A NEW INSIGHT INTO THE MODELING AND ESTIMATION OF EPIDEMIC SPREAD IN POPULATION NETWORKS USING COMPARTMENTAL ODES AND HIDDEN MARKOV MODELS*

SOOJEAN HAN , SOON-JO CHUNG , SHUYUE YU , AND JOHN C. DOYLE†

Abstract. Modeling of epidemic spread is a prevalent problem in the field of mathematical modeling and control and the recent global pandemic of the novel SARS-2 coronavirus (SARS-CoV-2) has sparked widespread renewed interest. This paper presents a novel two-part model: a large-scale multi-group compartmental model with ordinary differential equation (ODE) dynamics used to emulate disease spread across large-scale populations, and a small-scale compartmental Hidden Markov Model (HMM) used to estimate parameters at an individual level, which are used in the ODE dynamics. The novelty in our proposed technique is the combination and extensions of two models which are traditionally used separately in the study of epidemics. First, while the traditional ODE dynamics model the movement of the population as a whole, the multi-group compartmental model segregates it into groups based on individual features so that it can account for population heterogeneity. Second, extensions are made to the usual forward-backward algorithm for HMMs in that multiple simultaneous observation sequences can be incorporated, and time-varying parameters can be estimated. Additionally, virus super-spreaders are considered through a stochastic setting, which incorporate Gaussian white and Poisson jump noise in the dynamics. Various different experiments are performed to demonstrate our model on two datasets, constructed based on real contact-tracing COVID-19 data from USA, China, and South Korea.

Key words. Mathematical modeling, Population dynamics, Epidemiology, Parametric inference, Hidden Markov models, Application models in control theory, Stochastic systems

AMS subject classifications. 92-10, 93-10, 92D25, 92D30, 93C95, 62F15, 62M05, 93E03

1. Introduction.

Accurate modeling of the dynamics in open-loop plays an important role when it comes to the problem of epidemic tracking and control. For example, [17] designed a model to predict population mobility through census data in order to determine which targeted regions to vaccinate, thereby containing the spread of viruses. Other similar models have been shown to be quite effective in the regulation of various past outbreaks such as influenza pandemics [13, 14] and the SARS epidemic of 2003 [23]. The recent global outbreak of the novel SARS-2 coronavirus and the associated COVID-19 disease has caused a resurge of interest in this problem of epidemic control. An added challenge is that no vaccine is available yet, placing high emphasis on non-medical intervention strategies like social distancing and mask-wearing. Organized active efforts of mitigation such as frequent contact tracing and close monitoring of individuals have also been proven effective in mitigating the situation [18]. More draconian strategies like lockdowns, while perhaps the most effective in preventing spread, cannot be enforced long-term due to the devastating impact it has on the economy, leaving millions of people without work. This further emphasizes the need to develop models which can emulate the behavior of spread to a reasonable degree of accuracy.

There is a wide body of literature associated with the study of the mathematical model of epidemic dynamics. The seminal work [27] in 1927 introduced the original "compartmental model", which moves the individuals in a population through the different phases of illness undergone upon exposure to the virus. Mathematically, this movement is represented by a set of ordinary differential equations (ODEs). Since

*Submitted to the editors Nov 1st, 2020.

†Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125, USA. ({soojean, sjchung, syu5, doyle}@caltech.edu).

then, a diverse motley of extensions have been made to this traditional ODE-based approach: [31] provides a simple stochastic formulation for measles assuming recovered individuals gain permanent immunity, [11] includes more general nonlinearities in the equations to mimic more realistic interactions among infected and non-infected individuals, and [25] investigates periodic, sustained oscillations in the scenario where immunity is only temporary. More recently, for COVID-19, many similar compartmental models have been proposed and published (often as online preprints) in the wake of the pandemic, e.g., [2, 5, 30]. However, a couple notable limitations to this basic compartmental model may yield approximations which are low in resolution: 1) the population is assumed to be homogeneous, and 2) the entire population is collectively moved through a single compartmental model.

An alternative approach models the network at an individual level using hidden Markov models (HMMs), e.g., [16, 29, 32] Since each individual can be made to have different characteristics, heterogeneity can be incorporated and finer-grained approximations are obtained. An added benefit to the HMM module is that there are standard algorithms which enable the inference of unknown parameters; in the context of epidemics, they may be the death rate, recovery rate, etc. A majority of existing HMM-based models do not consider using multiple separate observation processes and time-varying parameters. One time-varying scheme for estimation in HMMs is provided by [7], which used Taylor approximation to transform the time-varying parameter function into a polynomial in time. Furthermore, while increasing computational power and more available data over the years have made these complex HMM-based models more approachable than ever, they are still difficult to implement for large-scale populations and it is also impractical to keep track of each individual's unknown parameters separately.

This paper proposes a method that combines the two model types: 1) a larger-scale nonlinear ODE module, constructed out of partitioning the large-scale network into disjoint groups then assigning each group to a compartmental model, is used to propagate the dynamics of the population, while 2) within each disjoint group, multiple smaller-scale HMM-based models are used to perform parameter estimation to be used in the ODE module. Machine learning and data-driven approaches have also been considered for COVID-19, e.g., [6, 35], but often suffer from well-known limitations such as massive amounts of offline training, which is not suitable for settings where we have insufficient data. Additionally, in models such as [22], grid search may be used to estimate the unknown parameters, but for an initial grid space which is too large, this takes up computation time. With our proposed model, we aim for the development of a more model-based analytical approach. We will henceforth refer to the large-scale propagation module of 1) simply as the *ODE module* and the small-scale parameter estimation module of 2) as the *HMM module*.

1.1. Paper Organization. In Section 2, we describe the graph notation we will be using throughout the paper, and in Section 2.1, we illustrate the underlying compartmental model which will serve as the foundation for both the ODE module and the HMM module. In Section 3, we outline the ODE module, part one of our two-part model, and mathematically define the parameters and the ODE dynamics in which it abides by in Section 3.1. An extension to the stochastic framework, which is used to account for the emergence of super-spreaders, is designed in Section 3.2. Next, in Section 4, we outline the HMM module, part two of our two-part model, and mathematically define the parameters and the HMM transition matrix per individual in Section 4.1. The Viterbi algorithm for estimating the true health state of an

individual will be used as a performance metric for our simulations, and is detailed in Section 4.2. Our algorithm for parameter estimation using multiple observation processes is detailed in Section 4.3, and the additional modification to time-varying parameters is described in Section 4.4. We then apply our overall two-part model to two datasets and present our simulation experiments in Section 5. Finally, the conclusion of the paper is made in Section 6.

