22)–(5.24) that for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, an approximation of $f_P(\mathcal{Y})$, denoted as $\hat{f}_P(\mathcal{Y})$, can be obtained in $O(n_I(t_2 - t_1 + 1)n^2)$ time, where $n_I \in \mathbb{Z}_{\geq 1}$ is the number of points used for the numerical integration method with respect to θ, as we described above.[4] Furthermore, in what follows, we may assume that $\hat{f}_P(\cdot)$ satisfies $|\hat{f}_P(\mathcal{Y}) - f_P(\mathcal{Y})| \leq \varepsilon/2$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$ (with $\hat{f}_P(\emptyset) = 0$), where $\varepsilon \in \mathbb{R}_{\geq 0}$.[5] We are now ready to introduce the greedy algorithm given in Algorithm 5.1 to solve the PEMS problem, where $\hat{f}_P(\cdot)$ can be either of $\hat{f}_{Pa}(\cdot)$ or $\hat{f}_{Pd}(\cdot)$, and $\hat{f}_P(\cdot)$ is the approximation of $f_P(\cdot)$, as we described above. From the definition of Algorithm 5.1, we see that the number of function calls of $\hat{f}_P(\cdot)$ required in the algorithm is $O(|\bar{\mathcal{M}}|^2)$, and thus the overall time complexity of Algorithm 5.1 is given by $O(n_I(t_2 - t_1 + 1)n^2|\bar{\mathcal{M}}|^2)$.

We proceed to analyze the performance of Algorithm 5.1 when applied to the PEMS problem. First, one can observe that $f_{Pd}(\mathcal{Y}) = \ln\det(F_p + H(\mathcal{Y})) - \ln\det(F_p)$ in problem (P) shares a form similar to that in [31]. Thus, using arguments similar to those in [31], one can show that $f_{Pd}(\cdot)$ is monotone nondecreasing and submodular with $f_{Pd}(\emptyset) = 0$. Noting the assumption that $|\hat{f}_{Pd}(\mathcal{Y}) - f_{Pd}(\mathcal{Y})| \leq \varepsilon/2$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, one can show that $y^\star$ given in line 6 of Algorithm 5.1 satisfies that $\frac{f_{Pd}(\{y^\star\} \cup \mathcal{Y}_2) - f_{Pd}(\mathcal{Y}_2) + \varepsilon}{c(y^\star)} \geq \frac{f_{Pd}(\{y\} \cup \mathcal{Y}_2) - f_{Pd}(\mathcal{Y}_2) - \varepsilon}{c(y)}$ for all $y \in \mathcal{C}$. Moreover, one can show that $\max_{y \in \bar{\mathcal{M}}} f_{Pd}(y) \leq f_{Pd}(\mathcal{Y}_1) + \varepsilon$, where $\mathcal{Y}_1$ is given in line 3 in Algorithm 5.1. One can then use arguments similar to those for Theorem 1 in [17] and obtain the following result; the detailed proof is omitted for conciseness.

[4] We assume that n_I is polynomial in the parameters of the PEMS instance.

[5] Note that ε is related to the approximation error of the numerical integration method, and ε will decrease if n_I increases; see, e.g., [29] for more details about the numerical integration method.

THEOREM 5.4. *Consider problem* (P) *with the objective function* $f_{Pd} : 2^{\bar{\mathcal{M}}} \to \mathbb{R}_{\geq 0}$ *given by* (5.24). *Then Algorithm* 5.1 *yields a solution, denoted as* $\mathcal{Y}_d^g$, *to problem* (P) *that satisfies*

$$f_{Pd}(\mathcal{Y}_d^g) \geq \frac{1}{2}(1 - e^{-1}) f_{Pd}(\mathcal{Y}_d^\star) - \left(\frac{B}{c_{\min}} + \frac{3}{2}\right)\varepsilon, \tag{5.26}$$

where $\mathcal{Y}_d^\star \subseteq \bar{\mathcal{M}}$ *is an optimal solution to problem* (P), $c_{\min} = \min_{y \in \bar{\mathcal{M}}} c(y)$,[6] *and* $\varepsilon \in \mathbb{R}_{\geq 0}$ *satisfies* $|\hat{f}_{Pd}(\mathcal{Y}) - f_{Pd}(\mathcal{Y})| \leq \varepsilon/2$ *for all* $\mathcal{Y} \subseteq \bar{\mathcal{M}}$.

Algorithm 5.1. Greedy algorithm for PEMS.

Input: An instance of PEMS transformed into the form of (P)
Output: $\mathcal{Y}_g$
Find $\mathcal{Y}_1 \in \arg\max_{y \in \bar{\mathcal{M}}} \hat{f}_P(y)$
Initialize $\mathcal{Y}_2 = \emptyset$ and $\mathcal{C} = \bar{\mathcal{M}}$
while $\mathcal{C} \neq \emptyset$ **do**
  Find $y^\star \in \arg\max_{y \in \mathcal{C}} \frac{\hat{f}_P(\{y\} \cup \mathcal{Y}_2) - \hat{f}_P(\mathcal{Y}_2)}{c(y)}$
  if $c(y^\star) + c(\mathcal{Y}_2) \leq B$ **then**
    $\mathcal{Y}_2 = \{y^\star\} \cup \mathcal{Y}_2$
  end if
  $\mathcal{C} = \mathcal{C} \setminus \{y^\star\}$
end while
$\mathcal{Y}_g = \arg\max_{\mathcal{Y} \in \{\mathcal{Y}_1, \mathcal{Y}_2\}} \{\hat{f}_P(\mathcal{Y}_1), \hat{f}_P(\mathcal{Y}_2)\}$

It is worth noting that in general, the problem of maximizing a submodular function under a cardinality constraint *cannot be approximated within* $(1 - 1/e)$ (if P $\neq$ NP) [9]. In contrast to $f_{Pd}(\cdot)$, the objective function $f_{Pa}(\cdot)$ is not submodular in general (see, e.g., [18]). In fact, one can construct instances of the PEMS problem where the objective function $f_{Pa}(\mathcal{Y}) = \mathrm{tr}((F_p)^{-1}) - \mathrm{tr}\left((F_p + H(\mathcal{Y}))^{-1}\right)$ is not submodular. Hence, in order to provide performance guarantees of the greedy algorithm when applied to problem (P) with $f(\cdot) = f_{Pa}(\cdot)$, we will extend the analysis in [17] to nonsubmodular settings. To proceed, note that for all $\mathcal{A} \subseteq \mathcal{B} \subseteq \bar{\mathcal{M}}$, we have $F_p + H(\mathcal{A}) \preceq F_p + H(\mathcal{B})$, which implies $(F_p + H(\mathcal{A}))^{-1} \succeq (F_p + H(\mathcal{B}))^{-1}$ and $\mathrm{tr}(F_p + H(\mathcal{A}))^{-1}) \geq \mathrm{tr}(F_p + H(\mathcal{B}))^{-1})$. Therefore, the objective function $f_{Pa}(\cdot)$ is monotone nondecreasing with $f_{Pa}(\emptyset) = 0$. We then characterize how close $f_{Pa}(\cdot)$ is to being submodular by introducing the following definition.

