

Estimation and Distributed Eradication of SIR Epidemics on Networks

Ciyuan Zhang, Humphrey Leung, Brooks A. Butler, and Philip E. Paré*

Abstract—This work examines a discrete-time networked SIR (susceptible-infected-recovered) epidemic model, where the infection, graph, and recovery parameters may be time-varying. We propose a stochastic framework to estimate the system states from observed testing data and provide an analytic expression for the error of the estimation algorithm. We validate some of our assumptions for the stochastic framework with real COVID-19 testing data. We identify the system parameters with the system states from our estimation algorithm. Employing the estimated system states, we provide a novel eradication strategy that guarantees at least exponential convergence to the set of healthy states. We illustrate the results via simulations over northern Indiana, USA.

I. INTRODUCTION

The main goal of epidemic model development is to identify conditions that lead to the eradication of the pathogen and then leverage the knowledge of these conditions to design mitigation strategies. Various infection models have been proposed, based on characteristics of individual pathogens, and studied in the literature, the most basic including susceptible-infected-susceptible (SIS), susceptible-infected-removed (SIR), and susceptible-infected-removed-susceptible (SIRS) [1], [2].

The SIR model is the basis of a class of widely-used epidemic models for diseases which hosts will recover with permanent or close to permanent immunity after infection [3]. Diseases belonging to this category contributed to a wide range of airborne diseases, including but not limited to Spanish Flu, SARS [4], MERS [5], Influenza [6], and COVID-19 [7]. In particular, seasonal Influenza claimed 28,000 lives in the US during the 2018-2019 flu season [8], a more extreme example of the outbreak of Spanish Flu in 1918, which caused 25–50 million deaths [9], and the more recent SARS-CoV-2 virus, which claimed 5.1 million lives and infected 255 million individuals worldwide [10].

Other SIR model variants, such as susceptible-exposed-infection-recovered (SEIR), susceptible-asymptomatic-infection-recovered (SAIR) [11], and SIDARTHE-V [12], model different immune responses and public health reactions to epidemic spread. These multi-compartmental models are accurate in estimating the propagation of specific diseases with sufficient data. However, recent research has demonstrated that viruses, such as SARS-CoV-2, are not detected accurately during the incubation period [13]. As a result, inferring the states associated with the exposed/asymptomatic phases is not possible due to the lack of credible data/widespread testing.

Therefore, in this paper, we focus on the SIR epidemic model, capturing the delay with a stochastic observation model. We expand on the SIR model by exploring mutating viruses over time-varying networks, the estimation of the underlying states, and the development of distributed eradication strategies.

The patchwork response to some of the above diseases, such as MERS [14] and COVID-19 [15], gives rise to susceptible community subpopulations, with heterogeneous time-varying factors not previously explored by the SIR model. Extensions on the SIS model, studied in [16], [17], augment the compartmental epidemic models originated in [18] to include interactions between subpopulations of susceptible communities. Additionally, various advanced epidemic models consider time-varying factors [19]–[21].

In this paper, we study the estimation of the system states of a networked SIR epidemic, where we focus on the effect of the delay between infection and testing data collection on the state estimation accuracy. The delay can be caused by factors such as the incubation period of a virus [14], [22], people's willingness to obtain a test [23], [24], different testing strategies [25], etc. We use COVID-19 as our case study for the epidemic state estimation problem, but our estimation framework can be applied to any SIR-type outbreak. The delay in onset of COVID-19 has led to large asymptomatic infectious populations, estimated between 17 – 81% [26]–[29], and the delay in test results [30] compromise the ability for accurate estimation of the current infection prevalence. An estimation algorithm that incorporated a constant delay between the change in infection proportion and testing data was introduced in [31]. They studied the inference problem by using a Bayesian approach. Inspired by the delay characterization suggested in [31], we propose a stochastic delay to model the unpredictability of incubation period of any virus and its testing strategies. We have developed methods for estimating the underlying epidemic states from testing data with a delay sampled from a geometric distribution, which cannot be completely filtered by the method suggested in [31]. We study the aggregated effect of the individual delays on the trajectory of confirmed cases and devise a method for estimating the underlying epidemic states of an SIR model from these delayed measurements. The geometric distribution which we utilize is the discrete analogy of the exponential distribution in the limit, and recall that standard compartmental models assume an exponential holding time of a host in each compartment [32]. However, our proposed stochastic framework is distinctive from the compartmental model with exponential holding time, as our model investigates the confirmed cases as the observation from the system in discrete time. We also investigate the proposed method's estimation error, which provides insights for achieving an

*Ciyuan Zhang, Humphrey Leung, Brooks A. Butler, and Philip E. Paré are with the Elmore Family School of Electrical and Computer Engineering at Purdue University. Emails: {zhan3375, leung61, brooksbutler, philpare}@purdue.edu. This work was funded in part by the C3.ai Digital Transformation Institute sponsored by C3.ai Inc. and the Microsoft Corporation and in part by the National Science Foundation, grants NSF-CNS #2028738 and NSF-ECCS #2032258.

accurate estimation of the system states. We then employ this more realistic estimation to strategically eradicate a disease.

As proven in [31], the SIR epidemic model converges to a healthy state, however, an exponential convergence is not shown. Combining the modeling and inference approach allows us to develop a distributed control algorithm capable of eradicating epidemic spread exponentially, at an equilibrium with a higher proportion of susceptible population. Decreasing the removed (recovered) and increasing the susceptibility proportion is of exceptional importance for controlling a pandemic, as long term and severe health complications have been documented in the recovered populations, including impaired cognition [33] and damage to cardiac tissue [34]. Our main result shows that by applying the estimated susceptible states of each node, our proposed eradication strategy will guarantee global exponential stability of a healthy state.

A. Paper Contributions

We summarize the main contributions of this paper:

- We propose a stochastic framework which estimates the trajectories of the system states of the networked SIR model from testing data.
- We provide a closed form solution for the error of the estimation algorithm we propose; see Proposition 1.
- We utilize real data to validate our choice of the geometric distribution for our estimation algorithm; see Section IV.
- We identify the system parameters with the estimated system states; see Section V.
- We propose a distributed eradication strategy for adjusting healing rates that is based on the inferred system states. The eradication method guarantees that the virus is eradicated at an exponential rate see Theorem 1.

B. Paper Outline

We organize this paper as follows: Section II lays down some basic assumptions and restates the well-known SIR model in the networked fashion, and it presents the main problems studied in this paper. Section III covers the proposed techniques of estimating hidden epidemics states with the stochastic delay of tested individuals and testing data. Using real COVID-19 testing data, Section IV validates the choice of the geometric distribution for the stochastic framework proposed in Section III. Section V presents the method which identifies the system parameters using the estimated system states from Section III. Section VI covers the distributed control strategy which ensures that the system converges to a healthy state exponentially fast. Section VII illustrates the results from Section III, V, and VI with numerical simulations. Finally, in Section VIII, we summarize the main conclusions of this paper and discuss future directions.

C. Notation

We denote the set of real numbers, the non-negative integers, and the positive integers as $\mathbb{R}$, $\mathbb{Z}_{\geq 0}$, and $\mathbb{Z}_{\geq 1}$, respectively. For any positive integer n , we have $[n] = \{1, 2, \dots, n\}$. The spectral radius of a matrix $A \in \mathbb{R}^{n \times n}$ is $\rho(A)$. A diagonal matrix is denoted as $\text{diag}(\cdot)$. The transpose of a vector $x \in \mathbb{R}^n$ is $x^\top$. The Euclidean norm is denoted by $\|\cdot\|$. We use I

to denote identity matrix. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the vectors whose entries all equal 0 and 1, respectively. The dimensions of the vectors are determined by context. Given a matrix A , $A \succ 0$ (resp. $A \succeq 0$) indicates that A is positive definite (resp. positive semidefinite), whereas $A \prec 0$ (resp. $A \preceq 0$) indicates that A is negative definite (resp. negative semidefinite). Let $G = (\mathbb{V}, \mathbb{E})$ denote a graph or network where $\mathbb{V} = \{v_1, v_2, \dots, v_n\}$ is the set of subpopulations, and $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$ is the set of edges. We denote the expectation of a random variable as $\mathbb{E}[\cdot]$. We use *w.p.* to represent *with probability* in equations.

II. MODEL AND PROBLEM FORMULATION

Consider a time-varying epidemic network of n subpopulations, where the size of subpopulation i is $N_i \in \mathbb{Z}_{>0}$, and the infection rates and healing rates could be time-varying. We denote $\beta_{ij}(k)$ as the infection rate from node j to node i at time step k , we denote $\gamma_i(k)$ as the healing rate of node i at time step k . The proportions of the subpopulation at node i which are susceptible, infected, and recovered at time step k are denoted by $s_i(k)$, $x_i(k)$, and $r_i(k)$, respectively. For a small sampling time $h > 0$, the discrete-time evolution of the SIR epidemic is given by

$$s_i(k+1) = s_i(k) + h[-s_i(k) \sum_{j=1}^n \beta_{ij}(k)x_j(k)], \quad (1a)$$

$$x_i(k+1) = x_i(k) + h[s_i(k) \sum_{j=1}^n \beta_{ij}(k)x_j(k) - \gamma_i(k)x_i(k)], \quad (1b)$$

$$r_i(k+1) = r_i(k) + h\gamma_i(k)x_i(k). \quad (1c)$$

Equation (1b) can be rewritten as

$$x(k+1) = x(k) + h[S(k)B(k) - \Gamma(k)]x(k), \quad (2)$$

where $S(k) = \text{diag}(s(k))$, $B(k)$ is the matrix with (i, j) th entry $\beta_{ij}(k)$, and $\Gamma(k) = \text{diag}(\gamma_i(k))$. The spread of a virus over a network can be captured using a graph $G = (\mathbb{V}, \mathbb{E})$, where $\mathbb{E} = \{(v_i, v_j) | \beta_{ij}(k) \neq 0\}$ is the set of directed edges.

We make the following assumptions in order for the system in (1) to be well defined.

Assumption 1. For every $i \in [n]$, $h\gamma_i(k) > 0$ and $\forall j \in [n]$, $\beta_{ij}(k) \geq 0$, for every $k \in \mathbb{Z}_{\geq 0}$.

Assumption 2. For every $i \in [n]$, $h\gamma_i(k) \leq 1$ and $h \sum_j \beta_{ij}(k) \leq 1$, for every $k \in \mathbb{Z}_{\geq 0}$.