2. Notation and Setup. We denote $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to be the undirected interaction network of individuals in the population; connections are based on the geographical location of each individual. The neighbor set $\mathcal{N}(n) := \{m \in \mathcal{V} | (n, m) \in \mathcal{E}\}$ denotes the set of all people who come into contact with individual $n \in \mathcal{V}$. For the ODE module, the nodes $\mathcal{V}$ are partitioned into K disjoint groups, and the collection of nodes which belong in Group k is denoted $\mathcal{V}_k \subset \mathcal{V}$. Furthermore, the corresponding edge subsets are denoted as follows: $\mathcal{E}_k := \{(n, m) \in \mathcal{E} | n, m \in \mathcal{V}_k\}$ is the set of edges which join members of the same Group k, while $\mathcal{E}_{jk} := \{(n, m) \in \mathcal{E} | (n \in \mathcal{V}_j \wedge m \in \mathcal{V}_k) \vee (n \in \mathcal{V}_k \wedge m \in \mathcal{V}_j)\}$ is the set of edges which join members of Group j and members of Group k, for $j < k$. The evolution of the disease spread within each group is modeled according to a *SEIRDS compartmental model* to be detailed in the following subsection.

Remark 2.1 (Group Partitioning Criteria). In this paper, we specifically focus on partitioning the population by geographical location. However, we remark that groupings can be created based on alternative sources of nonhomogeneity, e.g., by community membership, which varies the rate of contact, or by age, which varies the rate of death and recovery. The different metrics of partitioning are more relevant to the subject of network clustering where there is a wealth of literature on, e.g., [34]. We emphasize that the question of how the groups are created is outside the scope of our paper, but once they are created, the two-part model would work well for any suitable clustering, as long as the groups are disjoint.

2.1. The Compartmental Model. The underlying model of consideration throughout this work for both the HMM module and the ODE module, is the *compartmental model*. The two-part model can be used with any type of compartmental model, e.g. SIR or SIS, but we choose the five-compartment *SEIRDS model* because we believe they capture the most description about the COVID-19 disease to-date.

The five compartments are denoted $\{S, E, I, R, D\}$ and the dynamics between them are described as follows. A majority of the population begins as *susceptible* (S) individuals, meaning they do not test positive for traces of virus on their bodies. In the next phase, individuals are *exposed* (E) to the virus, meaning they have traces of the virus on their bodies but have not yet developed any noticeable symptoms of the disease. We argue that maintaining a separate compartment for (E) is helpful because many common viruses in the past have been greatly propagated by the asymptomatic exposed, and SARS-CoV-2 is no exception: COVID-19 symptoms typically present within 2-14 days upon exposure [4], and in the case of Singapore, as of March 17, 2020, it is reported that roughly 48% of infected individuals had been infected from exposed people alone [15]. After a certain period of time, individuals would either 1) return to becoming susceptible (S) again after all lingering traces of the virus disappear, or 2) begins to display symptoms, upon which they transition to the *ill* (I) phase. Finally, any individual in (I) would either 1) succumb to death by the illness, transition to the *death* (D) phase, and remain there for the rest of time, or 2) successfully fight the disease and transition to the *recovered (R) phase*, where they are invincible to the

virus. However, disease immunity may not be long-term in some cases; in the case of COVID-19, we have a growing amount of evidence of people becoming re-infected after recovery [20]. To account for this possibility, we allow for recovered individuals to transition back to the (S) phase after some stochastic period of time. The ODE version of the compartmental model, where the parameters correspond to the transition rates, is visualized in Figure 1 while the HMM version, where the parameters correspond to probabilities, is given in Figure 2. A more detailed description of each model and what the parameters will be are provided in Sections 3.1 and 4.1 for the ODE and the HMM modules, respectively.

DEFINITION 2.2 (Infectious). *An **infectious** individual is one in which any other individual in (S) is liable to becoming exposed via social contact with him/her. These individuals test positive for the presence of virus on their bodies, but may not always display symptoms of the disease. In this specific setting, both (E) and (I) individuals are considered infectious.*

Assumption 2.3. We do not take into account the natural births and natural deaths, as we are primarily focused on the effects of the disease itself on the existing population.

Assumption 2.4. For all $k \in \{1, \cdots, K\}$, the members of group k remain as members of group k. Thus, all susceptible individuals of k transition to becoming infectious members of k, from which they either die from the disease, or become a recovered member of group k. This assumption may be relaxed for time-varying networks, but for the main focus of this paper, we assume static topologies.

3. ODE Module: Large-Scale Disease Propagation. We begin by designing the ODE module, which will be used to emulate the virus spread throughout a given population network. We emphasize two key differences between traditional compartmental ODEs and the approaches we present here: in Section 3.1, we set up the multi-group compartmental model for each disjoint partition of the network to account for heterogeneous population characteristics, and in Section 3.2, we provide a stochastic formulation to account for the case of super-spreaders.

3.1. Basic Framework. For each of the five compartments we have, we denote $S(t)$ for the number of susceptible individuals in the population at time t, and likewise $E(t), I(t), R(t)$, and $D(t)$ for the number of ill, recovered, and deceased individuals at time t, respectively. With Assumption 2.3 in place, we have that for all time t, $S(t) + E(t) + I(t) + R(t) + D(t) = C$, where the constant C denotes the total population. As mentioned in Section 2, the population network is partitioned into K disjoint groups such that for each group $k \in [1, K]$, we have $S(t) = \sum_{k=1}^{K} S_k(t)$, and likewise for $E_k(t), I_k(t), R_k(t), D_k(t)$. By Assumption 2.4, the number of people per group remains the same for all time. Denote the size of subset k by the constant C_k, so that the total population is $C = \sum_{k=1}^{K} C_k$.

There are parameters associated with the transitions that occur between the five phases. For the ODE module, the full vector of parameters is given by $\theta_k := \begin{bmatrix} \{\beta_{kj}\} & \alpha_k^{(e)} & \alpha_k^{(s)} & \gamma_k^{(r)} & \gamma_k^{(d)} & \nu_k \end{bmatrix}^T$. We will denote $[\beta_{kj}] \in \mathbb{R}^{K \times K}$ (with $\{\beta_{kj}\}$ its vectorized form) to be the contact rate between each group of the network. Namely, β_{kk} denotes the rate of contact among susceptible and infectious members which belong in the same group k, while β_{kj}, denotes the average rate of contact among susceptible individuals in group k and infectious individuals in group j for all other groups $j \neq k$. The intuition for this definition is that there are K different groups,

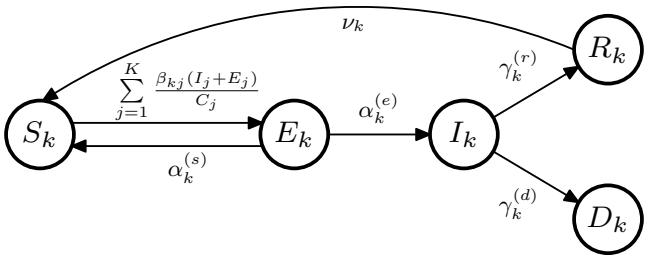

Fig. 1: SEIRDS for the ODE module for a single group of individuals, using transition rates as parameters.

and so a susceptible person in some group k has K different ways of becoming infected (i.e., by coming into contact with a person of any other group). $\alpha_k^{(z)}$ is the rate at which exposed individuals in group k leave their exposed status; they become ill if $z = e$ or susceptible if $z = s$. Similarly, $\gamma_k^{(z)}$ is the rate that ill individuals of group k leave their ill status, becoming recovered if $z = r$ or dead if $z = d$. Lastly, ν_k is the rate at which recovered people of group k become susceptible again.