DEFINITION 5.5. *Consider problem* (P) *with* $f_P(\cdot) = f_{Pa}(\cdot)$, *where* $f_{Pa} : 2^{\bar{\mathcal{M}}} \to \mathbb{R}_{\geq 0}$ *is defined in* (5.21). *Suppose Algorithm* 5.1 *is applied to solve problem* (P). *For all* $j \in \{1, \ldots, |\mathcal{Y}_2|\}$, *let* $\mathcal{Y}_2^j = \{y_1, \ldots, y_j\}$ *denote the set that contains the first* j *elements added to set* $\mathcal{Y}_2$ *in Algorithm* 5.1, *and let* $\mathcal{Y}_2^0 = \emptyset$. *The type-*1 *greedy submodularity ratio of* $f_{Pa}(\cdot)$ *is defined to be the largest* $\gamma_1 \in \mathbb{R}$ *that satisfies*

$$\sum_{y \in \mathcal{A} \setminus \mathcal{Y}_2^j} \left(f_{Pa}(\{y\} \cup \mathcal{Y}_2^j) - f_{Pa}(\mathcal{Y}_2^j)\right) \geq \gamma_1 \left(f_{Pa}(\mathcal{A} \cup \mathcal{Y}_2^j) - f_{Pa}(\mathcal{Y}_2^j)\right) \tag{5.27}$$

for all $\mathcal{A} \subseteq \bar{\mathcal{M}}$ *and for all* $j \in \{0, \ldots, |\mathcal{Y}_2|\}$. *The type-*2 *greedy submodularity ratio of* $f_{Pa}(\cdot)$ *is defined to be the largest* $\gamma_2 \in \mathbb{R}$ *that satisfies*

$$f_{Pa}(\mathcal{Y}_1) - f_{Pa}(\emptyset) \geq \gamma_2 \left(f_{Pa}(\{y\} \cup \mathcal{Y}_2^j) - f_{Pa}(\mathcal{Y}_2^j)\right) \tag{5.28}$$

[6] Note that we can assume without loss of generality that $c(y) \leq B$ for all $y \in \bar{\mathcal{M}}$.

for all $j \in \{0, \ldots, |\mathcal{Y}_2|\}$ *and for all* $y \in \bar{\mathcal{M}} \setminus \mathcal{Y}_2^j$ *such that* $c(y) + c(\mathcal{Y}_2^j) > B$, *where* $\mathcal{Y}_1 \in \arg\max_{y \in \bar{\mathcal{M}}} \hat{f}_{Pa}(y)$.

Remark 5.6. Note that $f_{Pa}(\cdot)$ is monotone nondecreasing as argued above. Noting the definition of γ_1 in (5.27), one can use arguments similar to those in [5] and show that $\gamma_1 \in [0, 1]$; if $f_{Pa}(\cdot)$ is submodular, then $\gamma_1 = 1$. Similarly, one can show that $\gamma_2 \geq 0$. Supposing that $\mathcal{Y}_1 \in \arg\max_{y \in \bar{\mathcal{M}}} f_{Pa}(y)$, one can further show that if $f_{Pa}(\cdot)$ is submodular, then $\gamma_2 \geq 1$.

Note that since we approximate $f_{Pa}(\cdot)$ using $\hat{f}_{Pa}(\cdot)$, we may not be able to obtain the exact values of γ_1 and γ_2 from Definition 5.5. Moreover, finding γ_1 may require an exponential number of function calls of $f_{Pa}(\cdot)$ (or $\hat{f}_{Pa}(\cdot)$). Nonetheless, it will be clear from our analysis below that obtaining lower bounds on γ_1 and γ_2 suffices. In particular, given $\mathcal{Y}_2^j$ for all $j \in \{0, \ldots, |\mathcal{Y}_2|\}$ from Algorithm 5.1, one can show that a lower bound on γ_2 (defined in (5.28)) can be obtained via $O(|\bar{\mathcal{M}}|^2)$ function calls of $\hat{f}_{Pa}(\cdot)$ (see [36] for details). We defer our analysis for lower bounding γ_1 to the end of this section, which requires more careful analysis. Based on Definition 5.5, the following result extends the analysis in [16, 17] and characterizes the performance guarantees of Algorithm 5.1 for problem (P) with $f_P(\cdot) = f_{Pa}(\cdot)$.

THEOREM 5.7. *Consider problem* (P) *with the objective function* $f_{Pa} : 2^{\bar{\mathcal{M}}} \to \mathbb{R}_{\geq 0}$ *given by* (5.21). *Then Algorithm* 5.1 *yields a solution, denoted as* $\mathcal{Y}_a^g$, *to problem* (P) *that satisfies*

$$f_{Pa}(\mathcal{Y}_a^g) \geq \frac{\min\{\gamma_2, 1\}}{2}(1 - e^{-\gamma_1}) f_{Pa}(\mathcal{Y}_a^\star) - \left(\frac{B + c_{\max}}{c_{\min}} + 1\right)\varepsilon, \tag{5.29}$$

where $\mathcal{Y}_a^\star \subseteq \bar{\mathcal{M}}$ *is an optimal solution to problem* (P), $\gamma_1 \in \mathbb{R}_{\geq 0}$ *and* $\gamma_2 \in \mathbb{R}_{\geq 0}$ *are defined in Definition* 5.5, $c_{\min} = \min_{y \in \bar{\mathcal{M}}} c(y)$, $c_{\max} = \max_{y \in \bar{\mathcal{M}}} c(y)$, *and* $\varepsilon \in \mathbb{R}_{\geq 0}$ *satisfies* $|\hat{f}_{Pa}(\mathcal{Y}) - f_{Pa}(\mathcal{Y})| \leq \varepsilon/2$ *for all* $\mathcal{Y} \subseteq \bar{\mathcal{M}}$.