We have the following result which shares the same idea as the time-invariant model, proven in [31].

Lemma 1. Suppose $s_i(0), x_i(0), r_i(0) \in [0, 1]$, $s_i(0) + x_i(0) + r_i(0) = 1$, and Assumptions 1 and 2 hold. Then, for all $k \in \mathbb{Z}_{\geq 0}$,

- 1) $s_i(k), x_i(k), r_i(k) \in [0, 1]$,
- 2) $s_i(k) + x_i(k) + r_i(k) = 1$,
- 3) $s_i(k+1) \leq s_i(k)$, and
- 4) there exists $\bar{k}$, such that $x_i(k)$ converges linearly to 0 for all $k \geq \bar{k}$, $i \in [n]$.

Definition 1. We define the set of healthy states of (1) as $\{s_i^*(k), x_i^*(k), r_i^*(k) : i \in [n], k \in \mathbb{Z}_{\geq 0}\}$, where $x_i^*(k) = 0$,

$s_i^*(k) \in [0, 1]$, $r_i^*(k) \in [0, 1]$, and $s_i^*(k) + r_i^*(k) = 1$, for all $i \in [n]$.

Given a network that is infected by a virus, our goal is to guarantee that the convergence rate of each subpopulation v_i to the set of healthy states is exponentially fast regardless of the initial conditions. From Lemma 1 we know that the well-defined SIR networked model converges to the set of healthy states linearly, on the other hand, the importance of exponential convergence can be interpreted as a speedy recovery from the outbreak and fewer individuals becoming infected over the duration of the epidemic. Here are the questions being studied in this paper:

- (i) Given the testing data, how can the stochastic framework be constructed to estimate the susceptible, infected, and recovered proportions, denoted by $\hat{s}_i(k)$, $\hat{x}_i(k)$, and $\hat{r}_i(k)$, respectively, for each subpopulation v_i in the network?
- (ii) What is the estimation error of the stochastic framework that we proposed?
- (iii) How does our proposed stochastic framework match with real data?
- (iv) How do we identify the system parameters with the estimated system states?
- (v) Given $\hat{s}_i(k)$ inferred from testing data, how can we devise a dynamic control algorithm which applies new healing rates $\hat{\gamma}_i(k)$ to each agent in (2) so that the epidemic is eradicated at an exponential rate?

III. STATE ESTIMATION FROM TESTING DATA

In this section, we study how to estimate the epidemic states $(s(k), x(k), r(k))$ from testing data in order to design a feedback controller in the following section. One of the challenges of estimating the underlying system states is that the testing data on a given day does not capture the new infections on the same day. Instead, the testing data is a delayed representation of the change in the system. Characterizing the delay of each individual is difficult, because the delay is determined by numerous factors such as the incubation period of the virus, the duration of obtaining test results, the willingness of each individual to get tested, etc. Therefore, we propose a stochastic framework in this section to capture the factors which cause the testing delay.

Definition 2. The testing delay τ_i is the length of time between when an individual from subpopulation v_i is infected and when their positive test result is reported.

In our discrete-time model, we assume that $\tau_i \in \mathbb{Z}_{\geq 0}$. We model the testing delay of each infected individual τ_i as two aggregate components to represent the uncertainty in the testing process:

$$\tau_i = \eta_i + \mathcal{Y}_i, \quad (3)$$

where $\eta_i \in \mathbb{Z}_{\geq 0}$ is a constant and $\mathcal{Y}_i$ is a discrete time random variable whose measurable space is $\mathbb{Z}_{\geq 0}$.

Remark 1. In (3), the constant component η_i can be interpreted as the length of time needed to acquire testing results. The random variable can be interpreted as the incubation

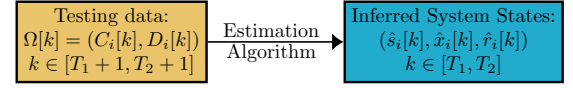


FIGURE 1: Estimation of system states from testing data, where $C_i(k)$ is the number of confirmed cases and $D_i(k)$ is the number of removed (recovered) cases at time k

period and/or the amount of time that it takes an individual to get tested after becoming infected.

First, we denote the set of estimated system states for subpopulation v_i at time k as $\hat{\Theta}_i(k) = (\hat{s}_i(k), \hat{x}_i(k), \hat{r}_i(k))$, we denote the set of testing data recorded at time k to be $\Omega_i(k) = (C_i(k), D_i(k))$, where $C_i(k)$ is the number of confirmed cases at time k , and $D_i(k)$ represents the number of removed (recovered) cases at time k . In addition, the cumulative number of confirmed and removed cases at node v_i are written as $\mathbb{C}_i(k) = \sum_{j=0}^k C_i(j)$ and $\mathbb{D}_i(k) = \sum_{j=0}^k D_i(j)$, respectively. Therefore, the number of active cases is calculated by $\mathbb{A}_i(k) = \mathbb{C}_i(k) - \mathbb{D}_i(k)$. Recall that the size of each subpopulation is N_i ; we define $c_i(k) = \frac{C_i(k)}{N_i}$ and $d_i(k) = \frac{D_i(k)}{N_i}$ as the proportion of daily confirmed cases and removal, respectively. Note that $c_i(k), d_i(k) \in [0, 1]$, for all $i \in [n], k \in \mathbb{Z}_{\geq 0}$. The estimation procedure is illustrated in Figure 1.

We then study how to relate $c_i(k)$ to the underlying states. We define a vector space Π_{T_1} as the space of all the proportions of daily number of confirmed cases, $c_i(k)$, from time step $k = T_1$ to time step $k = T_2 + 1$. We define Ξ_{T_1} as the vector of all the decreases in the proportion of susceptible individuals, $-\Delta s_i(k)$, from time step $k = T_1$ to time step $k = T_2 + 1$. We denote $\Phi(T_1, T_2)$ as the transfer matrix:

$$\Pi_{T_1} = \Phi(T_1, T_2) \Xi_{T_1}, \quad (4)$$

where $\Phi(T_1, T_2)$ is a $(T_2 - T_1 + 2) \times (T_2 - T_1 + 2)$ matrix, which depends on the SIR dynamics, the testing strategies, and the delay.

When the testing delay is a constant, i.e., $\tau_i = \eta_i > 0$, the only non-zero entries in $\Phi(T_1, T_2)$ are: $\Phi_{l+\eta_i, l} = 1, l \in [T_2 - T_1 + 2 - \eta_i]$. Since $c_i(k) = 0$ when $k \in [T_1, T_1 + \eta_i - 1]$, we can write $c_i(k)$ as

$$c_i(k) = -\Delta s_i(k - \eta_i), \quad (5)$$

for all $k \in [T_1 + \eta_i, T_2 + 1]$. When the delay $\eta_i = 0$, the transfer matrix $\Phi(T_1, T_2) = I$, as for all $k \in [T_1, T_2 + 1]$, we can write that

$$c_i(k) = -\Delta s_i(k). \quad (6)$$

We now propose a stochastic testing framework to capture the delay between when an individual is infected and when they receive a positive test result. We first let $\eta_i = 0$ in (3), without the loss of generality. Furthermore, we assume that each infected individual at node v_i has an equal probability $p_i^x \in (0, 1]$ of receiving a diagnostic test each day starting from the day after they are infected. Therefore, we model $\mathcal{Y}_i$ in (3) as a random variable following the geometric distribution, with the probability of an infected individual acquiring a positive test δ days after infection being:

$$P(\mathcal{Y}_i = \delta) = p_i^x (1 - p_i^x)^{\delta-1} \quad (7)$$

for $\delta \in \mathbb{Z}_{\geq 1}$. The geometric distribution of the testing delay models the number of days before an infected individual obtains a diagnostic test which represents the incubation period of the virus and/or the unwillingness of each individual getting a test. We now make the following assumptions for our stochastic model:

- We assume that the delay distributions are independent and identically distributed (i.i.d.) for each infected individual. In other words, the delay of one infected individual does not effect other infected individuals' delays nor does it change due to the delays of other patients, and each infected individual has an identical delay distribution.
- We assume that only the infected individuals obtain a test and the susceptible individuals who are not infected do not acquire a test; this assumption aims to represent the shortage of testing equipment in the early stage of an unknown virus outbreak. Alternatively, we could assume that the negative tests are not reported by test organizations during the data collection.
- Furthermore, we assume that an infected individual can be tested only once due to the deficiency of available tests at the beginning of the spread. In our proposed framework, duplicate test results would double count infected individuals. Alternatively, we could assume that once an infected individual gets a positive test result, any subsequent test results will not be included in the daily counts.
- We assume that all the tests collected are PCR (Polymerase Chain Reaction) tests and an infected individual has equal probability of acquiring a PCR test before and after recovery.
- In addition, we assume that a recovered individual will still test positive after recovering from the virus. Using COVID-19 as an example, patients would still test positive with PCR tests after recovering from the SARS-CoV-2 virus, albeit showing no symptoms, and being not contagious [35].
- We also assume that all the tests generate accurate results.

We will justify our choice of geometric distribution with real COVID-19 data from [13] in Section IV.

We now relate the proportion of confirmed cases $c_i(k)$ with the underlying states of the system. We define a binary random variable $\mathcal{X}_i(\nu)$, with $\mathcal{X}_i(\nu) = 1$ (resp. $\mathcal{X}_i(\nu) = 0$) if a randomly chosen individual from subpopulation v_i became (resp. did not become) infected at time ν . Its pmf can be written as

$$\mathcal{X}_i(\nu) = \begin{cases} 1 & \text{w.p. } -\Delta s_i(\nu) \\ 0 & \text{w.p. } 1 + \Delta s_i(\nu), \end{cases} \quad (8)$$

where, from (1), $-\Delta s_i(\nu) = s_i(\nu - 1) - s_i(\nu) = h s_i(\nu - 1) \sum_j \beta_{ij} x_j(\nu - 1) \geq 0$ for all $\nu \geq 0$.