The diagram of a single group of the ODE module is shown in Figure 1. The overall model can be visualized as K identical copies of this diagram, connected to each other according to the network topology. The sole exception is that all deaths are placed into a single (D) compartment because we are only interested in the migration patterns of individuals who can be impacted by the disease.

Mathematically, for each k, the dynamics are governed by the following set of differential equations:

$$\frac{dS_k(t)}{dt} = -\sum_{j=1}^{K} \frac{\beta_{kj}(I_j(t) + E_j(t))}{C_j(t)} S_k(t) + \nu_k R_k(t) + \alpha_k^{(s)} E_k(t) \tag{3.1a}$$

$$\frac{dE_k(t)}{dt} = \sum_{j=1}^{K} \frac{\beta_{kj}(I_j(t) + E_j(t))}{C_j(t)} S_k(t) - (\alpha_k^{(s)} + \alpha_k^{(e)}) E_k(t) \tag{3.1b}$$

$$\frac{dI_k(t)}{dt} = \alpha_k^{(e)} E_k(t) - (\gamma_k^{(r)} + \gamma_k^{(d)}) I_k(t) \tag{3.1c}$$

$$\frac{dR_k(t)}{dt} = \gamma_k^{(r)} I_k(t) - \nu_k R_k(t) \tag{3.1d}$$

where the parameters are defined previously, and the single death compartment has its own collective equation over all groups.

$$\frac{dD(t)}{dt} = \sum_{k=1}^{K} \gamma_k^{(d)} I_k(t)$$

3.2. A Stochastic Extension. Oftentimes, there are effects that the deterministic ODE module of the previous subsection cannot capture. Most notable is the effect of super-spreaders, which typically arises when a small handful of infectious people come in contact with an abnormally large number of susceptible individuals. A notable example of this effect is the individual known as Patient 31, who is believed to have been a primary cause for the sudden spike in cases of COVID-19 in

South Korea, which eventually lead to hundreds of deaths between late February and early March of 2020 [26]. Modeling phenomenons like these requires some modification to the entries of the $[\beta_{kj}]$ matrix. While white-noise perturbations have been accounted for in prior work [9, 10, 21] to mimic slight variations in the daily number of contacts, its influence is small and occurs steadily over some measurable duration of time. White noise is insufficient to represent a broad class of sudden impulsive perturbations like the phenomenon of super-spreaders.

Our work in [24] discusses the shortcomings of a pure Gaussian white noise model in the context of robotic control, and models these impulsive disturbances as *Poisson-distributed shot noise*. Using shot noise, the evolution of $\beta_{kj}(t)$ over time for each $k, j \in [1, K]$ can be characterized by the stochastic differential equation (SDE):

$$\beta_{kj}dt = \beta_{kj,0}dt + \sigma_{kj}dW(t) + \xi_{kj}dN_{kj}(t) \tag{3.2}$$

where $\beta_{kj,0}$ is the baseline constant value of the parameter, W is standard white noise, N_{kj} is standard Poisson noise with intensity λ_{kj}, and σ_{kj} and ξ_{kj} are their corresponding variances. This generic equation (3.2) replaces the β_{kj} or β_{kk} of every interaction between the groups of equation (3.1) where it is suspected that a higher-than-average frequency of contact might occur (e.g., community meetings, clubs). For example, when super-spreaders emerge from interactions within group k only, β_{kk} abides by equation (3.2) and the deterministic dynamics equation (3.1) become modified as follows.

$$\begin{aligned} dS_k(t) = &\left(-Z(t) - \frac{\beta_{kk,0}(I_k(t)+E_k(t))}{C_k(t)}S_k(t) + \nu_k R_k(t) + \alpha_k^{(s)}E_k(t)\right)dt \\ &- \frac{(I_k(t)+E_k(t))S_k(t)}{C_k}(\sigma dW(t) + \xi dN(t)) \end{aligned} \tag{3.3a}$$

$$\begin{aligned} dE_k(t) = &\left(Z(t) + \frac{\beta_{kk,0}(I_k(t)+E_k(t))}{C_k}S_k(t) - (\alpha_k^{(s)} + \alpha_k^{(e)})E_k(t)\right)dt \\ &+ \frac{(I_k(t)+E_k(t))S_k(t)}{C_k}(\sigma dW(t) + \xi dN(t)) \end{aligned} \tag{3.3b}$$

where the variable Z refers to the sum:

$$Z(t) := \sum_{j \neq k} \frac{\beta_{kj}(I_j(t)+E_j(t))}{C_j(t)}S_k(t)$$

and the rest of the equations are kept the same.

4. HMM Module: Individual Parameter Estimation. Now that we have a method to depict disease propagation across the entire network, we design the HMM module. We assign a SEIRDS compartmental model to each single individual in each group $k \in [1, K]$ of the population, and perform parameter estimation at the finest, individual-level resolution before aggregating all the parameters in the group and substituting them into the rates in equation (3.1).

In Section 4.1, we outline the basic framework and relevant notations. While the forward-backward and expectation-maximization procedures are standard to the HMM literature, we emphasize two simultaneous extensions that are made for the purposes of this paper: 1) multiple different time series of observations can be incorporated at once, and 2) time-varying parameters are handled. In doing so, we address two specific questions. Given a complete sequence of observed symptoms over time:

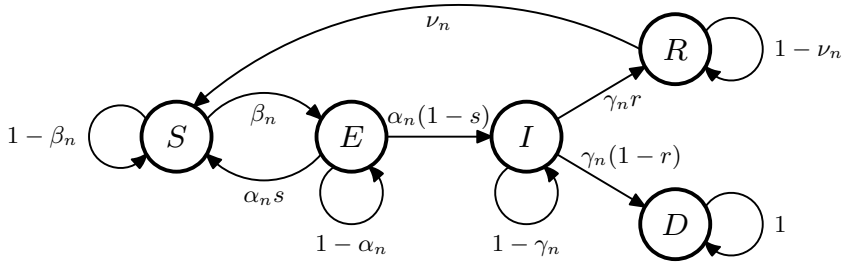

Fig. 2: SEIRDS for the HMM module for a single individual, using transition probabilities as parameters.