Proof. Noting that (5.29) holds trivially if $\gamma_1 = 0$ or $\gamma_2 = 0$, we assume that $\gamma_1 > 0$ and $\gamma_2 > 0$. In this proof, we drop the subscript of $f_{Pa}(\cdot)$ (resp., $\hat{f}_{Pa}(\cdot)$) and denote $f(\cdot)$ (resp., $\hat{f}(\cdot)$) for notational simplicity. First, recall that for all $j \in \{1, \ldots, |\mathcal{Y}_2|\}$, we let $\mathcal{Y}_2^j = \{y_1, \ldots, y_j\}$ denote the set that contains the first j elements added to set $\mathcal{Y}_2$ in Algorithm 5.1 and let $\mathcal{Y}_2^0 = \emptyset$. Now, let j_l be the first index in $\{1, \ldots, |\mathcal{Y}_2|\}$ such that a candidate element $y^\star \in \arg\max_{y \in \mathcal{C}} \frac{\hat{f}(\{y\} \cup \mathcal{Y}_2^{j_l}) - \hat{f}(\mathcal{Y}_2^{j_l})}{c(y)}$ for $\mathcal{Y}_2$ (given in line 6 of Algorithm 5.1) cannot be added to $\mathcal{Y}_2$ due to $c(y^\star) + c(\mathcal{Y}_2^{j_l}) > B$. In other words, for all $j \in \{0, \ldots, j_l - 1\}$, any candidate element $y^\star \in \arg\max_{y \in \mathcal{C}} \frac{\hat{f}(\{y\} \cup \mathcal{Y}_2^j) - \hat{f}(\mathcal{Y}_2^j)}{c(y)}$ for $\mathcal{Y}_2$ satisfies $c(y^\star) + c(\mathcal{Y}_2^j) \leq B$ and can be added to $\mathcal{Y}_2$ in Algorithm 5.1. Noting that $|\hat{f}_P(\mathcal{Y}) - f_P(\mathcal{Y})| \leq \varepsilon/2$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, one can then show that the following holds for all $j \in \{0, \ldots, j_l - 1\}$:

$$\frac{f(\mathcal{Y}_2^{j+1}) - f(\mathcal{Y}_2^j) + \varepsilon}{c(y_{j+1})} \geq \frac{f(\{y\} \cup \mathcal{Y}_2^j) - f(\mathcal{Y}_2^j) - \varepsilon}{c(y)} \quad \forall y \in \bar{\mathcal{M}} \setminus \mathcal{Y}_2^j. \tag{5.30}$$

Now, considering any $j \in \{0, \ldots, j_l - 1\}$, we have the following:

$$f(\mathcal{Y}_a^\star \cup \mathcal{Y}_2^j) - f(\mathcal{Y}_2^j) \le \frac{1}{\gamma_1} \sum_{y \in \mathcal{Y}_a^\star \setminus \mathcal{Y}_2^j} c(y) \cdot \frac{f(\{y\} \cup \mathcal{Y}_2^j) - f(\mathcal{Y}_2^j)}{c(y)} \tag{5.31}$$

$$\le \frac{1}{\gamma_1} \sum_{y \in \mathcal{Y}_a^\star \setminus \mathcal{Y}_2^j} c(y) \Big(\frac{f(\mathcal{Y}_2^{j+1}) - f(\mathcal{Y}_2^j) + \varepsilon}{c(y_{j+1})} + \frac{\varepsilon}{c(y)} \Big) \tag{5.32}$$

$$\le \frac{B}{\gamma_1} \cdot \frac{f(\mathcal{Y}_2^{j+1}) - f(\mathcal{Y}_2^j)}{c(y_{j+1})} + \frac{\varepsilon}{\gamma_1} \sum_{y \in \mathcal{Y}_a^\star \setminus \mathcal{Y}_j^2} \Big(\frac{c(y)}{c(y_{j+1})} + 1 \Big) \tag{5.33}$$

$$\le \frac{B}{\gamma_1} \cdot \frac{f(\mathcal{Y}_2^{j+1}) - f(\mathcal{Y}_2^j)}{c(y_{j+1})} + \frac{\varepsilon}{\gamma_1} \Big(\frac{B}{c(y_{j+1})} + |\mathcal{Y}_a^\star| \Big), \tag{5.34}$$

where (5.31) follows from the definition of γ_1 in (5.27), and (5.32) follows from (5.30). To obtain (5.33), we use the fact $c(\mathcal{Y}_a^\star) \le B$. Similarly, we obtain (5.34). Noting that $f(\cdot)$ is monotone nondecreasing, one can further obtain from (5.34) that

$$f(\mathcal{Y}_2^{j+1}) - f(\mathcal{Y}_2^j) \ge \frac{\gamma_1 c(y_{j+1})}{B} \big(f(\mathcal{Y}_a^\star) - f(\mathcal{Y}_2^j) \big) - \varepsilon \Big(1 + |\mathcal{Y}_a^\star| \frac{c(y_{j+1})}{B} \Big). \tag{5.35}$$

To proceed, let $y' \in \arg\max_{y \in \mathcal{C}} \frac{\hat{f}(\{y\} \cup \mathcal{Y}_2^{j_l}) - \hat{f}(\mathcal{Y}_2^{j_l})}{c(y)}$ be the (first) candidate element for $\mathcal{Y}_2$ that cannot be added to $\mathcal{Y}_2$ due to $c(y') + c(\mathcal{Y}_2^{j_l}) > B$, as we argued above. Similarly to (5.30), one can see that $\frac{f(\{y'\} \cup \mathcal{Y}_2^{j_l}) - f(\mathcal{Y}_2^{j_l}) + \varepsilon}{c(y')} \ge \frac{f(\{y\} \cup \mathcal{Y}_2^{j_l}) - f(\mathcal{Y}_2^{j_l}) - \varepsilon}{c(y)}$ holds for all $y \in \bar{\mathcal{M}} \setminus \mathcal{Y}_2^{j_l}$. Letting $\bar{\mathcal{Y}}_2^{j_l+1} \triangleq \{y'\} \cup \mathcal{Y}_2^{j_l}$ and following arguments similar to those leading up to (5.35), we have

$$f(\bar{\mathcal{Y}}_2^{j_l+1}) - f(\mathcal{Y}_2^{j_l}) \ge \frac{\gamma_1 c(y')}{B} \big(f(\mathcal{Y}_a^\star) - f(\mathcal{Y}_2^{j_l}) \big) - \varepsilon \Big(1 + |\mathcal{Y}_a^\star| \frac{c(y')}{B} \Big). \tag{5.36}$$

Denoting $\Delta_j \triangleq f(\mathcal{Y}_a^\star) - f(\mathcal{Y}_2^j)$ for all $j \in \{0, \dots, j_l\}$ and $\Delta_{j_l+1} \triangleq f(\mathcal{Y}_a^\star) - f(\bar{\mathcal{Y}}_2^{j_l+1})$, we obtain from (5.35) and (5.36) the following:

$$\Delta_j \le \Delta_{j-1} \Big(1 - \frac{c(y_j)\gamma_1}{B} \Big) + \varepsilon + \frac{c(y_j)|\mathcal{Y}_a^\star|}{B} \varepsilon \quad \forall j \in [j_l + 1]. \tag{5.37}$$