We define the binary random variable $\mathcal{T}_i(\mu, \delta)$, with $\mathcal{T}_i(\mu, \delta) = 1$ if a randomly chosen individual acquired a positive test at time μ and was infected δ days before μ . Now we rewrite ν , in (8), as $\mu - \delta$. From (7), the conditional probability $P(\mathcal{T}_i(\mu, \delta) = 1 | \mathcal{X}_i(\mu - \delta) = 1)$ is given by the geometric probability mass function (pmf) $p_i^x(1 - p_i^x)^{\delta-1}$ and

represents the probability of an infected individual acquiring a positive test specifically δ days after being infected. Hence, the joint pmf of the two random variables $\mathcal{X}_i(\mu - \delta)$ and $\mathcal{T}_i(\mu, \delta)$ is written as:

$$P_{\mathcal{X}_i, \mathcal{T}_i}(\mu - \delta, \mu) = P(\mathcal{X}_i(\mu - \delta) \cap \mathcal{T}_i(\mu, \delta)), \quad (9)$$

which is interpreted as the probability that a randomly chosen individual became infected at time $\mu - \delta$ and acquired a positive test at time μ , where $\mu - \delta, \mu \in [T_1, T_2]$.

Therefore, the joint pmf $P_{\mathcal{X}_i, \mathcal{T}_i}(\mu - \delta, \mu)$ is calculated as:

$$\begin{aligned} P_{\mathcal{X}_i, \mathcal{T}_i}(\mathcal{X}_i(\mu - \delta) = 1 \cap \mathcal{T}_i(\mu, \delta) = 1) \\ = P(\mathcal{T}_i(\mu, \delta) = 1 | \mathcal{X}_i(\mu - \delta) = 1) P(\mathcal{X}_i(\mu - \delta) = 1) \\ = p_i^x(1 - p_i^x)^{\delta-1} [-\Delta s_i(\mu - \delta)], \end{aligned} \quad (10)$$

$$\begin{aligned} P_{\mathcal{X}_i, \mathcal{T}_i}(\mathcal{X}_i(\mu - \delta) = 0 \cap \mathcal{T}_i(\mu, \delta) = 1) \\ = P(\mathcal{T}_i(\mu, \delta) = 1 | \mathcal{X}_i(\mu - \delta) = 0) P(\mathcal{X}_i(\mu - \delta) = 0) \\ = 0[1 + \Delta s_i(\mu - \delta)] = 0, \end{aligned} \quad (11)$$

since we assume that the test results are accurate. Similarly,

$$\begin{aligned} P_{\mathcal{X}_i, \mathcal{T}_i}(\mathcal{X}_i(\mu - \delta) = 1 \cap \mathcal{T}_i(\mu, \delta) = 0) \\ = [1 - p_i^x(1 - p_i^x)^{\delta-1}] [-\Delta s_i(\mu - \delta)], \end{aligned}$$

$$P_{\mathcal{X}_i, \mathcal{T}_i}(\mathcal{X}_i(\mu - \delta) = 0 \cap \mathcal{T}_i(\mu, \delta) = 0) = 1 + \Delta s_i(\mu - \delta).$$

Let $\mathcal{W}_i(\mu)$ be the marginal distribution of $\mathcal{T}_i(\mu, \delta)$ over the set of feasible delays, δ , with its pmf being the probability of a random individual acquiring a positive test at time μ :

$$\begin{aligned} P_{\mathcal{W}_i}(\mathcal{W}_i(\mu) = 1) \\ = \sum_{\delta=1}^{\mu-T_1} P_{\mathcal{X}_i, \mathcal{T}_i}(\mathcal{X}_i(\mu - \delta) = 1 \cap \mathcal{T}_i(\mu, \delta) = 1) \\ = \sum_{\delta=1}^{\mu-T_1} p_i^x(1 - p_i^x)^{\delta-1} [-\Delta s_i(\mu - \delta)], \end{aligned}$$

by combining (10) and (11). Therefore, the expected number of confirmed cases at time k is calculated as

$$\begin{aligned} C_i(k) &= \mathbb{E} \left[\sum_{l=1}^{N_i} \mathcal{W}_i(k) \right] \\ &= \sum_{l=1}^{N_i} \mathbb{E}[\mathcal{W}_i(k)] \end{aligned} \quad (12)$$

$$= N_i \sum_{\delta=1}^{k-T_1} p_i^x(1 - p_i^x)^{\delta-1} [-\Delta s_i(k - \delta)], \quad (13)$$

where (12) holds because of the linearity of expectation and since the testing delays are i.i.d.. We now connect the daily expected proportion of confirmed cases with the actual proportion of confirmed cases at time k , by first demonstrating that the $\mathcal{W}_i(\mu)$ are i.i.d..

Lemma 2. $\mathcal{W}_1(\mu), \dots, \mathcal{W}_{N_i}(\mu)$ are i.i.d. when the size of subpopulation approaches infinity.

Proof. The distribution of $\mathcal{W}_i(\mu)$ is a function of the random variables: $\mathcal{X}_i(\nu)$ and $\mathcal{T}_i(\mu, \delta)$. We know that $\mathcal{T}_i(\mu, \delta)$, for all $i \in [N_i]$, are i.i.d. by assumption. We also obtain that $\mathcal{X}_i(\nu)$ are i.i.d. for each individual, because as the number

of subpopulation N_i goes to infinity, the probability of a randomly chosen individual testing positive should be equal for each individual and should not affect others. $\square$

Lemma 3. *By assuming $c_i(k) = 0$ for all $k \notin [T_1 + 1, T_2 + 1]$, we obtain that the proportion of daily new cases almost surely equals*

$$c_i(k) = p_i^x(-\Delta s_i(k-1)) + (1 - p_i^x)c_i(k-1), \quad (14)$$

as the subpopulation size approaches infinity, where $k \in [T_1 + 1, T_2 + 1]$.

Proof. We denote $\bar{W}_{N_i}(k)$ as the actual proportion of confirmed cases at time k , which can be calculated as:

$$\bar{W}_{N_i}(k) = \frac{1}{N_i} \sum_{i=1}^{N_i} \mathcal{W}_i(k). \quad (15)$$

Based on the strong law of large numbers [36], Lemma 2, and the definition of $c_i(k) = \frac{C_i(k)}{N_i}$, we obtain that the average proportion of daily confirmed cases converges almost surely to the expected value when the size of the subpopulation goes to infinity, namely:

$$Pr\left(\lim_{N_i \rightarrow \infty} \frac{1}{N_i} \sum_{i=1}^{N_i} \mathcal{W}_i(k) = c_i(k)\right) = 1, \quad (16)$$

where we can interpret $\lim_{N_i \rightarrow \infty} \frac{1}{N_i} \sum_{i=1}^{N_i} \mathcal{W}_i(k)$ as the actual proportion of confirmed cases at time k when $N_i \rightarrow \infty$. Therefore, we obtain that as the subpopulation size goes to infinity for all $i \in [n]$, the actual proportion is equal to the expected value almost surely:

$$Pr\left(\lim_{N_i \rightarrow \infty} \bar{W}_{N_i}(k) = \sum_{\delta=1}^{k-T_1} p_i^x(1 - p_i^x)^{\delta-1}[-\Delta s_i(k-\delta)]\right) = 1. \quad (17)$$

Hence, we acquire that

$$\begin{aligned} c_i(k) &= \sum_{\delta=1}^{k-T_1} p_i^x(1 - p_i^x)^{\delta-1}[-\Delta s_i(k-\delta)] \\ &= p_i^x(-\Delta s_i(k-1)) \\ &\quad + \sum_{\delta=2}^{k-T_1} p_i^x(1 - p_i^x)^{\delta-1}[-\Delta s_i(k-\delta)] \\ &= p_i^x(-\Delta s_i(k-1)) \\ &\quad + (1 - p_i^x) \left\{ \sum_{\delta=2}^{k-T_1} p_i^x(1 - p_i^x)^{\delta-2}[-\Delta s_i(k-\delta)] \right\} \\ &= p_i^x(-\Delta s_i(k-1)) + (1 - p_i^x)c_i(k-1) \end{aligned} \quad (18)$$

almost surely for all $k \in [T_1 + 1, T_2 + 1]$, as the subpopulation size N_i goes to infinity, where (18) follows from the definition of $c_i(k-1)$. $\square$

Finally, we relate the the daily number of recoveries, i.e., $D_i(k)$, with the underlying states. In the data collected, $D_i(k)$ corresponds to the change in the number of recovered individuals and the total number of known active cases $\mathbb{A}_i(k-1)$. We assume

$$D_i(k) \sim \text{Bin}\left(\mathbb{A}_i(k-1), h\gamma_i(k-1)\right). \quad (19)$$

Namely, each known active case recovers with healing rate $h\gamma_i(k-1)$.

The above analysis links the collected data proportions with the underlying states of the system. If we acquire the parameter: p_i^x , we will be able to estimate the state systems.

Definition 3. *We assume that: $\hat{x}_i(k) = \hat{x}_i(0)$, $\hat{r}_i(k) = \hat{r}_i(0)$, and $\hat{s}_i(k) = \hat{s}_i(0)$, where $\hat{x}_i(0), \hat{r}_i(0), \hat{s}_i(0) \in [0, 1]$ for all $i \in [n], k < T_1$. Given the testing data set $\Omega_i(k)$ collected from time step $T_1 + 1$ to $T_2 + 1$, according to (14), we define the estimated proportion of new infections at node v_i as*

$$-\widehat{\Delta s}_i(k) = \frac{c_i(k+1) - (1 - p_i^x)c_i(k)}{p_i^x}, k \in [T_1, T_2]. \quad (20)$$

Notice that when $p_i^x = 1$, (20) becomes:

$$-\widehat{\Delta s}_i(k) = c_i(k+1), k \in [T_1, T_2],$$

which can be interpreted as: every infected individual will be tested the day after being infected. Hence, the estimated change in proportion of infection on a given day k exactly equals to the fraction of the number of positive cases on the next day $k+1$.