1. how can we infer the state of the HMM, i.e., the health status?
2. how can we estimate the unknown parameters of the HMM, such as the probability of death or recovery?

We address Question 1) in Section 4.2, then address Question 2) in Section 4.3 which discusses the algorithm for multiple observation sequences, and Section 4.4 which extends it to time-varying parameters.

4.1. Basic Framework. Within each node n of the network lies a hidden Markov Model (HMM) with hidden state $X_n \in \mathcal{X}$ representing the individual's health status. The state space for each n consists exactly of the five compartments $\mathcal{X} := \{S, E, I, R, D\}$.

The full vector of parameters is given by $\theta_n(t) := \begin{bmatrix}\beta_n(t) & \alpha_n & \gamma_n & \nu_n & s & r\end{bmatrix}^T$, where the parameters are described analogously to the ODE version in Section 3, with probabilities instead of rates. Namely, $\beta_n(t)$ is the probability that a susceptible n contracts the virus via exposure to his/her infectious neighbors. $\beta_n(t)$ varies with time because it is dependent on the number of infectious neighbors of n $\kappa_n(t) < |\mathcal{N}_n|$, which varies with time. α_n is the probability that exposed individuals leave their exposed status, and s is the proportion of those who return to being susceptible, while the remaining $1-s$ become ill. γ_n is the probability that ill individuals leave their ill status; r is the proportion who make a recovery, while the remaining $1-r$ die. ν_n is the probability that a recovered person becomes susceptible again. For simplicity, we assume s and r are universally constant; the analysis would not change significantly if they varied with n as well. The visualization of the HMM model for a single individual n is shown in Figure 2.

The transition probability matrix is then given as follows, for individual n with the state ordering as $\{S, E, I, R, D\}$.

$$P_n^{(t)} = \begin{bmatrix} 1-\beta_n(t) & \beta_n(t) & 0 & 0 & 0 \\ \alpha_n s & 1-\alpha_n & \alpha_n(1-s) & 0 & 0 \\ 0 & 0 & 1-\gamma_n & \gamma_n r & \gamma_n(1-r) \\ \nu_n & 0 & 0 & 1-\nu_n & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where the parameters are described previously.

Remark 4.1 (Either HMM or ODE as the Emulation Module). A relationship between the HMM module representation and the ODE module representation can be illustrated as follows: an individual in group k who has health state S at time

t is counted in the number $S_k(t)$. When (s)he transitions to E at time $t+1$, we decrement $S_k(t+1) = S_k(t) - 1$ and increment $E_k(t+1) = E_k(t) + 1$. Thus, as alluded to in Section 1, it is possible for the HMM module to be used as the emulation module instead of the ODE module by using the Markov chain dynamics $X_n(t+1) = P_n^{(t)} * X_n(t)$, and counting the number of individuals which belong in each state, i.e., $S_k(t) = \sum_{n \in \mathcal{V}_k} \mathbb{1}\{X_n(t) = S\}$ and likewise for the other four compartments. However, the key point is that when handling large-scale populations, individual-level modeling is not ideal due to computation time incurred from the large number of parameters that need to be estimated; hence, we use the coarser group partitions of the ODE module to propagate the emulation instead. In Section 5.3, we illustrate how the HMM module can be used as the emulation module on our smaller dataset of ~ 200 nodes.

Remark 4.2 (Multiple Groups for HMMs). While, like the ODE model, it is possible to define subscript n to refer to a group rather than a single individual, that requires the state space $\mathcal{X}$ to capture more than 5 states. In fact, if the node was a group of 2 people, we would need 25 different states $((S,S),(S,E),(S,I),\cdots,(E,S),(E,E),...)$, and in general, a group of m people would take 5^m states to represent. On the other hand, using the regular proposed HMM model for a network of c individuals, it is impractical to keep track of all 6^c parameters separately. If we formed groups for the HMM model by age or prior health condition, the same set of parameters could be used for all individuals within a single group. However, this also requires the ODE model to be grouped by the same metric, rather than geographical location.

In practice, the true health status of an individual can be difficult to discern, especially when (s)he is in the (E) state, and we may only infer his/her health via observation of the symptoms (s)he exhibits. Suppose that there are a total of M notable disease symptoms that an individual can present. For simplicity, we assume that they occur independently of each other. We define binary random variables $Y_{n,j}, j = 1,\cdots,M$, which take value 0 if symptom i is not present in individual n, and 1 otherwise; the scalar realizations of these random variables are denoted $y_{n,j} \in \{0,1\}$. Further denote the matrix $O_n(y_{n,\cdot}|\cdot) \in \mathbb{R}^{M \times |\mathcal{X}|}$ for individual n, with entries $O_n(y_{n,j}|x_n)$ denoting the probability of observing the positive presence symptom j, meaning $y_{n,j} = 1$, in individual n. Note that each row of this matrix does not necessarily need to sum to 1, unlike the standard notation, since the entries correspond only to the case $y_{n,j} = 1$.

We will denote random vector $\mathbf{X}_n^{(t_1:t_2)} := (X_n(t_1), X_n(t_1+1), \cdots, X_n(t_2))$ for $0 < t_1 < t_2$ and the vector of deterministic values as $\mathbf{x}_n^{(t_1:t_2)} := (x_n(t_1), x_n(t_1+1), \cdots, x_n(t_2))$. When $t_1 = 0$, we instead use the shorthand notation $\mathbf{X}_n^{(t_2)}, \mathbf{x}_n^{(t_2)}$. All mentioned notation also extends to $\mathbf{Y}_n, \mathbf{y}_n$ in the following way:

$$\{\mathbf{Y}_n^{(t)} = \mathbf{y}_n^{(t)}\} \iff \left\{ \begin{bmatrix} \mathbf{Y}_{n,1}^{(t)} \\ \vdots \\ \mathbf{Y}_{n,M}^{(t)} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{n,1}^{(t)} \\ \vdots \\ \mathbf{y}_{n,M}^{(t)} \end{bmatrix} \right\}$$

4.2. Estimating the True State. In this subsection, we will answer the first question posed in the beginning of this section by applying the standard Viterbi algorithm (e.g., [19]) to each separate observation sequence, then aggregating them using the method described below.