Unrolling (5.37) yields

$$\Delta_{j_l+1} \le \Delta_0 \Bigg(\prod_{j=1}^{j_l} \Big(1 - \frac{c(y_j)\gamma_1}{B} \Big) \Bigg) \Big(1 - \frac{c(y')\gamma_1}{B} \Big) + \Big(j_l + 1 + \frac{c(\bar{\mathcal{Y}}_2^{j_l+1})|\mathcal{Y}_a^\star|}{B} \Big) \varepsilon \tag{5.38}$$

$$\implies \Delta_{j_l+1} \le \Delta_0 \Bigg(\prod_{j=1}^{j_l} \Big(1 - \frac{c(y_j)\gamma_1}{B} \Big) \Bigg) \Big(1 - \frac{c(y')\gamma_1}{B} \Big) + \frac{2(B + c_{\max})}{c_{\min}} \varepsilon. \tag{5.39}$$

To obtain (5.38), we use the facts that $(1 - \frac{c(y_j)\gamma_1}{B}) \le 1$ for all $j \in [j_l + 1]$ and $(1 - \frac{c(y')\gamma_1}{B}) \le 1$, since $\gamma_1 \in (0, 1]$, as we argued in Remark 5.6. To obtain (5.39), we first note from the way we defined j_l that $j_l + 1 \le c(\bar{\mathcal{Y}}_2^{j_l+1})/c_{\min} \le (B + c_{\max})/c_{\min}$. Also noting that $|\mathcal{Y}_a^\star| \le B/c_{\min}$, we then obtain (5.39).

Now, one can show that $\big(\prod_{j=1}^{j_l}(1-\frac{c(y_j)\gamma_1}{B})\big)(1-\frac{c(y')\gamma_1}{B}) \leq \prod_{j=1}^{j_l+1}(1-\frac{c(\bar{\mathcal{Y}}_2^{j_l+1})\gamma_1}{(j_l+1)B}) \leq e^{-\gamma_1\frac{c(\bar{\mathcal{Y}}_2^{j_l+1})}{B}}$ (see, e.g., [16]). We then have from (5.39) the following:

$$f(\mathcal{Y}_a^\star) - f(\bar{\mathcal{Y}}_2^{j_l+1}) \leq f(\mathcal{Y}_a^\star)e^{-\gamma_1\frac{c(\bar{\mathcal{Y}}_2^{j_l+1})}{B}} + \frac{2(B+c_{\max})}{c_{\min}}\varepsilon$$

$$\implies f(\bar{\mathcal{Y}}_2^{j_l+1}) \geq (1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{2(B+c_{\max})}{c_{\min}}\varepsilon, \tag{5.40}$$

where (5.40) follows from $c(\bar{\mathcal{Y}}_2^{j_l+1}) > B$. To proceed with the proof of the theorem, we note from the definition of γ_2 in Definition 5.5 that $f(\{y'\}\cup\mathcal{Y}_2^{j_l}) - f(\mathcal{Y}_2^{j_l}) \leq \frac{1}{\gamma_2}f(\mathcal{Y}_1)$ with $\gamma_2 > 0$, which together with (5.40) implies that

$$f(\mathcal{Y}_2^{j_l}) + \frac{1}{\gamma_2}f(\mathcal{Y}_1) \geq f(\bar{\mathcal{Y}}_2^{j_l+1}) \geq (1-e^{\gamma_1})f(\mathcal{Y}_a^\star) - \frac{2\bar{B}}{c_{\min}}\varepsilon, \tag{5.41}$$

where $\bar{B} \triangleq B + c_{\max}$. Since $f(\cdot)$ is monotone nondecreasing, we obtain from (5.41)

$$f(\mathcal{Y}_2) + \frac{1}{\gamma_2}f(\mathcal{Y}_1) \geq (1-e^{\gamma_1})f(\mathcal{Y}_a^\star) - \frac{2\bar{B}}{c_{\min}}\varepsilon. \tag{5.42}$$

We will then split our analysis into two cases. First, supposing $\gamma_2 \geq 1$, we see from (5.42) that at least one of $f(\mathcal{Y}_2) \geq \frac{1}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\bar{B}}{c_{\min}}\varepsilon$ and $f(\mathcal{Y}_1) \geq \frac{1}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\bar{B}}{c_{\min}}\varepsilon$ holds. Recalling that $|\hat{f}(\mathcal{Y}) - f(\mathcal{Y})| \leq \varepsilon/2$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, it follows that at least one of $\hat{f}(\mathcal{Y}_2) \geq \frac{1}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\bar{B}}{c_{\min}}\varepsilon - \frac{\varepsilon}{2}$ and $\hat{f}(\mathcal{Y}_1) \geq \frac{1}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\bar{B}}{c_{\min}}\varepsilon - \frac{\varepsilon}{2}$ holds. Second, supposing $\gamma_2 < 1$ and using similar arguments, we have that at least one of $\hat{f}(\mathcal{Y}_2) \geq \frac{1}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\bar{B}}{c_{\min}}\varepsilon - \frac{\varepsilon}{2}$ and $\hat{f}(\mathcal{Y}_1) \geq \frac{\gamma_2}{2}(1-e^{-\gamma_1})f(\mathcal{Y}_a^\star) - \frac{\gamma_2\bar{B}}{c_{\min}}\varepsilon - \frac{\varepsilon}{2}$ holds. Now, we note from line 12 of Algorithm 5.1 that $\hat{f}(\mathcal{Y}_a^g) \geq \max\{\hat{f}(\mathcal{Y}_1), \hat{f}(\mathcal{Y}_2)\}$, which implies $f(\mathcal{Y}_a^g) \geq \max\{\hat{f}(\mathcal{Y}_1), \hat{f}(\mathcal{Y}_2)\} - \frac{\varepsilon}{2}$. Combining the above arguments together, we obtain (5.29). ∎

Remark 5.8. Note that (5.29) becomes $f_{Pa}(\mathcal{Y}_a^g) \geq \frac{1}{2}(1-e^{-\gamma_1})f_{Pa}(\mathcal{Y}_a^\star) - (\frac{B+c_{\max}}{c_{\min}} + 1)\varepsilon$ if $\gamma_2 \geq 1$. Also note that $\gamma_2 \geq 1$ can hold when the objective function $f_{Pa}(\cdot)$ is not submodular, as we will see later in our numerical examples.

Remark 5.9. The authors of [32] also extended the analysis of Algorithm 5.1 to nonsubmodular settings, when the objective function can be exactly evaluated (i.e., $\varepsilon = 0$). They obtained a performance guarantee for Algorithm 5.1 that depends on a submodularity ratio defined in a different manner. One can show that the submodularity ratios defined in Definition 5.5 are lower bounded by the one defined in [32], which further implies that the performance bound (when ε is 0) for Algorithm 5.1 given in Theorem 5.7 is tighter than that provided in [32].

Finally, we aim to provide a lower bound on γ_1 that can be computed in polynomial time. The lower bounds on γ_1 and γ_2 together with Theorem 5.7 will also provide performance guarantees for the greedy algorithm.