Moreover, we let $\widehat{\Delta r}_i(k) = 0$ for $k = T_1$. Note that the following equality holds from the formulation of the SIR model:

$$\widehat{\Delta s}_i(k) + \widehat{\Delta x}_i(k) + \widehat{\Delta r}_i(k) = 0.$$

We further define, for $k \in [T_1, T_2]$,

$$\begin{aligned} \hat{s}_i(k) &= \hat{s}_i(k-1) + \widehat{\Delta s}_i(k), \\ \hat{x}_i(k) &= \hat{x}_i(k-1) + \widehat{\Delta x}_i(k), \\ \hat{r}_i(k) &= \hat{r}_i(k-1) + \widehat{\Delta r}_i(k). \end{aligned} \quad (21)$$

Similar to the proof of Lemma 3, based on (19) and the strong law of large numbers [36], when the number of active cases $\mathbb{A}_i \rightarrow \infty$, i.e., the size of subpopulation i , N_i approaches infinity, the number of daily removed cases $D_i(k)$ is almost surely equal to the expected value of the binomial distribution in Eq. (19): $h\gamma_i(k-1)\mathbb{A}_i(k-1)$. Hence, we acquire that the number of daily removed cases $D_i(k)$ and the expected value $h\gamma_i(k-1)\mathbb{A}_i(k-1)$ are almost surely equivalent, and $h\gamma_i(k-1)$ can be calculated by $D_i(k)/\mathbb{A}_i(k-1)$ when the subpopulation size goes to infinity for all $i \in [n]$. Therefore, the change in the proportion of recovered individuals at node v_i can be inferred as

$$\widehat{\Delta r}_i(k) = \frac{D_i(k)}{\mathbb{A}_i(k-1)} \hat{x}_i(k-1), k \in [T_1 + 1, T_2], \quad (22)$$

where $\hat{x}_i(k-1)$ is calculated from (20) and (21). When $\mathbb{A}_i(k-1) = 0$, we assume $\widehat{\Delta r}_i(k) = 0$.

Therefore, if the testing data $\Omega_i(k)$ is available over an interval $k \in [T_1 + 1, T_2 + 1]$, we can estimate the states of the system by repetitively applying (20), (21), and (22) with the initial conditions, i.e., $\hat{s}_i(0)$, $\hat{x}_i(0)$, and $\hat{r}_i(0)$, assumed for the geometric distribution model. This addresses question (i) in Section II.

Assumption 3. *We assume that $c_i(k) = 0$ for all $k \in [T_1] \cup \{0\}$ and the estimated initial susceptible proportion is indicated as $\hat{s}_i(0)$.*

Assumption 3 can be interpreted as follows: even though there possibly exist positive infection cases in subpopulation i

before the testing starts, we may not be able to collect all the data on these positive cases. We now present the close form solution for the estimation error that accounts for this discrepancy.

Proposition 1. *Under Assumption 3, the error of the inference method in (20)-(22), for all $k \geq T_1$, is almost surely:*

$$|\hat{s}_i(k) - s_i(k)| = \left| \hat{s}_i(0) - s_i(0) - \sum_{l=1}^{T_1-1} \Delta s_i(l) \right|, \quad (23)$$

as the subpopulation size N_i approaches infinity.

Proof. From (1a), we first represent $s_i(k)$ by:

$$s_i(k) = s_i(0) + \sum_{l=1}^k \Delta s_i(l). \quad (24)$$

Now, we characterize $\hat{s}_i(k)$:

$$\begin{aligned} \hat{s}_i(k) &= \hat{s}_i(0) + \sum_{l=1}^k \widehat{\Delta s}_i(l) \\ &= \hat{s}_i(0) + \sum_{l=T_1}^k \widehat{\Delta s}_i(l) \end{aligned} \quad (25)$$

$$= \hat{s}_i(0) - \frac{c_i(k+1)}{p_i^x} - \sum_{l=T_1+1}^k c_i(l), \quad (26)$$

where (25) is written because $-\widehat{\Delta s}_i(l) = 0$ for all $l \leq T_1 - 1$ in (20). We acquired (26) through representing each $\widehat{\Delta s}_i(l)$, $l \geq T_1$ by (20) and following Assumption 3. By applying (14), we calculate the $\sum_{l=T_1+1}^k c_i(l)$ on the R.H.S. of (26) as

$$\begin{aligned} \sum_{l=T_1+1}^k c_i(l) &= -p_i^x \sum_{l=T_1}^{k-1} \Delta s_i(l) + (1 - p_i^x) \sum_{l=T_1+1}^{k-1} c_i(l) \\ &= -p_i^x \sum_{l=T_1}^{k-1} \Delta s_i(l) + (1 - p_i^x) \left[\sum_{l=T_1+1}^k c_i(l) - c_i(k) \right], \end{aligned} \quad (27)$$

since $\sum_{l=T_1+1}^{k-1} c_i(l) = \sum_{l=T_1+1}^k c_i(l) - c_i(k)$. We can reorganize (28) and acquire:

$$\sum_{l=T_1+1}^k c_i(l) = - \sum_{l=T_1}^{k-1} \Delta s_i(l) - \frac{1 - p_i^x}{p_i^x} c_i(k). \quad (28)$$

Hence, we replace $\sum_{l=T_1+1}^k c_i(l)$ on the R.H.S. of (26) with (29) and obtain:

$$\hat{s}_i(k) = \hat{s}_i(0) - \frac{c_i(k+1)}{p_i^x} + \sum_{l=T_1}^{k-1} \Delta s_i(l) + \frac{1 - p_i^x}{p_i^x} c_i(k) \quad (30)$$

$$= \hat{s}_i(0) + \sum_{l=T_1}^k \Delta s_i(l), \quad (31)$$

where (31) follows from writing $c_i(k+1)$ in (30) as $p_i^x(-\Delta s_i(k)) + (1 - p_i^x)c_i(k)$, using (14). Therefore, we can calculate $|\hat{s}_i(k) - s_i(k)|$ by comparing (24) with (31) and yield the result. $\square$

Prop. 1 provides a closed form solution to the estimation error given the initial susceptible level assumed and the start

testing time. Hence, Prop. 1 solves question (ii) in Section II. Note that in Prop. 1, the time interval is $k \geq T_1$ instead of $k \in [T_1, T_2]$, as there may not be an end time for data collection, hence $T_2 \rightarrow \infty$ in this case.

Corollary 1. *In Prop. 1, if $\hat{s}_i(0) \geq s_i(0)$, then $\hat{s}_i(k) \geq s_i(k)$, for all $k \geq T_1$. Moreover, if $\hat{s}_i(0) = s_i(0)$, and $T_1 = 1$, then $\hat{s}_i(k) = s_i(k)$, for all $k \geq T_1$.*

Corollary 2. *Under Assumption 3,*

$$|\hat{s}_i(k) - s_i(k)| = |\hat{s}_i(0) - s_i(T_1 - 1)|, \quad (32)$$

for all $k \geq T_1$.

Remark 2. *The error expression presented in Corollary 2 consists of two parts: $\hat{s}_i(0)$ and $s_i(T_1 - 1)$. The first component depends on the initial condition for the estimation algorithm. The second component depends on the day we begin testing. Therefore, more data does not compensate the loss of measurement in the change of susceptible level, it is crucial to begin testing early for an accurate estimation of the states of an outbreak.*

We will explore this error via simulations in Section VII.

By estimating the proportion of infected individuals in a subpopulation of a network, we are able to acquire the estimation of the infection prevalence in the whole system. These inferred states provide an understanding of the epidemic and important factors for designing eradication schemes for infectious diseases.

IV. VALIDATION OF GEOMETRIC DISTRIBUTION

In this section, we employ four sets of real COVID-19 data [37]–[40] to justify the choice of stochastic observation model proposed in Section III and learn the parameters of the geometric distribution for all data sets. We generate a random sample of 100 observations from a geometric distribution with a delay η_i and compare the sample mean to the real data. We then find the optimal parameters for each data set by solving the following minimization problem:

$$\min_{\eta_i, p_i^x} \sum_{k=0}^{T_i} \|C_i(k) - \hat{C}_i(k)\|^2, \quad (33)$$

where T_i represents the number of days of each data set, $C_i(k)$ is the number of cases on the k -th day in the i -th data set, and $\hat{C}_i(k)$ is the average number of cases on the k -th day from the 100 Monte Carlo simulations.

In Figure 2, we calculate the optimal parameters for each data set by solving Eq. (33). From Figure 2, we can see that the data generated by the geometric distribution captures the behavior of the real collected data, thus, our stochastic framework can model practical scenarios sufficiently well.

In Table I, we evaluate the fit of the geometric distribution by calculating the optimal cost function for each data set. We see that the data set from [38] has the best fit (under this metric). Moreover, we can see that the delay η_i and p_i^x generated from each data set are distinct from those generated from the other data sets. The different values of p_i^x, η_i for each data set validate that distinct regions have heterogeneous

Data set	[37]	[38]	[39]	[40]
$\sum_{k=0}^{T_i} \ C_i(k) - \hat{C}_i(k)\ ^2$	461.43	34.16	173.78	174.74
$\frac{\sum_{k=0}^{T_i} \ C_i(k) - \hat{C}_i(k)\ ^2}{\sum_k C_i(k)}$	2.24	0.61	1.66	1.12
η_i	3	2	1	2
p_i^x	0.110	0.130	0.069	0.074

TABLE I: The performance of using the geometric distribution for fitting each data set in Figure 2 and the learned optimal parameters for each data set

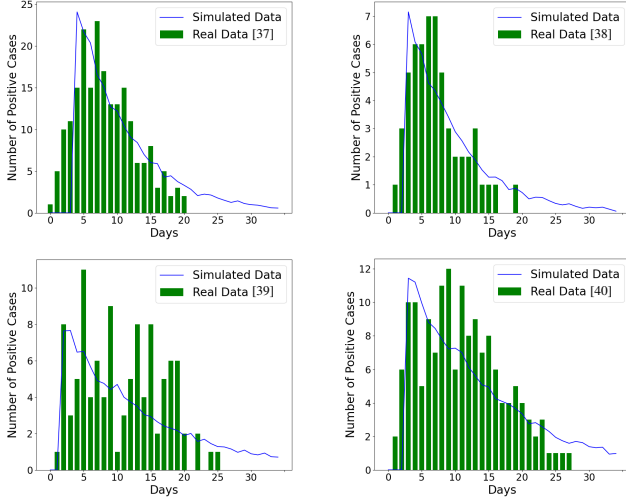


FIGURE 2: Comparing the real COVID-19 cases data with the cases generated by the geometric distribution proposed in (7) with a delay η_i . The learned parameters are: $\eta_1 = 3, p_1^x = 0.110$ (upper left); $\eta_2 = 2, p_2^x = 0.130$ (upper right); $\eta_3 = 1, p_3^x = 0.069$ (lower left); $\eta_4 = 2, p_4^x = 0.074$ (lower right). The corresponding costs can be found in Table I

testing delays caused by shortages of testing kits in the early outbreak, individuals' unwillingness to obtain a test, and/or the incubation period of patients. Therefore, the stochastic modeling framework proposed in Section III is a practical method for capturing the delay between the infection and testing data collection times and it enables the estimation of the system states networked SIR epidemics when the system parameters are unknown.