The probability of observing a specific sequence of states is given by:

$$\begin{aligned}&\mathbb{P}(\{X_n(t) = x_n(t), n \in \mathcal{V}, t \in [0,T]\}) \\ &\quad = \prod_{n\in\mathcal{V}} q_n(x_n(0)) \prod_{\substack{t\in[0,T-1]\\ n\in\mathcal{V}}} \mathbb{P}(X_n(t+1)|X_n(t), \{X_m(t), m \in \mathcal{N}(n)\})\end{aligned}$$

where, $q_n(x)$ denotes the initial probability that individual n starts off at state x.

Based on the observations of an individual's symptoms, we can determine the best-fitting sequence of states by invoking the *Viterbi algorithm*. For a single observation sequence $j \in \{1, \cdots, M\}$:

$$\begin{aligned}\delta_{n,j}(0,x) &= q_n(x)O_n(y_{n,j}(0)|x) \\ \delta_{n,j}(t,x) &= \max_{z\in\mathcal{X}} \delta_{n,j}(t-1,z)P_n^{(t-1)}(z,x)O_n(y_{n,j}(t)|x),\ t \geq 1\end{aligned}$$

Then for the specific observation sequence i, the optimal sequence of states is given by $x^*_{n,j}(t) := \text{argmax}_{z\in\mathcal{X}}\delta_{n,j}(t,z)$. The overall optimum sequence is determined according to whichever state value occurs as the majority. Ties are broken according to the state which is more "harmful" to the network, e.g. if the most likely state is tied between susceptible (S) or exposed (E), then we take the individual to be exposed because (s)he is liable to infecting more people in the network.

4.3. Multiple Simultaneous Observation Processes. Now we will address the second question: given $\{\mathbf{Y}_n^{(T)} = \mathbf{y}_n^{(T)}\}$, how can we estimate $\{q_n, P_n, O_n\}$? We will do this by modifying standard HMM procedures, namely the *forward-backward algorithm* (e.g., [28]) and the *Baum-Welch algorithm*, alternatively known as the procedure of *Expectation-Maximization* (EM) (e.g., [8]), in two simultaneous ways: 1) by incorporating multiple observation processes, and 2) by considering parameters which are time-varying.

We will begin by presenting the forward-backward algorithm which incorporates multiple observation processes. We derive the posterior distribution of the current state $X_n(t)$ given past observations $\mathbf{Y}_n^{(t)}$ and future observations $\mathbf{Y}_n^{(t+1):T}$ by applying the forward-backward algorithm for a single observation process on each $j \in \{1, \cdots, M\}$. Define $f_{n,j}(t,x) := \mathbb{P}(X_n(t) = x, \mathbf{Y}_{n,j}^{(t)} = \mathbf{y}_{n,j}^{(t)})$ to be the probability that the individual is in state $x \in \mathcal{X}$ at time t and the past observed symptom sequence is given by $\mathbf{Y}_{n,j}^{(t)} = \mathbf{y}_{n,j}^{(t)}$, and define $b_{n,j}(t,x) := \mathbb{P}(\mathbf{Y}_{n,j}^{(t+1:T)} = \mathbf{y}_{n,j}^{(t+1:T)}|X_n(t) = x)$ to be the probability of observing a future sequence of symptoms $\mathbf{Y}_{n,j}^{(t+1:T)} = \mathbf{y}_{n,j}^{(t+1:T)}$ given we know the individual is in state x. The recursive equations for $f_{n,j}$ and $b_{n,j}$ are then given by:

(4.1a)
$$f_{n,j}(t,x) = \sum_{z\in\mathcal{X}} f_{n,j}(t-1,z)O_n(y_{n,j}(t)|x)P_n^{(t-1)}(z,x), \quad f_{n,j}(0,x) := q_n(x)O_n(y_{n,j}(0)|x)$$

(4.1b)
$$b_{n,j}(t,x) = \sum_{z\in\mathcal{X}} b_{n,j}(t+1,z)P_n^{(t)}(x,z)O_n(y_{n,j}(t+1)|z), \quad b_{n,j}(T,x) = 1\ \forall x \in \mathcal{X}$$

Now define $g_{n,j}(t,x)$ to be the probability that the state of individual n at time t is x given observation sequence j, and $h_{n,j}(t,x,z)$ to be the probability that the state

of individual n makes a transition from x to z at time t. That is:

(4.2a)
$$g_{n,j}(t,x) = \mathbb{P}(X_n(t) = x | \mathbf{Y}_{n,j}^{(T)} = \mathbf{y}_{n,j}^{(T)}) = \frac{f_{n,j}(t,x) b_{n,j}(t,x)}{\sum_{z \in \mathcal{X}} f_{n,j}(t,z) b_{n,j}(t,z)}$$

(4.2b)
$$\begin{aligned} h_{n,j}(t,x,z) &:= \mathbb{P}(X_n(t) = x, X_n(t+1) = z | \mathbf{Y}_{n,j}^{(T)} = \mathbf{y}_{n,j}^{(T)}) \\ &= \frac{f_{n,j}(t,x) P_n^{(t)}(x,z) O_n(\mathbf{y}_{n,j}(t+1)|z) b_{n,j}(t+1,z)}{\sum_{u,w \in \mathcal{X}} f_{n,j}(t,u) P_n^{(t)}(u,w) O_n(\mathbf{y}_{n,j}(t+1)|w) b_{n,j}(t+1,w)} \end{aligned}$$

For each variable $f_{n,j}, b_{n,j}, g_{n,j}, h_{n,j}$, we will apply Baum-Welch Algorithm to determine the estimate $\hat{\theta}_{n,j}(T)$ for θ_n based on the single observation sequence j. We refer to $\hat{\theta}_{n,j}(T)$ as the *auxiliary variable* and define a corresponding auxiliary function as:

$$Q_{n,j}(T) := \mathbb{E}\left[\log\left(p_{n,j}^{(c)}(\mathbf{X}_n^{(T)}, \mathbf{Y}_{n,j}^{(T)} | \theta_n) \right) \middle| \mathbf{y}_{n,j}^{(T)}, \hat{\theta}_{n,j}(T) \right] \tag{4.3}$$

where $p_{n,j}^{(c)}$ denotes the joint probability distribution of observing a complete set of data $\{\mathbf{x}_n^{(T)}, \mathbf{y}_{n,j}^{(T)}\}$ for individual n:

$$p_{n,j}^{(c)}(\mathbf{x}_n^{(T)}, \mathbf{y}_{n,j}^{(T)} | \theta_n) = q_n(x_n(0)) \prod_{t=0}^{T-1} P_n^{(t)}(x_n(t), x_n(t+1)) \prod_{t=0}^{T} O_n(y_{n,j}(t) | x_n(t))$$