LEMMA 5.10 (see [11]). *For any positive semidefinite matrices $P, Q \in \mathbb{R}^{n\times n}$, $\lambda_1(P) \leq \lambda_1(P+Q) \leq \lambda_1(P) + \lambda_1(Q)$, and $\lambda_n(P+Q) \geq \lambda_n(P) + \lambda_n(Q)$.*

We have the following result; the proof is included in section 7.1 in the appendix.

LEMMA 5.11. *Consider the set function* $f_{Pa} : 2^{\bar{\mathcal{M}}} \to \mathbb{R}_{\geq 0}$ *defined in* (5.21). *The type-*1 *greedy submodularity ratio of* $f_{Pa}(\cdot)$ *given by Definition* 5.5 *satisfies*

$$\gamma_1 \geq \min_{j\in\{0,\ldots,|\mathcal{Y}_2|\}} \frac{\lambda_2(F_p + H(\mathcal{Y}_2^j))\lambda_2(F_p + H(\{z_j\} \cup \mathcal{Y}_2^j))}{\lambda_1(F_p + H(\mathcal{Y}_2^j))\lambda_1(F_p + H(\{z_j\} \cup \mathcal{Y}_2^j))}, \tag{5.43}$$

where $\mathcal{Y}_2^j$ *contains the first* j *elements added to* $\mathcal{Y}_2$ *in Algorithm* 5.1 *for all* $j \in \{1,\ldots,|\mathcal{Y}_2|\}$ *with* $\mathcal{Y}_2^0 = \emptyset$, F_p *is given by* (5.13), $H(\mathcal{Y}) = \sum_{y\in\mathcal{Y}} H_y$ *for all* $\mathcal{Y} \subseteq \bar{\mathcal{M}}$ *with* $H_y \succeq \mathbf{0}$ *defined in* (5.22), *and* $z_j \in \arg\min_{y\in\bar{\mathcal{M}}\setminus\mathcal{Y}_2^j} \frac{\lambda_2(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}{\lambda_1(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}$ *for all* $j \in \{1,\ldots,|\mathcal{Y}_2|\}$.

Recalling our arguments at the beginning of this section, we may only obtain approximations of the entries in the (2 by 2) matrix $F_p + H(\mathcal{Y})$ for $\mathcal{Y} \subseteq \bar{\mathcal{M}}$ using, e.g., numerical integration, where $H(\mathcal{Y}) = \sum_{y\in\mathcal{Y}} H_y$, and H_y (resp., F_p) is defined in (5.22) (resp., (5.13)). Specifically, for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, let $\hat{H}(\mathcal{Y}) = (F_p + H(\mathcal{Y})) + E(\mathcal{Y})$ be the approximation of $F_p + H(\mathcal{Y})$, where each entry of $E(\mathcal{Y}) \in \mathbb{R}^{2\times 2}$ represents the approximation error of the corresponding entry in $F_p + H(\mathcal{Y})$. Since F_p and $H(\mathcal{Y})$ are positive semidefinite matrices, $E(\mathcal{Y})$ is a symmetric matrix. Now, using a standard eigenvalue perturbation result, e.g., Corollary 6.3.8 in [11], one can obtain that $\sum_{i=1}^{2} |\lambda_i(F_p + H(\mathcal{Y})) - \lambda_i(\hat{H}(\mathcal{Y})|^2 \leq \|E(\mathcal{Y})\|_F^2$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$, where $\|E(\mathcal{Y})\|_F \triangleq \sqrt{\mathrm{tr}(E(\mathcal{Y})^T E(\mathcal{Y}))}$. It then follows that

$$\frac{\lambda_2(F_p + H(\mathcal{Y}))}{\lambda_1(F_p + H(\mathcal{Y}))} \geq \frac{\lambda_2(\hat{H}(\mathcal{Y})) - \|E(\mathcal{Y})\|_F}{\lambda_1(\hat{H}(\mathcal{Y})) + \|E(\mathcal{Y})\|_F} \geq \frac{\lambda_2(\hat{H}(\mathcal{Y})) - \varepsilon'}{\lambda_1(\hat{H}(\mathcal{Y})) + \varepsilon'} \quad \forall \mathcal{Y} \subseteq \bar{\mathcal{M}},$$

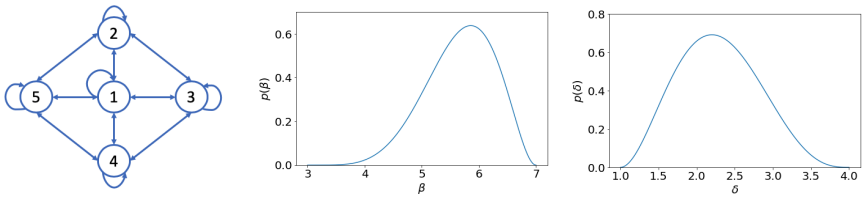
where $\varepsilon' \in \mathbb{R}_{\geq 0}$ and satisfies $\|E(\mathcal{Y})\|_F \leq \varepsilon'$ for all $\mathcal{Y} \subseteq \bar{\mathcal{M}}$. Combining the above arguments with (5.43) in Lemma 5.11, we obtain a lower bound on γ_1 that can be computed using $O(|\bar{\mathcal{M}}|^2)$ function calls of $\hat{H}(\cdot)$.

5.3.1. Illustrations. Note that one can further obtain from (5.43)

$$\gamma_1 \geq \min_{j\in\{0,\ldots,|\mathcal{Y}_2|\}} \frac{\lambda_2(F_p) + \lambda_2(H(\mathcal{Y}_2^j))}{\lambda_1(F_p) + \lambda_1(H(\mathcal{Y}_2^j))} \cdot \frac{\lambda_2(F_p) + \lambda_2(H(z_j)) + \lambda_2(H(\mathcal{Y}_2^j))}{\lambda_1(F_p) + \lambda_1(H(z_j)) + \lambda_1(H(\mathcal{Y}_2^j))}, \tag{5.44}$$

where $z_j \in \arg\min_{y\in\bar{\mathcal{M}}\setminus\mathcal{Y}_2^j} \frac{\lambda_2(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}{\lambda_1(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}$. Supposing F_p is fixed, we see from (5.44) that the lower bound on γ_1 would potentially increase if $\lambda_2(H(z_j))/\lambda_1(H(z_j))$ and $\lambda_2(H(\mathcal{Y}_2^j))/\lambda_1(H(\mathcal{Y}_2^j))$ increase. Recall that F_p given by (5.13) encodes the prior knowledge that we have about $\theta = [\beta\ \delta]^T$. Moreover, recall from (5.22) that $H(y)$ depends on the prior pdf $p(\theta)$ and the dynamics of the SIR model in (2.1). Thus, the lower bound given by Lemma 5.11 and thus the corresponding performance bound of Algorithm 5.1 given in Theorem 5.7 depend on the prior knowledge that we have on $\theta = [\beta\ \delta]^T$ and the dynamics of the SIR model. Also note that the performance bounds given in Theorem 5.7 are *worst-case* performance bounds for Algorithm 5.1. Thus, in practice the ratio between a solution returned by the algorithm and an optimal solution can be smaller than the ratio predicted by Theorem 5.7. As we will see in our simulations in the next section, the greedy algorithm provides solutions that are close to optimal in practice. Moreover, instances with tighter performance bounds potentially imply better performance of the algorithm when applied to those instances. Similar arguments hold for the performance bounds provided in Theorem 5.4.