V. PARAMETER IDENTIFICATION

Our control strategy proposed in the next section assumes the infection parameters are known. Leveraging the ideas from [41], we can estimate the spreading parameters from time series data. Assuming we have time series data for $k \in [T_1, T_2]$, the parameters are static, and by factoring β_{ij} into $\beta_i a_{ij}$, the dynamics in (1) can be rewritten as

$$\begin{bmatrix} x_i(T_1 + 1) - x_i(T_1) \\ \vdots \\ x_i(T_2) - x_i(T_2 - 1) \\ r_i(T_1 + 1) - r_i(T_1) \\ \vdots \\ r_i(T_2) - r_i(T_2 - 1) \end{bmatrix} = \begin{bmatrix} -s_i(T_1) \sum_{j=1}^n a_{ij} x_j(T_1) & -x_i(T_1) \\ \vdots & \vdots \\ -s_i(T_2 - 1) \sum_{j=1}^n a_{ij} x_j(T_2 - 1) & -x_i(T_2 - 1) \\ 0 & x_i(T_1) \\ \vdots & \vdots \\ 0 & x_i(T_2 - 1) \end{bmatrix} \begin{bmatrix} h\beta_i \\ h\gamma_i \end{bmatrix}. \quad (34)$$

Therefore, as long as the data matrix on the right-hand side is full column rank and we know the a_{ij} 's, the spreading

parameters for subpopulation i can be uniquely identified [41]. When there is noise in the state measurements, parameter identification is less accurate, but still can produce viable results [42]. We illustrate how the parameter identification performs when there is error in the estimated states via simulations in Section VII.

VI. DISTRIBUTED ERADICATION STRATEGY

In this section we focus on how to eradicate the virus at an exponential rate by adjusting the healing rates. This approach can be understood as boosting the healing rate of each subpopulation separately by providing effective medication, medical supplies, and/or healthcare workers. For example, research has quantified the reduction of recovery/removal time from COVID-19 with certain types of treatment such as taking baricitinib plus remdesivir, receiving noninvasive ventilation, or inhaling high-flow oxygen [43]. In addition, monoclonal antibodies were also confirmed to be able to speed up the recovery from COVID-19 as since they targeted the SARS-CoV-2 virus directly [44]. Moreover, as the healing rate can represent the inverse of the average duration of being infectious, the healing rates can be increased by enforcing isolation and quarantine for the patients so that they are not able to infect other individuals. These are just several examples of special medicines/treatments for the SARS-CoV-2 virus; there are validated methods for other infectious diseases as well. For instance, when controlling the epidemic spread over animals, rapid livestock destruction of all the infected animals is considered as an effective removal method [45]. The control's literature on networked epidemics often employs the healing rate as an actuator [21], [46], [47].

We propose a distributed strategy that employs the estimated states and guarantees the eradication of the virus in at least exponential time. We propose the following healing rate to control the epidemic spread over the network:

$$\tilde{\gamma}_i(k) = \hat{s}_i(k) \sum_{j=1}^n \beta_{ij}(k) + \epsilon_i, \quad i \in [n], \quad (35)$$

where $\hat{s}_i(k)$ is the inferred susceptible rate from (21) and $\epsilon_i > 0$, for each $i \in [n]$.

Theorem 1. Consider the system in (1) and assume that

- 1) $0 \leq h \sum_j \beta_{ij}(k) < 1, \forall i \in [n] \text{ and } \forall k \in \mathbb{Z}_{\geq 0}$,
- 2) $B(k)$ is symmetric and irreducible $\forall k \in \mathbb{Z}_{\geq 0}$,
- 3) $\exists \epsilon_i > 0$ such that $h\tilde{\gamma}_i(k) < 1, \forall i \in [n] \text{ and } \forall k \in \mathbb{Z}_{\geq 0}$,
- 4) $\hat{s}_i(0) \geq s_i(0), \forall i \in [n]$.

Then (35) guarantees GES of the set of healthy states.

Proof. By substituting (35) into (1), we obtain

$$\begin{aligned} x_i(k+1) &= x_i(k) + \\ &h \left\{ s_i(k) \sum_{j=1}^n \beta_{ij}(k) x_j(k) - [\hat{s}_i(k) \sum_{j=1}^n \beta_{ij}(k) + \epsilon_i] x_i(k) \right\}. \end{aligned} \quad (36)$$

The state transition matrix of (36) can be written as

$$\tilde{M}(k) = I + h[S(k)B(k) - (\hat{S}(k)\text{diag}(B(k)\mathbf{1}_{n \times 1}) + \text{diag}(\epsilon_i))], \quad (37)$$

where $\hat{S}(k) = \text{diag}(\hat{s}_i(k))$. For any $i, j \in [n], j \neq i$, the entries of the i -th row of $\tilde{M}(k)$ are

$$\tilde{m}_{ii}(k) = 1 + h[s_i(k)\beta_{ii}(k) - \hat{s}_i(k) \sum_{j=1}^n \beta_{ij}(k) - \epsilon_i], \quad (38)$$

$$\tilde{m}_{ij}(k) = h s_i(k) \beta_{ij}(k), \quad (39)$$

which satisfies the following inequality

$$\tilde{m}_{ii}(k) + \sum_{j \neq i} \tilde{m}_{ij}(k) \leq 1 - h \min\{\epsilon_i\}, \forall i \in n, \quad (40)$$

since from Corollary 1 we know that when we assume that $\hat{s}_i(0) \geq s_i(0)$, $\forall i \in [n]$ and $\forall k \in \mathbb{Z}_{\geq 0}$, we obtain $\hat{s}_i(k) \geq s_i(k)$ for all $i \in [n]$ and $k \in \mathbb{Z}_{\geq 0}$. Therefore, by Gershgorin circle theorem, the spectral radius of $\tilde{M}(k)$ is upper bounded by $1 - h \min\{\epsilon_i\}$:

$$\rho(\tilde{M}(k)) \leq 1 - h \min\{\epsilon_i\}. \quad (41)$$

Since we have $x(k+1) = \tilde{M}(k)x(k)$ and $x(k) \geq 0$ for all k , we can write that $\|x(k+1)\| \leq [1 - h \min\{\epsilon_i\}]\|x(k)\|$ for all k . Since $\epsilon_i > 0, \forall i \in n$, we obtain that, for all $x_i(0) \in [0, 1]^n$,

$$\|x(k)\| \leq [1 - h \min\{\epsilon_i\}]^k \|x(0)\| \leq e^{-kh \min\{\epsilon_i\}} \|x(0)\|, \quad (42)$$

where the second inequality holds by Bernoulli's inequality [48],

$$e^x = \lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n \geq 1 + x. \quad (43)$$

Hence, $x(k)$ converges to $\mathbf{0}$ with an exponential rate of at least $h \min\{\epsilon_i\}$. Therefore, the set of healthy states is GES. $\square$

Remark 3. The control strategy proposed in Theorem 1 can be interpreted as follows: if the healing rate of each subpopulation is appropriately increased according to its estimated susceptible proportion, for example by distributing effective medication, medical supplies, and/or healthcare workers to each subpopulation, then the epidemic will be eradicated with at least an exponential rate. This theorem provides decision makers insight into, given sufficient resources, how to allocate medical supplies and healthcare workers to different subpopulations so that the epidemic can be eradicated quickly. Furthermore, Theorem 1 provides sufficient conditions for guaranteeing an exponentially decreasing $\|x(k)\|$ for all k when the conditions apply. In other words, implementing the control strategy in Theorem 1 at full length will prevent the potential upcoming waves of the epidemic in the 2-norm sense of $x(k)$. We note that the implementation of (35) relies on the knowledge of infection rates $\beta_{ij}(k)$ for all $k \in \mathbb{Z}_{\geq 0}$, which can be learned from the method in Eq. (34), and we will illustrate that the distributed eradication works efficiently with the identified parameter $\hat{\beta}_{ij}(k)$ in Section VII.

Theorem 1 has proven that given the estimated susceptible state the distributed eradication strategy proposed eradicates the virus with at least an exponential rate. Therefore, question (v) from Section II has been addressed here.

In this section, we have presented a distributed eradication strategy based on the estimated system states. The strategy ensure that the SIR epidemics converge to the sets of healthy states exponentially. We illustrate the eradication strategy with numerical simulations in Section VII, and study how a system will react if the eradication strategy is removed too early.

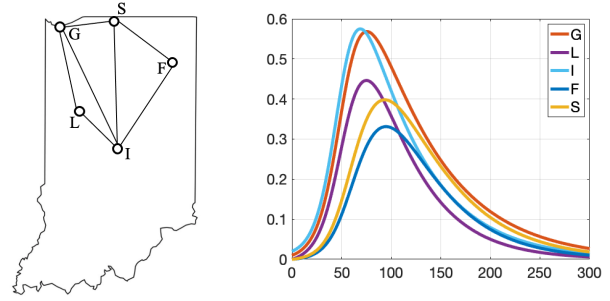


FIGURE 3: Graph topology in the map of the state of Indiana [49] analyzed and the evolution of infected proportion in each city

β_{ij}	G	L	I	F	S
G	0.08	0.15	0.24	0	0.06
L	0.15	0.12	0.13	0	0
I	0.24	0.13	0.25	0.05	0.04
F	0	0	0.05	0.11	0.15
S	0.06	0	0.04	0.14	0.09
γ_i	0.075	0.115	0.085	0.125	0.1
N_i	50000	16000	90000	35000	30000

TABLE II: Simulation Network Parameters

VII. SIMULATIONS

In this section, we simulate a virus spreading over a static network with 5 nodes in Fig 3 to illustrate our results. The nodes are modeled after the five metropolitan areas with a population over 150,000 in northern Indiana, U.S.: Gary (G), Lafayette (L), Indianapolis (I), Fort Wayne (F) and South Bend (S). Two nodes are neighbors if there is a major highway connecting them. We set the initially infected proportion to be 0.02 at node I and 0.01 at node G and 0 elsewhere. The infection rates, healing rates, and the size of each subpopulation are static and presented in Table. II. The evolution of the infected proportion for each city is shown in Figure 3.

Considering the stochastic framework, we simulate testing data using (14) and (19), with $p_i^x = 0.2, \forall i \in \{G, L, I, F, S\}$ from $T_1 = 6$ to $T_2 = 300$. The number of daily and cumulative confirmed cases and removed (recovered) cases over time at node L are shown in Fig. 4. When $k \geq 80$, the proportion of infected individuals at node L begins to decrease in Fig. 3, which reduces the number of active cases in Fig. 4.