Maximizing the equation (4.3) subject to three regularity conditions $\sum_{u \in \mathcal{X}} \hat{P}_{n,j}^{(T)}(x,u) = 1$, $\sum_{y_j \in \{0,1\}} \hat{O}_{n,j}^{(T)}(y_j|x) = 1$, and $\sum_{x \in \mathcal{X}} \hat{q}_{n,j}^{(T)}(x) = 1$ for all $x \in \mathcal{X}, j \in \{1, \cdots, M\}$, yields the optimal point:

$$\hat{q}_{n,j}^{(T)}(x) = g_{n,j}(0,x) \tag{4.4a}$$

$$\hat{P}_{n,j}^{(T)}(x,z) = \left(\sum_{t=0}^{T-1} h_{n,j}(t,x,z) \right) \left(\sum_{t=0}^{T} g_{n,j}(t,x) \right)^{-1} \tag{4.4b}$$

$$\hat{O}_{n,j}^{(T)}(y|x) = \left(\sum_{t=0}^{T} g_{n,j}(t,x) \mathbb{1}\{y_{n,j}(t) = y\} \right) \left(\sum_{t=0}^{T} g_{n,j}(t,x) \right)^{-1} \tag{4.4c}$$

After a short sequence of new observations are accumulated for a time duration of ΔT, the entire procedure can be repeated to recompute equation (4.4) at time $T + \Delta T$ We will denote the overall simulation horizon as T_{sim}. For a single individual $n \in \mathcal{V}$, the inference problem for observation sequence j can be solved according to the standard single-observation Baum-Welch algorithm.

Once we have variable estimates $\hat{\theta}_{n,j}$ for each $j \in \{1, \cdots, M\}$, we need to aggregate them to obtain a single definitive estimate $\hat{\theta}_n$ which uses information from all M observations. Define a new weighted-average auxiliary function below.

$$Q_n(T) = \sum_{j=1}^{M} w_j Q_{n,j}(T) \tag{4.5}$$

where $\mathbf{w}$ is a vector of weights such that the components satisfy $\sum w_j = 1$ and $Q_{n,j}$ is defined previously in equation (4.3). Intuitively, the assignment of weights is chosen via two metrics: 1) the observation sequences are statistically correlated with each other, or 2) one observation sequence yields more information about a state than another, e.g., observing a fever on an individual may be more reflective of his/her ill state than a runny nose. For simplicity, we assume that these weights are known beforehand and that our observation processes are independent of each other, meaning that the weights are only chosen according to how well they represent the true state.

The previous work of [33] estimates the full observation matrix $\hat{O}_n^{(T)}(y|x)$ by averaging the $\hat{O}_{n,j}^{(T)}(y|x)$, like it does for the two estimates $\hat{q}_n^{(T)}$ and $\hat{P}_n^{(T)}$. This implies that an estimate based off a single observation process j estimates rows of the full matrix beyond just row j. However, for our assumption of independence among observation processes, this is not a reasonable thing to do; estimates based off a runny nose cannot be used to deduce estimates based off a fever. Instead, our approach contrasts this by using the jth observation process to obtain only an estimate for values in the jth row of the true O_n matrix. Then rows $j = 1$ to M are stacked up in order to obtain the full estimate $\hat{O}_n^{(T)}(y|x)$ of the original $O_n(y|x)$.

With this in mind, maximizing the auxiliary function equation (4.5) subject to the same three regularity conditions gives us the modified expressions for $\{\hat{q}_n^{(T)}, \hat{P)}_n^{(T)}\}$:

$$\hat{q}_n^{(T)}(x) = \sum_{j=1}^{M} w_j q_{n,j}^{(T)}(x) \tag{4.6a}$$

$$\hat{P}_n^{(T)}(x,z) = \left(\sum_{j=1}^{M}\sum_{t=0}^{T} w_j h_{n,j}(t,x,z)\right)\left(\sum_{j=1}^{M}\sum_{t=0}^{T} w_j g_{n,j}(t,x)\right)^{-1} \tag{4.6b}$$

4.4. Time-Varying Baum-Welch. In this section, we extend equation (4.6) to account for time-varying parameters. Recall that the rate $\beta_n(t)$ is dependent on the number of infectious neighbors of n, $\kappa_n(t)$, which varies with time.

First, in the forward-backward recursion and the Baum-Welch procedure for each observation process j, we apply a discounting factor $a \in (0, 1]$ which weights the values of past estimates less the further back in the past they were observed. This modification is due to the fact that it no longer makes sense for estimates over the full time horizon $[0, T]$ to be used to get $\{\hat{P}_{n,j}^{(T)}, \hat{O}_{n,j}^{(T)}\}$. While the expression for $\hat{q}_{n,j}^{(T)}$ is the same as in equation (4.4), the equation for the estimates of the transition probabilities and each row of the observation matrix change as follows:

$$\hat{P}_{n,j}^{(T)}(x,z) = \left(\sum_{t=0}^{T-1} a^{T-t} h_{n,j}(t,x,z)\right)\left(\sum_{t=0}^{T} a^{T-t} g_{n,j}(t,x)\right)^{-1} \tag{4.7a}$$

$$\hat{O}_{n,j}^{(T)}(y|x) = \left(\sum_{t=0}^{T} a^{T-t} g_{n,j}(t,x)\mathbb{1}\{y_{n,j}(t) = y\}\right)\left(\sum_{t=0}^{T} a^{T-t} g_{n,j}(t,x)\right)^{-1} \tag{4.7b}$$

The method of aggregation is the same as in equation (4.6).

Second, there is a specific expression for $\beta_n(t)$ in terms of $\kappa_n(t)$, which we can use once we have an estimate of $\beta_n(t)$ in the corresponding matrix entry from the forward-backward recursion. For now, we will consider the case where the network is

closed, in that no interaction with any individual outside of the network is allowed. Let $\beta_n^{(1)} \in [0,1]$ denote the probability of a susceptible person becoming exposed upon contact with an infectious individual, with superscript 1 to emphasize that the parameter is defined for a *single* susceptible individual and a *single* infectious individual. Then the probability of individual n transitioning from susceptible at time t to exposed at time $t+1$, given we know the states of all his current neighbors $\{X_m, m \in \mathcal{N}(n)\}$ is given by:

$$\beta_n(t) := \mathbb{P}(X_n(t+1) = E | X_n(t) = S) = 1 - \left(1 - \beta_n^{(1)}\right)^{\kappa_n(t)} \tag{4.8}$$

In order to estimate $\kappa_n(t)$, individual n runs the Viterbi algorithm on each of his/her neighbors according to the symptoms (s)he observes from them, then (s)he simply counts the number of people (s)he infers to be exposed or ill. One caveat is that since the individual has no past history of his/her neighbors (i.e., (s)he only collects observations about his/her neighbors at the current timestep), it simply reduces to a maximum likelihood problem. Then we estimate $\beta_n^{(1)}$ using our estimate of $\kappa_n(t)$.