5.3.2. Simulations. To validate the theoretical results in Theorems 5.4 and 5.7, and Lemma 5.11, we consider various PEMS instances.[7] The directed network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is given in Figure 1(a). According to the existing literature about the estimated infection and recovery rates for the COVID-19 pandemic (see, e.g., [27]), we assume that the infection rate β and the recovery rate δ lie in the intervals $[3, 7]$ and $[1, 4]$, respectively. Let the prior pdf of β (resp., δ) be a (linearly transformed) Beta distribution with parameters $\alpha_1 = 6$ and $\alpha_2 = 3$ (resp., $\alpha_1 = 3$ and $\alpha_2 = 4$), where β and δ are also assumed to be independent. The prior pdfs of β and δ are then plotted in Figures 1(b) and 1(c), respectively. We set the sampling parameter to be $h = 0.1$. We randomly generate the weight matrix $A \in \mathbb{R}^{5\times 5}$ such that Assumptions 3.1–3.2 are satisfied, where each entry of A is drawn (independently) from certain uniform distributions. The initial condition is set to be $s_1[0] = 0.95$, $x_1[0] = 0.05$ and $r_1[0] = 0$, and $s_i[0] = 0.99$, $x_i[0] = 0.01$, and $r_i[0] = 0$ for all $i \in \{2, \ldots, 5\}$. In the pmfs of measurements $\hat{x}_i[k]$ and $\hat{r}_i[k]$ given in (5.16) and (5.17), respectively, we set $N_i^x = N_i^r = 100$ and $N_i = 1000$ for all $i \in \mathcal{V}$, where N_i is the total population at node i.



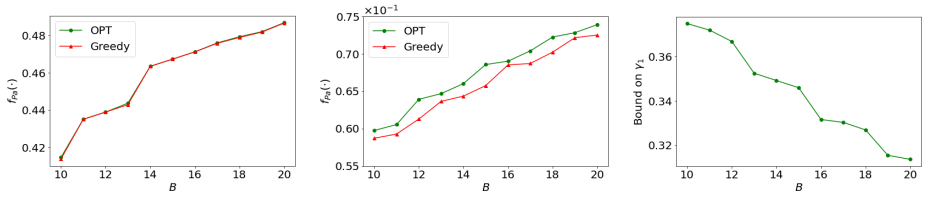
(a) Network (b) Prior pdf of β (c) Prior pdf of δ

FIG. 1. *Network structure and prior pdfs of β and δ.*

First, let us consider PEMS instances with a relatively smaller size. In such instances, we set the time steps $t_1 = t_2 = 5$; i.e., we only consider collecting measurements at time step $k = 5$. In the sets $\mathcal{C}_{5,i} = \{\zeta c_{5,i} : \zeta \in (\{0\} \cup [\zeta_i])\}$ and $\mathcal{B}_{5,i} = \{\eta b_{5,i} : \eta \in (\{0\} \cup [\eta_i])\}$, we let $c_{5,i} = b_{5,i}$ and $\zeta_i = \eta_i = 2$ for all $i \in \mathcal{V}$ and draw $c_{5,i}$ and $b_{5,i}$ uniformly randomly from $\{1, 2, 3\}$. Here, we can choose to perform 0, 100, or 200 virus (or antibody) tests at a node $i \in \mathcal{V}$ and at $k = 5$. Since the set $\bar{\mathcal{M}}$ defined in (5.20) has size 20, it allows us to compare the solution returned by the greedy algorithm (Algorithm 5.1) to the optimal solution (found by brute force). In Figure 2(a), we consider the objective function $f_{Pd}(\cdot)$, given by (5.24), in the PEMS instances constructed above, and plot the greedy solutions and the optimal solutions to the PEMS instances under different values of budget B. Note that for all the simulation results in this section, we obtain the averaged results from 50 randomly generated A matrices as described above for each value of B. As shown in Theorem 5.4, the greedy algorithm yields a $\frac{1}{2}(1 - e^{-1}) \approx 0.31$ approximation for $f_{Pd}(\cdot)$ (in the worst case), and the results in Figure 2(a) show that the greedy algorithm performs near optimally for the PEMS instances generated above. Similarly, in Figure 2(b), we plot the greedy solutions and the optimal solutions to the PEMS instances constructed above under different values of B, when the objective function is $f_{Pa}(\cdot)$ given in (5.21). Again, the results in Figure 2(b) show that the greedy algorithm performs well for the constructed PEMS instances. Moreover, according to Lemma 5.11, we plot the lower bound on the submodularity ratio γ_1 of $f_{Pa}(\cdot)$ in Figure 2(c). Here,

[7] In our simulations, we neglect the approximation error corresponding to the numerical integrations discussed in section 5.3 since the error terms are found to be sufficiently small.

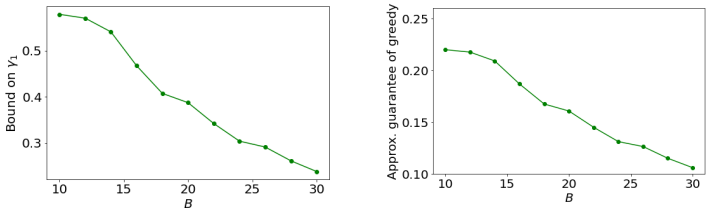
we note that the submodularity ratio γ_2 of $f_{Pa}(\cdot)$ is always greater than one in the PEMS instances constructed above. Hence, Theorem 5.7 yields a $\frac{1}{2}(1-e^{-\gamma_1})$ worst-case approximation guarantee for the greedy algorithm, where $\frac{1}{2}(1-e^{-0.3}) \approx 0.13$.