We now use the method proposed in Section III to estimate the susceptible proportion at node I. We assume that the initial

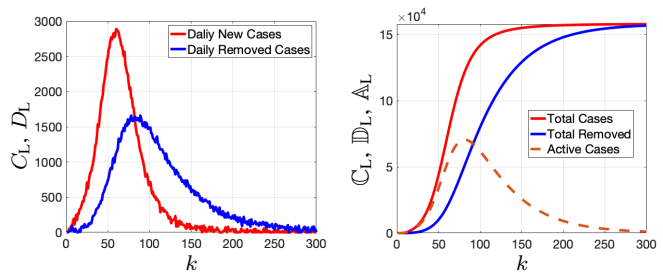


FIGURE 4: Simulated daily and cumulative number of cases for node L with $p_i^x = 0.2$

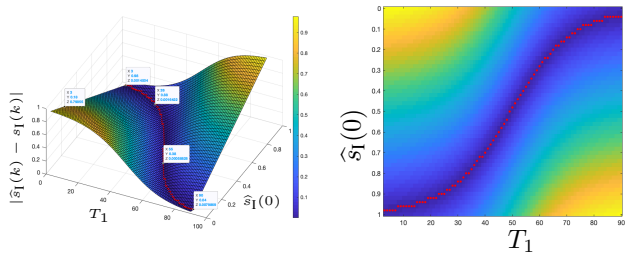


FIGURE 5: The absolute value of susceptible state estimation error at $k = 100$ with respect to start testing date and the initial susceptible level assumed at node I, where the red points represent the estimation error: $|\hat{s}_1(k) - s_1(k)| < 0.01$. Both of the plots are illustrations of Prop. 1

condition of the recovered state is $\hat{r}_1(0) = 0$. Hence, the initial infected state is written as: $\hat{x}_1(0) = 1 - \hat{s}_1(0)$. In Fig. 5, we plot the absolute value of the estimation error of the susceptible state at $k = 100$ versus the start testing time T_1 and initial condition assumed, $\hat{s}_1(0)$. It can be seen in Fig. 5 (left) that the estimation error increases linearly with the initial susceptible level assumed. When the initial condition is assumed correctly for node I, with a later start testing date, the estimation error at $k = 100$ builds up from 0 to $r_1(k)$ eventually. The increase in the estimation error with T_1 signifies the importance of an early testing during an outbreak: with appropriate initial conditions assumed, we should initiate testing as quickly as possible to improve the accuracy of the state estimation. Further, we can see from Fig. 5 that: if we start collecting the testing data late, we must compensate by assuming lower initial conditions for the susceptible proportion in order to acquire accurate estimation results. The intuition behind this finding is that since, by Definition 3, $\hat{s}_i(k) = \hat{s}_i(0)$, for all $k < T_1$, the lower initial condition can compensate for missed tests from $k \in [0, T_1 - 1]$, captured by the last term in (23). However, guessing $\hat{s}_i(0)$ correctly, namely $\hat{s}_i(0) = s_i(0) + \sum_{l=1}^{T_1-1} \Delta s_i(l)$, for $T_1 > 0$ is quite difficult. Additionally, if we assume that $\hat{s}_i(0) = 1$, the estimated $\hat{s}_i(k)$ is always larger than the true susceptible state in Fig. 5. The overestimation of susceptible level encourages us to design a stronger strategy to eradicate the virus, as will be seen in the subsequent simulations.

A. Stochastic Estimator vs Luenberger Observer

We now provide a thorough numerical comparison between a traditional observer design and the stochastic framework which we proposed in Section III. We build a traditional state observation model:

$$\begin{aligned} \hat{x}_i(k+1) = & \hat{x}_i(k) + h \left\{ \hat{s}_i(k) \sum_{j=1}^n \beta'_{ij} \hat{x}_j(k) - \gamma'_i \hat{x}_i(k) \right\} \\ & + L_i [y_i(k) - \hat{y}_i(k)], \end{aligned} \quad (44)$$

where $\hat{y}_i(k) = \hat{x}_i(k)$ and $y_i(k) = x_i(k)$ are our estimated and observed infection level at time k , respectively, β'_{ij} and γ'_i are our assumed infection and recovery parameters, respectively, L_i is the observer gain such that the estimator error converges to zero, $\hat{s}_i(k) = 1 - \hat{x}_i(k) - \hat{r}_i(k)$ and

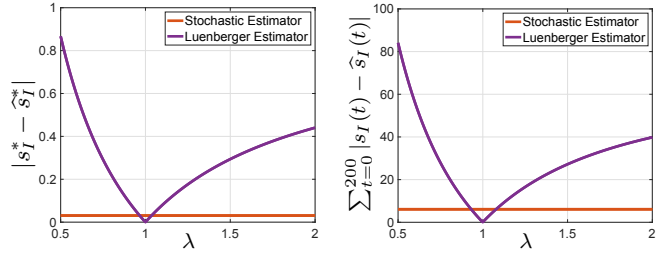


FIGURE 6: Each estimator's error at the healthy state (left) and total accumulated error (right) vs the scaling factor λ

$$\hat{r}_i(k) = h \sum_{q=0}^k \gamma'_i \hat{x}_i(q).$$

However, in our system model, we do not know any of the system parameters in the beginning of the outbreak. Hence, while performing the system states estimation through the Luenberger observer, we assume a set of the system parameters such that the estimated system states stay well-defined, namely $\hat{x}_i(k), \hat{s}_i(k), \hat{r}_i(k) \in [0, 1]$, for all $k \in \mathbb{Z}_{\geq 0}$. Furthermore, notice that the measurement $y_i(k)$ in Eq. (44) is the current value of the system state, assuming no delay in the measurement, while our estimation algorithm accounts for random delays caused by the incubation period of the virus or other factors.

We now investigate the robustness of the Luenberger observer with incorrect system parameters in comparison with our stochastic estimation framework. We introduce a scaling factor $\lambda > 0$ such that the assumed system parameters are:

$$\beta'_{ij} = \lambda \beta_{ij}, \gamma'_i = \frac{1}{\lambda} \gamma_i \quad (45)$$

for all $i, j \in [n]$. We choose to tune the original system parameters with the method in Eq. (45) so that the approximate reproduction number, namely $\frac{\beta'_{ij}}{\gamma'_i} = \lambda^2 \frac{\beta_{ij}}{\gamma_i}$, also changes accordingly with λ . We denote s_i^* as the susceptible level of the equilibrium at subpopulation i for the true system. We denote $\hat{s}_i^{*L}$ as the final estimated susceptible level via the Luenberger observer in (44) at subpopulation i . Finally, we denote $\hat{s}_i^{*P}$ as the final estimated susceptible level from our stochastic framework. We compare the offset of each estimator at the healthy state: $|s_i^* - \hat{s}_i^{*L}|, |s_i^* - \hat{s}_i^{*P}|$ against the scaling factor λ in Figure 6. We choose the range for the scaling factor to be $\lambda \in [0.5, 2]$ to represent the possible errors within 100% of the original system parameters' scales. The error of the Luenberger observer first decreases and then increases with the change of the scaling factor λ and reaches 0 when $\lambda = 1$, which means that the assumed parameters are correct and, as would be expected, there is no estimation error. Thus, unless you have accurate knowledge of the system parameters (in our example $\lambda \in [0.97, 1.03]$) and there is no delay between the change of infection level and its measurement, our stochastic estimation framework outperforms a standard observer framework.

B. Parameter Identification

By factoring the β_{ij} in Table II into $\beta_i a_{ij}$ and, for simplicity, letting $a_{ij} = \beta_{ij}$, assuming we acquire the network structure via, e.g., interstate traffic data, we can write $\beta_i = 1$ for all

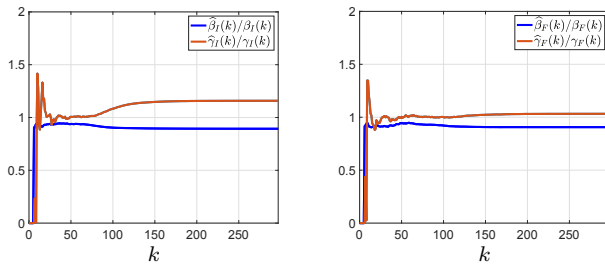


FIGURE 7: Identified parameters for nodes I (left) and F (right)

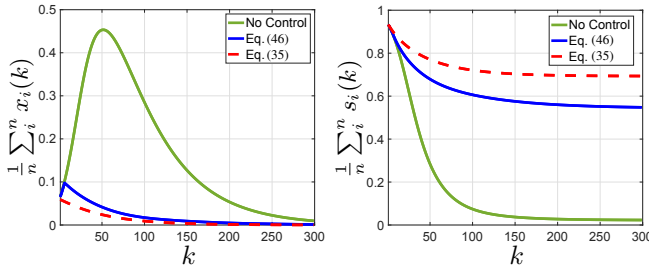


FIGURE 8: Average system states over time

$i \in [n]$. We then use the parameter identification method in Eq. (34) with the system states estimated from the data in Fig. 4 using the algorithm in (21) to learn $\hat{\beta}_i$ and $\hat{\gamma}_i$ with $T_1 = 6$ (recall, if $T_1 = 0$, the estimation error will be small by Prop 1) and $T_2 \in [7, 300]$. In Fig. 7 we plot the ratio between the estimated system parameters and the true parameters. We see the ratios at both nodes are close to 1 which demonstrates that our estimation algorithm can provide state inference that enables us to identify the system parameters. Notice that there is a slight decrease in $\frac{\hat{\beta}_i(k)}{\beta_i(k)}$ around $k = 100$, as the infection levels reach their peaks, and estimating the infection rates becomes difficult without becoming infected.