There are two cases we need to consider: $\kappa_n(t) > 0$ and $\kappa_n(t) = 0$. The simpler case is when $\kappa_n(t) = 0$. If individual n does not have any infectious neighbors, there is no way for him/her to transition to the exposed state. Thus, in this current transition probability matrix, $\beta_n(t) = 0$. For the case where $\kappa_n(t) > 0$, $\beta_n(t)$ is recorded by algebraic rearrangement: $\hat{P}_n^{(t)}(1,2) = 1 - \left(1 - \beta_n^{(1)}\right)^{\kappa_n(t)} \implies \beta_n^{(1)} = 1 - \left(1 - \hat{P}_n^{(t)}(1,2)\right)^{\frac{1}{\kappa_n(t)}}$, where $\hat{P}_n^{(t)}(1,2) > 0$, the $(1,2)$th entry of $\hat{P}_n^{(t)}$, is where $\beta_n(t)$ is recorded after estimation via Baum-Welch.

The combined estimation procedure developed throughout this section is presented below in Algorithm 4.1.

Algorithm 4.1 Expectation Maximization with Observation Sequence j

Input: $\theta_{n,j}^{(0)}, T_{\text{sim}}, a, \mathbf{w}$

Output: $\hat{\theta}_n^{(1:T_{\text{sim}})}$

1: Initialize parameter estimate with $\theta_{n,j}^{(0)}$.
2: **for** $T = 1 : T_{\text{sim}}$ **do**
3: Perform forward-backward equation (4.1) to obtain $f_{n,j}(t,x)$ and $b_{n,j}(t,x)$
4: Obtain $g_{n,j}(t,x)$ and $h_{n,j}(t,x,z)$ from equation (4.2).
5: Compute $\hat{q}_{n,j}^{(T)}, \hat{P}_{n,j}^{(T)}, \hat{O}_{n,j}^{(T)}(y|x)$ from equation (4.7).
6: Aggregate parameter estimates using equation (4.6).
7: **end for**

Finally, once the entries of the transition probability matrix are determined via Algorithm 4.1, the conversion from probability p to rate r for the ODE model is done through the standard formula $p := 1 - e^{-r}$, where the time cycle period is one day. The sole exception is the $\beta(t)$ parameter. Instead, the time-varying rate of susceptible individuals becoming ill in the ODE case is just a first-order Taylor approximation of the HMM case when $\beta_{jk} := \beta_j^{(1)}$ uniformly for all k. Using the representation outlined in Remark 4.1, we can equate $\kappa_n(t)$ from the HMM model with $I_n(t) + E_n(t)$ to the ODE model in the finest-grain case when each group represents a single individual, because both notations represent the number of infectious neighbors.

5. Simulation. We apply our two-part model to specific case studies pertaining to the ongoing pandemic of COVID-19 and demonstrate their performances via numerical simulation. Although the datasets we use for our experiments were based on real statistics provided online, we remark that other inaccuracies such as geographical network topology, population count, and missing true parameter entries could potentially contribute as sources of error.

First, in Section 5.1, we describe the construction of the two datasets, describe how we extrapolated the nominal parameter values for each individual node, and explain how we constructed the rows of the observation matrix of equation (4.7b). Next, the performance of Algorithm 4.1 is evaluated according to two performance metrics defined in Section 5.2, and the influence of the interaction topology among a small group of individuals is discussed in Section 5.3. The trajectories showing the evolution of the number of people belonging to each of the five compartments over time is determined by the ODE module and the estimated transition probabilities from the HMM module; the results are discussed in Section 5.4. Based on this deterministic result, the stochastic model is applied and simulated in Section 5.5 in the event where one of the cliques in the network experience the emergence of super-spreaders.

5.1. Data Processing. The first dataset, Dataset 1, is created from real time-series about the COVID-19 spread in South Korea, obtained from [1]. We choose 23 locations with the most number of people, and construct a closed network of 23 groups by adding each person and their neighbor, determined by the 'infected by' field, to the set of graph nodes $\mathcal{V}$, for a total of ~ 2000 nodes. Additional random edges were created among individuals based on 'city' and 'province'. Then we assigned each individual node n a set of nominal HMM module parameter values $\alpha_n, \beta_n, \nu_n, \gamma_n, s$, and r based on their 'age', 'symptom-onset-date', 'confirmed-date', and 'released-date' fields.

The topology of the second dataset, Dataset 2, is generated more artificially than Dataset 1. We create a graph with four groups, composed of $40, 50, 52$, and 58 individuals, respectively. An edge between two nodes in the same group is made with probability 0.8, while an edge between two nodes in separate groups is made with probability 0.25. We do not connect any edges between nodes in the pairs of groups (Group 2, Group 4) and (Group 3, Group 4). Most of the HMM module true parameters for each individual in a group were generated based on real values from China and USA in the comprehensive contact-tracing data collected from [3]. We choose $s = 0.2$ and $r = 0.979$.

For the observation matrix, we consider the symptoms of 1)fever/headache/migraine, 2) difficulty breathing/blockage in lungs, 3) sore throat/scratchy throat/coughing, 4) sneezing/runny nose/itchy nose, and 5) dead based on the real symptoms observed of a person infected with COVID-19. We assign the following concrete probability values, which were chosen based on the real-data statistics about the symptoms given by the CDC [12].

$$O_n(y_{n,j} = 1|x) = \begin{bmatrix} 0.1 & 0.1 & 0.9 & 0 & 0 \\ 0.05 & 0.05 & 0.65 & 0.01 & 0 \\ 0.07 & 0.07 & 0.73 & 0.01 & 0 \\ 0 & 0 & 0.03 & 0.01 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{(j,x)}$$

5.2. Performance of HMM Model: Parameter Estimation. To test the performance of our HMM model, we consider two performance metrics:

- Binary correlation between the predicted and true sequence of states, i.e., the ratio of correctly-inferred states determined by the Viterbi algorithm of Section 4.2 in comparison to the true health states of the individual.
- ℓ_1-norm difference between true and estimated initial vectors, and average entrywise absolute-value difference between true and estimated transition matrices and observation matrices:

$$\epsilon_P(T) := \frac{1}{|\mathcal{X}|^2} \sum_{x,y} |P_n^{(T)}(x,y) - \hat{P}_n^{(T)}(x,y)| \tag{5.1a}$$

$$\epsilon_O(T) := \frac{1}{|\mathcal{X}| \cdot M} \sum_{x,j} |O_n(y_{n,j} = 1|x) - \hat{O}_n^{(T)}(y_{n,j} = 1|x)| \tag{5.1b}$$

  For the matrices, we choose in this way, as opposed to the standard $\mathcal{L}_1$ or $\mathcal{L}_2$ norms, so that the deviation away from the true matrix caused by a massive error in just one single entry does not throw off the overall error.