(a) OPT vs. Greedy for $f_{Pd}(\cdot)$ (b) OPT vs. Greedy for $f_{Pa}(\cdot)$ (c) Bound on γ_1

FIG. 2. *Results for PEMS instances of medium size.*

We then investigate the performance of the greedy algorithm for PEMS instances of a larger size. Different from the smaller instances constructed above, we set $t_1 = 1$ and $t_2 = 5$. We let $\zeta_i = \eta_i = 10$ for all $i \in \mathcal{V}$ in $\mathcal{C}_{k,i} = \{\zeta c_{k,i} : \zeta \in (\{0\} \cup [\zeta_i])\}$ and $\mathcal{B}_{k,i} = \{\eta b_{k,i} : \eta \in (\{0\} \cup [\eta_i])\}$, where we also set $c_{k,i} = b_{k,i}$ and draw $c_{k,i}$ and $b_{k,i}$ uniformly randomly from $\{1, 2, 3\}$, for all $k \in [5]$ and for all $i \in \mathcal{V}$. Here, we can choose to perform 0, 100, 200, ..., or 1000 virus (or antibody) tests at a node $i \in \mathcal{V}$ and at a time step $k \in [5]$. It follows from (5.20) that $|\bar{\mathcal{M}}| = 500$. Moreover, we modify the parameter of the Beta distribution corresponding to the pdf of β to be $\alpha_1 = 8$ and $\alpha_2 = 3$. Since the optimal solution to the PEMS instances cannot be efficiently obtained when the size of the instances becomes large, we obtain the lower bound on the submodularity ratio γ_1 of $f_{Pa}(\cdot)$ provided in Lemma 5.11, which can be computed in polynomial time. In Figure 3(a), we plot the lower bound on γ_1 obtained from the PEMS instances constructed above. We note that the submodularity ratio γ_2 of $f_{Pa}(\cdot)$ is also always greater than one. Hence, Theorem 5.7 yields a $\frac{1}{2}(1-e^{-\gamma_1})$ worst-case approximation guarantee for the greedy algorithm. We plot in Figure 3(b) the approximation guarantee using the lower bound that we obtained on γ_1.



(a) Bound on γ_1 (b) Approx. guarantee of greedy

FIG. 3. *Results for PEMS instances of large size.*

6. Conclusion. We first considered the PIMS problem under the exact measurement setting, and showed that the problem is NP-hard. We then proposed an approximation algorithm that returns a solution to the PIMS problem that is within a certain factor of the optimal one. Next, we studied the PEMS problem under the noisy measurement setting. Again, we showed that the problem is NP-hard. We applied a greedy algorithm to solve the PEMS problem and provided performance guarantees on the greedy algorithm. We presented numerical examples to validate the obtained

performance bounds of the greedy algorithm and showed that the greedy algorithm performs well in practice.

7. Appendix.

7.1. Proof of Lemma 5.11. Noting the definition of γ_1 in Definition 5.5, we provide a lower bound on $\frac{\sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}(f_{Pa}(\{y\}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j))}{f_{Pa}(\mathcal{A}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j)}$ for all $\mathcal{A}\subseteq\bar{\mathcal{M}}$ and for all $\mathcal{Y}_2^j$, where we assume that $\mathcal{A}\setminus\mathcal{Y}_2^j\neq\emptyset$; otherwise, (5.27) would be satisfied for all $\gamma_1\in\mathbb{R}$. Recalling the expression of $f_{Pa}(\cdot)$ in (5.23), we lower bound $LHS\triangleq\sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\left(f_{Pa}(\{y\}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j)\right)$ in the following manner:

$$\begin{aligned}
LHS &= \sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\sum_{i=1}^{2}\frac{\lambda_i(F_p+H(\{y\}\cup\mathcal{Y}_2^j))-\lambda_i(F_p+H(\mathcal{Y}_2^j))}{\lambda_i(F_p+H(\mathcal{Y}_2^j))\lambda_i(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}\\
&\geq \sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\frac{\sum_{i=1}^{2}(\lambda_i(F_p+H(\{y\}\cup\mathcal{Y}_2^j))-\lambda_i(F_p+H(\mathcal{Y}_2^j)))}{\lambda_1(F_p+H(\mathcal{Y}_2^j))\lambda_1(F_p+H(\{z'\}\cup\mathcal{Y}_2^j))} \qquad (7.1)\\
&= \frac{\sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\operatorname{tr}(H_y)}{\lambda_1(F_p+H(\mathcal{Y}_2^j))\lambda_1(F_p+H(\{z'\}\cup\mathcal{Y}_2^j))}. \qquad (7.2)
\end{aligned}$$

To obtain (7.1), we let $z'\in\arg\max_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\lambda_1(F_p+H(\{y\}\cup\mathcal{Y}_2^j))$ and note that $\lambda_1(F_p+H(\{z'\}\cup\mathcal{Y}_2^j))\geq\lambda_i(F_p+H(\{y\}\cup\mathcal{Y}_2^j))$ for all $i\in\{1,2\}$ and for all $y\in\mathcal{A}\setminus\mathcal{Y}_2^j$. Next, we upper bound $f_{Pa}(\mathcal{A}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j)$ in the following manner:

$$\begin{aligned}
f_{Pa}(\mathcal{A}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j) &= \sum_{i=1}^{2}\frac{\lambda_i(F_p+H(\mathcal{A}\cup\mathcal{Y}_2^j))-\lambda_i(F_p+H(\mathcal{Y}_2^j))}{\lambda_i(F_p+H(\mathcal{Y}_2^j))\lambda_i(F_p+H(\mathcal{A}\cup\mathcal{Y}_2^j))}\\
&\leq \frac{\sum_{i=1}^{2}\left(\lambda_i(F_p+H(\mathcal{A}\cup\mathcal{Y}_2^j))-\lambda_i(F_p+H(\mathcal{Y}_2^j)\right)}{\lambda_2(F_p+H(\mathcal{Y}_2^j))\lambda_2(F_p+H(\{z'\}\cup\mathcal{Y}_2^j))} \qquad (7.3)\\
&= \frac{\sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}\operatorname{tr}(H_y)}{\lambda_2(F_p+H(\mathcal{Y}_2^j))\lambda_2(F_p+H(\{z'\}\cup\mathcal{Y}_2^j))}. \qquad (7.4)
\end{aligned}$$

To obtain (7.3), we note that $\lambda_i(F_p+H(\mathcal{A}\cup\mathcal{Y}_2^j))\geq\lambda_2(F_p+H(\mathcal{A}\cup\mathcal{Y}_2^j))\geq\lambda_2(F_p+\{z'\}\cup\mathcal{Y}_2^j)$ for all $i\in\{1,2\}$, where the second inequality follows from Lemma 5.10 with the fact $H(\mathcal{A}\cup\mathcal{Y}_2^j)-H(\{z'\}\cup\mathcal{Y}_2^j)\succeq\mathbf{0}$, and z' is defined above. Combining (7.2) and (7.4), and noting $z_j\in\arg\min_{y\in\bar{\mathcal{M}}\setminus\mathcal{Y}_2^j}\frac{\lambda_2(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}{\lambda_1(F_p+H(\{y\}\cup\mathcal{Y}_2^j))}$, we have

$$\frac{\sum_{y\in\mathcal{A}\setminus\mathcal{Y}_2^j}(f_{Pa}(\{y\}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j))}{f_{Pa}(\mathcal{A}\cup\mathcal{Y}_2^j)-f_{Pa}(\mathcal{Y}_2^j)}\geq\frac{\lambda_2(F_p+H(\mathcal{Y}_2^j))\lambda_2(F_p+H(\{z_j\}\cup\mathcal{Y}_2^j))}{\lambda_1(F_p+H(\mathcal{Y}_2^j))\lambda_1(F_p+H(\{z_j\}\cup\mathcal{Y}_2^j))}, \qquad (7.5)$$

which implies (5.43). ∎

REFERENCES

[1] *Centers for Disease Control and Prevention Test for Current Infection*, https://www.cdc.gov/coronavirus/2019-ncov/testing/diagnostic-testing.html, 2020.