C. Control Implementation

We simulate three scenarios over the network in Fig. 3 with the parameters of Table. II: no control, the distributed eradication strategy utilizing estimated states in (35), and the distributed eradication strategy in (35) but with $\beta_{ij}(k)$ replaced by identified parameters $\hat{\beta}_{ij}(k)$ from Fig. 7:

$$\tilde{\gamma}_i(k) = \hat{s}_i(k) \sum_{j=1}^n \hat{\beta}_{ij}(k) + \epsilon_i, \quad i \in [n]. \quad (46)$$

The inferred states were produced by the algorithm in Section III with $p_i^x = 0.5 \forall i \in \{G, L, I, F, S\}$. The average states for each scenario are plotted in Fig. 8. Note that both eradication strategies are able to eliminate the virus at a much higher speed than the case where no control strategy applied. Furthermore, when $k \geq 200$, the healthy states with the eradication strategies applied achieve higher susceptible fraction than the healthy state without control. We can interpret the higher susceptible proportion as fewer individuals becoming sick during the entire outbreak and the eradication algorithms prevent resurgences of the virus over the network. Even though our system parameter estimation is not completely accurate due to the state estimation error, the control strategy utilizing the identified parameters is still able to eradicate the virus at a

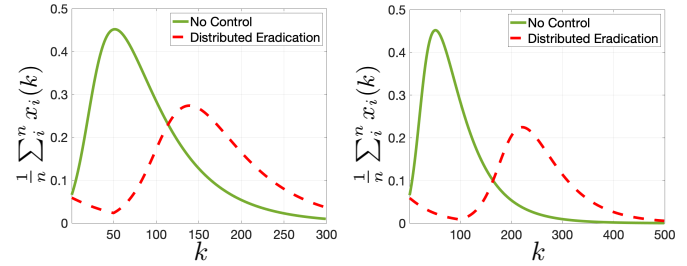


FIGURE 9: Average infection proportion of the virus over time with the eradication strategy enforced at $k \in [0, 50]$ (left) and at $k \in [0, 100]$ (right)

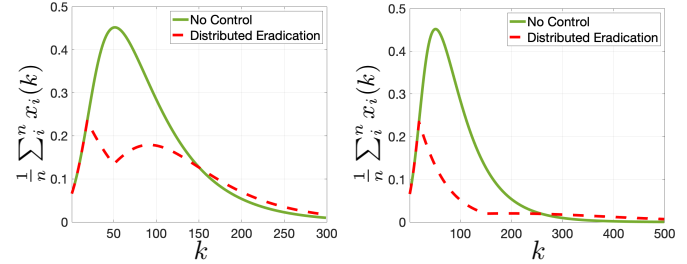


FIGURE 10: Average infection proportion of the virus over time with the eradication strategy imposed at $k \in [20, 50]$ (left) and at $k \in [20, 150]$ (right)

faster rate than no control, ensuring fewer individuals become infected over the course of the outbreak.

In Fig. 9, we remove the eradication strategy when $k = 50$ and $k = 100$ and do not reinstate it. It can be seen that the infection curve rises up immediately when $k \geq 50$ (resp. $k \geq 100$), and reach peaks before it slowly dies out. Fig. 9 can be interpreted as removing the allocation of resources and healthcare workers from a subpopulation too early during a pandemic, resulting in the increase in infection level and a potential outbreak. In Fig. 10, we only enforce our eradication strategy within time interval: $k \in [20, 50]$ and $k \in [20, 150]$, respectively. By implementing the control algorithm for more time, we are able to relatively flatten the infection curve. We can see that although the control strategy reduces the infection level, a resurgence of the outbreak occurs instantly upon the removal of the eradication strategy. Hence, policy makers are suggested to enforce the eradication strategy during the entire outbreak to avoid subsequent waves of infections.

In Fig. 11, we consider a time-varying system where the initial infection rates are set to the values from Table II and then increased linearly over time (by adding 0.0005 at each time step k) to capture the effect of a mutating virus. We utilize the eradication strategy proposed in (35) and the resulting average healing rates are shown in Figure 11 (right). We can see from Figure 11 (left) that our control strategy is able to flatten the curve and eradicate the epidemic.

We now explore how the algorithm performs if a resource constraint c is introduced:

$$\sum_{i=1}^n \tilde{\gamma}_i(k) \leq c.$$

To meet this constraint, we introduce a uniform limitation on

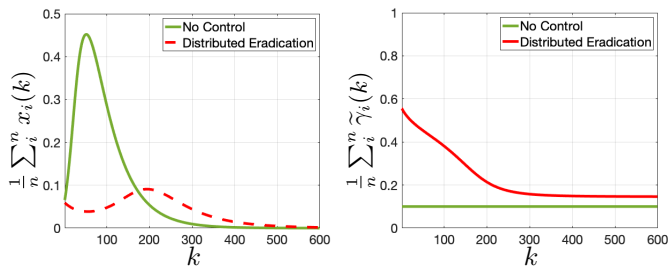


FIGURE 11: Average infection proportion (left) and average healing rate (right) of the virus over time with time-varying system parameters, with and without control from (35)

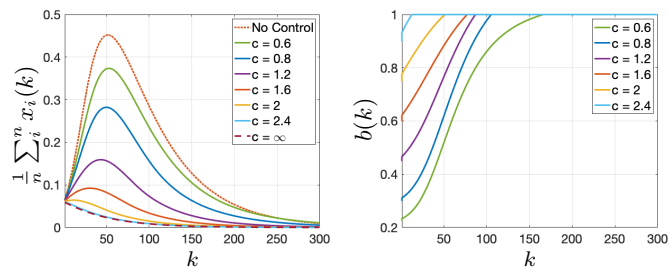


FIGURE 12: Average infection level when using (47) (left) and resource limitation coefficient $b(k)$ from (48) (right) over time with different resource limitations

the total amount of medical resources:

$$\tilde{\gamma}_i(k) = b(k)\tilde{\gamma}_i(k), \forall k \in \mathbb{Z}_{\geq 0}, \quad (47)$$

where $\tilde{\gamma}_i(k)$ is the eradication control strategy proposed in Eq. (35) and $b(k)$ is the resource limitation coefficient:

$$b(k) = \begin{cases} \frac{c}{\sum_{i=1}^n \tilde{\gamma}_i(k)} & \text{if } \sum_{i=1}^n \tilde{\gamma}_i(k) > c \\ 1 & \text{if } \sum_{i=1}^n \tilde{\gamma}_i(k) \leq c. \end{cases} \quad (48)$$

In Fig. 12, we choose $c = \{0.6, 0.8, 1.2, 1.6, 2, 2.4, \infty\}$ and plot the average infection level and $b(k)$ over time, notice that when $c = \infty$ the controller is identical to Eq. (35). We can see from Figure 12 that with more resources available, namely a larger value of c , the performance of the eradication strategy progresses as fewer individuals become infected over the outbreak. Therefore, when there are no constraints over medical resources, our proposed eradication strategy ensures speedy convergence to the set of healthy states.

VIII. CONCLUSION

This paper studied the inference and control of discrete time, time-varying SIR epidemics over networks. We proposed a stochastic framework for estimating the underlying epidemic states from collected testing data. We provided analytic expressions for the error of the estimation algorithm and validated some of our assumptions with real COVID-19 testing data. We identified the system parameters with the system states from the estimation algorithm proposed. We also proposed a distributed control strategy that is able to eradicate the virus exponentially fast. The control strategy provides insights for decision makers on how to eliminate an ongoing outbreak.

In future work, we plan to study the stability and control of models with more states than SIR such as SEIRS (susceptible-exposed-infected-recovered-susceptible) and SAIR (susceptible-asymptomatic-infected-recovered) as they can possibly capture the asymptomatic phase of COVID-19 better than the SIR model. In our stochastic testing framework, we did not consider the existence of inaccurate testing kits, which appear frequently and cause confusion for policy makers. Hence, we plan to include false positive/negative test results into our testing and estimation model and investigate the new model's estimation accuracy.

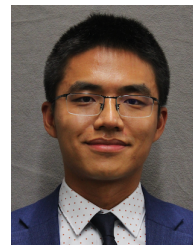
ACKNOWLEDGMENT

The authors would like to thank Ashish Hota, Paulina Babiak, and Baike She for useful conversations related to Section III.

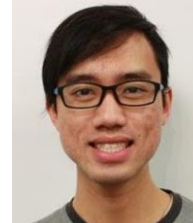
REFERENCES

- [1] K. Rock, S. Brand, J. Moir, and M. J. Keeling, "Dynamics of infectious diseases," *Reports on Progress in Physics*, vol. 77, no. 2, p. 026602, 2014.
- [2] W. Mei, S. Mohagheghi, S. Zampieri, and F. Bullo, "On the dynamics of deterministic epidemic propagation over networks," *Annual Reviews in Control*, vol. 44, pp. 116–128, 2017.
- [3] H. H. Weiss, "The SIR model and the foundations of public health," *Materials Mathematics*, pp. 0001–17, 2013.
- [4] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, "Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study," *BMC medicine*, vol. 5, no. 1, pp. 1–13, 2007.
- [5] H.-J. Chang, "Estimation of basic reproduction number of the Middle East respiratory syndrome coronavirus (MERS-CoV) during the outbreak in South Korea, 2015," *Biomedical Engineering Online*, vol. 16, no. 1, pp. 1–11, 2017.
- [6] D. Osthus, K. S. Hickmann, P. C. Caragea, D. Higdon, and S. Y. Del Valle, "Forecasting seasonal influenza with a state-space SIR model," *The Annals of Applied Statistics*, vol. 11, no. 1, p. 202, 2017.
- [7] G. C. Calafiore, C. Novara, and C. Possieri, "A modified SIR model for the COVID-19 contagion in Italy," in *Proc. 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 3889–3894.
- [8] "Estimated flu-related illnesses, medical visits, hospitalizations, and deaths in the United States-2018–2019 flu season," Sep 2021. [Online]. Available: <https://www.cdc.gov/flu/about/burden/2018-2019.html>
- [9] P. Spreuwenberg, M. Kroneman, and J. Paget, "Reassessing the global mortality burden of the 1918 influenza pandemic," *American J. of Epidemiology*, vol. 187, no. 12, pp. 2561–2567, 2018.
- [10] World Health Organization (WHO), "Global coronavirus (2019-nCoV)," <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, accessed: 2021-02-21.
- [11] C. Liu, X. Wu, R. Niu, X. Wu, and R. Fan, "A new SAIR model on complex networks for analysing the 2019 novel coronavirus (covid-19)," *Nonlinear Dynamics*, vol. 101, no. 3, pp. 1777–1787, 2020.
- [12] G. Giordano, M. Colaneri, A. Di Filippo, F. Blanchini, P. Bolzern, G. De Nicolao, P. Sacchi, P. Colaneri, and R. Bruno, "Modeling vaccination rollouts, SARS-CoV-2 variants and the requirement for non-pharmaceutical interventions in Italy," *Nature Medicine*, vol. 27, no. 6, pp. 993–998, 2021.
- [13] L. M. Kucirka, S. A. Lauer, O. Laeyendecker, D. Boon, and J. Lessler, "Variation in false-negative rate of reverse transcriptase polymerase chain reaction-based SARS-CoV-2 tests by time since exposure," *Annals of Internal Medicine*, vol. 173, no. 4, pp. 262–267, 2020.
- [14] V. Virlogeux, V. J. Fang, M. Park, J. T. Wu, and B. J. Cowling, "Comparison of incubation period distribution of human infections with MERS-CoV in South Korea and Saudi Arabia," *Scientific Reports*, vol. 6, no. 1, pp. 1–7, 2016.
- [15] R. L. Haffajee and M. M. Mello, "Thinking Globally, Acting Locally—The US Response to COVID-19," *New England J. of Med.*, vol. 382, no. 22, p. e75, 2020.
- [16] P. Van Mieghem, J. Omic, and R. Kooij, "Virus Spread in Networks," *IEEE/ACM Trans. on Net.*, vol. 17, no. 1, pp. 1–14, 2009.
- [17] H. J. Ahn and B. Hassibi, "Global dynamics of epidemic spread over complex networks," in *Proc. 52nd IEEE Conference on Decision and Control*, 2013, pp. 4579–4585.