[2] *Protect Purdue-Purdue University's response to COVID*-19, https://protect.purdue.edu, 2020.

[3] H. J. Ahn and B. Hassibi, *Global dynamics of epidemic spread over complex networks*, in Proceedings of the IEEE Conference on Decision and Control, 2013, pp. 4579–4585.
[4] E. Bendavid, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra, J. Tedrow, D. Tversky, A. Bogan, T. Kupiec, D. Eichner, R. Gupta, J. P.A. Ioannidis, and J. Bhattachary, *COVID-19 antibody seroprevalence in Santa Clara County, California*, (2020), https://doi.org/10.1101/2020.04.14.20062463.
[5] A. A. Bian, J. M. Buhmann, A. Krause, and S. Tschiatschek, *Guarantees for greedy maximization of non-submodular functions with applications*, in Proceedings of the International Conference on Machine Learning, 2017, pp. 498–507.
[6] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, *Epidemic thresholds in real networks*, ACM Trans. Inf. Syst. Secur., 10 (2008), pp. 1–26.
[7] S. P. Chepuri and G. Leus, *Sparsity-promoting sensor selection for non-linear measurement models*, IEEE Trans. Signal Process., 63 (2014), pp. 684–698.
[8] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms*, MIT Press, Cambridge, MA, 2009.
[9] U. Feige, *A threshold of* ln *n for approximating set cover*, J. ACM, 45 (1998), pp. 634–652.
[10] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, San Francisco, CA, 1979.
[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.
[12] A. Hota, J. Godbole, and P. Pare, *A closed-loop framework for inference, prediction, and control of SIR epidemics on networks*, IEEE Trans. Network Sci. Eng., 8 (2021), pp. 2262–2278.
[13] S. Joshi and S. Boyd, *Sensor selection via convex optimization*, IEEE Trans. Signal Process., 57 (2008), pp. 451–462.
[14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[15] D. Kempe, J. Kleinberg, and É. Tardos, *Maximizing the spread of influence through a social network*, in Proceedings of the International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146.
[16] S. Khuller, A. Moss, and J. S. Naor, *The budgeted maximum coverage problem*, Inform. Process. Lett., 70 (1999), pp. 39–45.
[17] A. Krause and C. Guestrin, *A Note on the Budgeted Maximization of Submodular Functions*, Center for Automated Learning and Discovery, Carnegie Mellon University, Pittsburgh, PA, 2005.
[18] A. Krause, A. Singh, and C. Guestrin, *Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies*, J. Mach. Learn. Res., 9 (2008), pp. 235–284.
[19] W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo, *On the dynamics of deterministic epidemic propagation over networks*, Ann. Rev. Control, 44 (2017), pp. 116–128.
[20] Y. Mo, R. Ambrosino, and B. Sinopoli, *Sensor selection strategies for state estimation in energy constrained wireless sensor networks*, Automatica J. IFAC, 47 (2011), pp. 1330–1338.
[21] M. E. Newman, *Spread of epidemic disease on networks*, Phys. Rev. E (3), 66 (2002), 016128.
[22] C. Nowzari, V. M. Preciado, and G. J. Pappas, *Analysis and control of epidemics: A survey of spreading processes on complex networks*, IEEE Control Syst. Mag., 36 (2016), pp. 26–46.
[23] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, and T. Başar, *Analysis, estimation, and validation of discrete-time epidemic processes*, IEEE Trans. Control Syst. Technol., 28 (2018), pp. 79–93.
[24] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic processes in complex networks*, Rev. Modern Phys., 87 (2015), pp. 925–979.
[25] M. Pezzutto, N. B. Rossello, L. Schenato, and E. Garone, *Smart testing and selective quarantine for the control of epidemics*, Annu. Rev. Control, 51 (2021), pp. 540–550.
[26] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas, *Optimal resource allocation for network protection against spreading processes*, IEEE Trans. Control Netw. Syst., 1 (2014), pp. 99–108.
[27] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, et al., *The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study*, The Lancet Public Health, 5 (2020), pp. E261–E270.

[28] F. Pukelsheim, *Optimal Design of Experiments*, SIAM, Philadelphia, 2006, https://doi.org/10.1137/1.9780898719109.

[29] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, Vol. 12, Springer-Verlag, New York, 2013.

[30] M. Streeter and D. Golovin, *An online algorithm for maximizing submodular functions*, in Advances in Neural Information Processing Systems, MIT Press, Cambridge, UK, 2009, pp. 1577–1584.

[31] T. H. Summers, F. L. Cortesi, and J. Lygeros, *On submodularity and controllability in complex dynamical networks*, IEEE Trans. Control Netw. Syst., 3 (2015), pp. 91–101.

[32] V. Tzoumas, L. Carlone, G. J. Pappas, and A. Jadbabaie, *LQG control and sensing co-design*, IEEE Trans. Automat. Control, 66 (2020), pp. 1468–1483.

[33] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part* I, John Wiley & Sons, New York, 2004.

[34] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part* IV: *Optimum Array Processing*, John Wiley & Sons, New York, 2004.

[35] D. Vrabac, P. E. Paré, H. Sandberg, and K. H. Johansson, *Overcoming challenges for estimating virus spread dynamics from data*, in Proceedings of the IEEE Conference on Information Sciences and Systems, 2020, pp. 1–6.

[36] L. Ye, P. E. Paré, and S. Sundaram, *Parameter Estimation in Epidemic Spread Networks Using Limited Measurements*, preprint, https://arxiv.org/abs/2105.04750, 2021.

[37] L. Ye, S. Roy, and S. Sundaram, *Resilient sensor placement for Kalman filtering in networked systems: Complexity and algorithms*, IEEE Trans. Control Netw. Syst., 7 (2020), pp. 1870–1881.

[38] L. Ye and S. Sundaram, *Sensor selection for hypothesis testing: Complexity and greedy algorithms*, in Proceedings of the IEEE Conference on Decision and Control, 2019, pp. 7844–7849.

[39] L. Ye, N. Woodford, S. Roy, and S. Sundaram, *On the complexity and approximability of optimal sensor selection and attack for Kalman filtering*, IEEE Trans. Automat. Control, 66 (2021), pp. 2146–2161.