- [18] N. T. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [19] M. Pascual and A. Dobson, "Seasonal patterns of infectious diseases," *PLoS Medicine*, vol. 2, no. 1, p. e5, 2005.
- [20] V. Bokharaie, O. Mason, and F. Wirth, "Spread of epidemics in time-dependent networks," in *Proc. International Symposium on Mathematical Theory of Networks and Systems (MTNS)*, vol. 5, no. 9, 2010.
- [21] S. Gracy, P. E. Paré, H. Sandberg, and K. H. Johansson, "Analysis and distributed control of periodic epidemic processes," *IEEE Trans. on Control of Network Systems*, vol. 8, no. 1, pp. 123–134, 2020.
- [22] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, "The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application," *Annals of Internal Medicine*, vol. 172, no. 9, pp. 577–582, 2020.
- [23] L. Dowson, C. Kober, N. Perry, M. Fisher, and D. Richardson, "Why some MSM present late for HIV testing: A qualitative analysis," *AIDS Care*, vol. 24, no. 2, pp. 204–209, 2012.
- [24] M. S. Islam, T. Sarkar, S. H. Khan, A.-H. M. Kamal, S. M. Hasan, A. Kabir, D. Yeasmin, M. A. Islam, K. I. A. Chowdhury, K. S. Anwar *et al.*, "COVID-19-related infodemic and its impact on public health: A global social media analysis," *The American J. of Tropical Medicine and Hygiene*, vol. 103, no. 4, p. 1621, 2020.
- [25] F. Spielberg, B. M. Branson, G. M. Goldbaum, D. Lockhart, A. Kurth, C. L. Celum, A. Rossini, C. W. Critchlow, and R. W. Wood, "Overcoming barriers to HIV testing: Preferences for new strategies among clients of a needle exchange, a sexually transmitted disease clinic, and sex venues for men who have sex with men," *J. of Acquired Immune Deficiency Syndromes*, vol. 32, no. 3, pp. 318–327, 2003.
- [26] O. Byambasuren, M. Cardona, K. Bell, J. Clark, M.-L. McLaws, and P. Glasziou, "Estimating the extent of true asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis," *J. of the Association of Medical Microbiology and Infectious Disease Canada*, vol. 5, no. 4, pp. 223–234, 2020.
- [27] D. Chang, G. Mo, X. Yuan, Y. Tao, X. Peng, F.-S. Wang, L. Xie, L. Sharma, C. S. Dela Cruz, and E. Qin, "Time kinetics of viral clearance and resolution of symptoms in novel coronavirus infection," *American J. of Respiratory and Critical Care Medicine*, vol. 201, no. 9, pp. 1150–1152, 2020.
- [28] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, "Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the diamond princess cruise ship, Yokohama, Japan, 2020," *Eurosurveillance*, vol. 25, no. 10, p. 2000180, 2020.
- [29] A. J. Ing, C. Cocks, and J. P. Green, "COVID-19: In the footsteps of Ernest Shackleton," *Thorax*, vol. 75, no. 8, pp. 693–694, 2020.
- [30] S. Bergquist, T. Otten, and N. Sarich, "COVID-19 pandemic in the United States," *Health Policy and Technology*, vol. 9, no. 4, pp. 623–638, 2020.
- [31] A. Hota, J. Godbole, and P. E. Paré, "A closed-loop framework for inference, prediction, and control of SIR epidemics on networks," *IEEE Trans. on Net. Sci. and Eng.*, vol. 8, no. 3, pp. 2262–2278, 2021.
- [32] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011.
- [33] E. M. Liotta, A. Batra, J. R. Clark, N. A. Shlobin, S. C. Hoffman, Z. S. Orban, and I. J. Koralnik, "Frequent neurologic manifestations and encephalopathy-associated morbidity in COVID-19 patients," *Annals of Clinical and Translational Neurology*, vol. 7, no. 11, pp. 2221–2230, 2020.
- [34] R. D. Mitrani, N. Dabas, and J. J. Goldberger, "COVID-19 cardiac injury: Implications for long-term surveillance and outcomes in survivors," *Heart Rhythm*, vol. 17, no. 11, pp. 1984–1990, 2020.
- [35] L. Lan, D. Xu, G. Ye, C. Xia, S. Wang, Y. Li, and H. Xu, "Positive rt-pcr test results in patients recovered from COVID-19," *JAMA*, vol. 323, no. 15, pp. 1502–1503, 2020.
- [36] M. Loeve, *Probability Theory*. Courier Dover Publications, 2017.
- [37] J. Zhao, Q. Yuan, H. Wang, W. Liu, X. Liao, Y. Su, X. Wang, J. Yuan, T. Li, J. Li *et al.*, "Antibody responses to SARS-CoV-2 in patients with novel coronavirus disease 2019," *Clinical Infectious Diseases*, vol. 71, no. 16, pp. 2027–2034, 2020.
- [38] R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe *et al.*, "Virological assessment of hospitalized patients with COVID-2019," *Nature*, vol. 581, no. 7809, pp. 465–469, 2020.
- [39] L. Guo, L. Ren, S. Yang, M. Xiao, D. Chang, F. Yang, C. S. Dela Cruz, Y. Wang, C. Wu, Y. Xiao *et al.*, "Profiling early humoral response to diagnose novel coronavirus disease (COVID-19)," *Clinical Infectious Diseases*, vol. 71, no. 15, pp. 778–785, 2020.
- [40] B. E. Young, S. W. X. Ong, S. Kalimuddin, J. G. Low, S. Y. Tan, J. Loh, O.-T. Ng, K. Marimuthu, L. W. Ang, T. M. Mak *et al.*, "Epidemiologic features and clinical course of patients infected with SARS-CoV-2 in Singapore," *JAMA*, vol. 323, no. 15, pp. 1488–1494, 2020.
- [41] P. E. Paré, C. L. Beck, and T. Başar, "Modeling, estimation, and analysis of epidemics over networks: An overview," *Annual Reviews in Control*, vol. 50, pp. 345–360, 2020.
- [42] D. Vrabac, P. E. Paré, H. Sandberg, and K. H. Johansson, "Overcoming challenges for estimating virus spread dynamics from data," in *Proc. 54th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2020, pp. 1–6.
- [43] A. C. Kalil, T. F. Patterson, A. K. Mehta, K. M. Tomashek, C. R. Wolfe, V. Ghazaryan, V. C. Marconi, G. M. Ruiz-Palacios, L. Hsieh, S. Kline *et al.*, "Baricitinib plus remdesivir for hospitalized adults with COVID-19," *New England J. of Med.*, vol. 384, no. 9, pp. 795–807, 2021.
- [44] P. C. Taylor, A. C. Adams, M. M. Hufford, I. De La Torre, K. Winthrop, and R. L. Gottlieb, "Neutralizing monoclonal antibodies for treatment of COVID-19," *Nature Reviews Immunology*, vol. 21, no. 6, pp. 382–393, 2021.
- [45] N. Muroga, Y. Hayama, T. Yamamoto, A. Kurogi, T. Tsuda, and T. Tsutsui, "The 2010 foot-and-mouth disease epidemic in Japan," *J. of Vet. Med. Science*, vol. 74, no. 4, pp. 399–404, 2012.
- [46] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas, "Optimal resource allocation for network protection against spreading processes," *IEEE Trans. on Control of Network Systems*, vol. 1, no. 1, pp. 99–108, 2014.
- [47] M. Ye, J. Liu, B. D. Anderson, and M. Cao, "Distributed feedback control on the SIS network model: An impossibility result," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 10955–10962, 2020.
- [48] N. L. Carothers, *Real Analysis*. Cambridge University Press, 2000.
- [49] Indiana state of USA. [Online]. Available: <https://www.vectorstock.com/royalty-free-vector/indiana-state-of-usa-solid-black-outline-map-of-vector-24335112>



Ciyuan Zhang is a Ph.D. student in the Elmore Family School of Electrical and Computer Engineering at Purdue University. He received his M.S. in Electrical Engineering from Columbia University in 2018 and his B.S. in Electrical Engineering from Xi'an Jiaotong University in 2016. He is a recipient of Purdue University's Bilsland Dissertation Fellowship. His research interests include estimating and controlling epidemic dynamics on networked systems.



Humphrey Leung is a Ph.D. student in the Elmore Family School of Electrical and Computer Engineering at Purdue University studying networked epidemic dynamics. He received his Bachelor of Science in 2017 and Master of Science in 2019 in Computer Science from Brigham Young University. His primary research interests are machine learning, deep learning, and network dynamics.



Brooks A. Butler is a Ph.D. student in the Elmore Family School of Electrical and Computer Engineering at Purdue University studying networked dynamic systems and safety-critical control. He received his M.S. in Computer Science and his B.S. in Applied Physics from Brigham Young University in 2020 and 2019, respectively. His primary research interests include automatic control of networks, safety guarantees for networked systems, and artificial intelligence.



Philip E. Paré is the Rita Lane and Norma Fries Assistant Professor in the Elmore Family School of Electrical and Computer Engineering at Purdue University. He received his Ph.D. in Electrical and Computer Engineering (ECE) from the University of Illinois at Urbana-Champaign (UIUC) in 2018, after which he went to KTH Royal Institute of Technology in Stockholm, Sweden to be a Post-Doctoral Scholar. He received his B.S. in Mathematics with University Honors and his M.S. in Computer Science from Brigham Young University in 2012 and 2014,

respectively.