We compute these metrics for Dataset 1, with the choices of $a = 0.4$ as the discount factor and uniform weighting across all observations. The performance metric values were accumulated over 10 Monte-Carlo trial simulations; the trajectories of one chosen trial are shown in the top figure of Figure 3. For a simulation time of $T_{\text{sim}} = 167$, the binary correlation from the Viterbi algorithm is 0.9269, and the ℓ_1-norm for the initial state vector, the error of estimation for q_n is given by 0.1452. The evolution over time of the error for estimating the time-varying transition probability matrix according to equation (5.1a) is shown in the bottom figure of Figure 3 for four randomly-chosen nodes in the network. We can see that the average difference is less than 0.15 and speculate that the downwards decline to around 0.01 occurs when the individual is either dead or recovered without any further transitions to the susceptible state; from then on, the transition matrix remains constant, giving the algorithm enough time for the estimate to settle around the true matrix.

5.3. Impact of Network Topology on Virus Propagation. The HMM model can itself be used to model the propagation of disease for graphs that are small enough. We additionally perform an experiment to determine the effect of the population network structure on the speed and breadth of the virus spread. We consider the following three specific edge connection topologies with the node grouping of Dataset 2:

1. *Dense Graph*: the number of edges in each group is between 60% and 80% among all possible edges in the group. The number of edges between any two groups is between 1% and 2% among all possible edges. More specifically, the three groups have 564, 1044, and 1274 many edges respectively.
2. *Sparse Graph*: the number of edges in each group is between 1% and 2% among all possible edges in the group. The number of edges between any two groups is between 1% and 2% among all possible edges. More specifically, the three groups have 12, 25, and 30 many edges respectively.
3. *Tree with Multiple Branches*: each group is a tree, and the number of branches for each node is between 3 and 6. This corresponds to a structure which is somewhere in between the densities of the Dense graph and the Sparse graph.

We visualize the trajectories of each compartment in Figure 4. Even for this smaller network of 200 nodes, the Sparse Graph took approximately an hour to run on a 2.2 GHz Intel Core i7 Macbook Air. This result provides further experimental motivation for why the HMM module should be used strictly for parameter estimation,

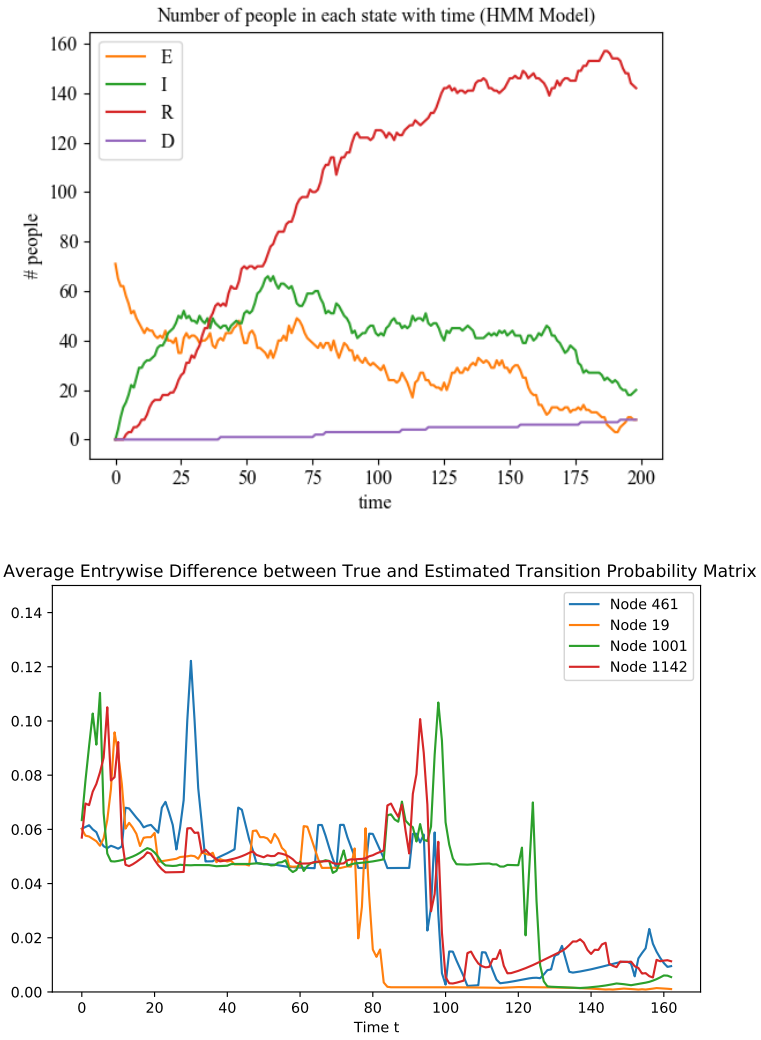

Fig. 3: [Top] A sample Monte-Carlo trial trajectory of the number of exposed (E), infected (I), recovered (R), and deceased (D) individuals over time. The number of susceptible people was omitted in the graph due to its larger magnitude, which overshadowed the rest of the curves. [Bottom] Evolution of average entrywise absolute-value difference over time for four different individuals in the population graph.

and that the ODE module might be better to use for the propagation, in accordance with our discussion in Remark 4.1.

We verify the intuitive hypothesis that reducing the frequency of contacts between people via interventions methods such as quarantine is helpful in preventing the spread of the virus. For a fixed simulation time of $T_{\text{sim}} = 2000$, the pandemic in the Dense Graph continues to spread until all the people have been infected at least once. For the Tree with Multiple Branches, the pandemic ends at time 860 with 146 people

	Group 1	Group 2	Group 3	Group 4
$\alpha^{(e)}$	$0.8 \times 4/14$	$0.8 \times 2/14$	$0.8 \times 1/14$	$0.8 \times 1/28$
$\alpha^{(s)}$	$0.2 \times 4/14$	$0.2 \times 2/14$	$0.2 \times 1/14$	$0.2 \times 1/28$
$\gamma^{(r)}$	$0.979/7$	$0.979/14$	$0.979/28$	$0.979/56$
$\gamma^{(d)}$	$0.021/7$	$0.021/14$	$0.021/28$	$0.021/56$
ν	0.04	0.02	0.01	0.005

Table 1: True parameter values for each group in Dataset 2.

infected at least once, and for the Sparse Graph, it ends at time 799 with only 116 people infected at least once.

5.4. ODE Model Results. Using the parameters estimated via the HMM model in Section 5.3, we now consider the propagation of disease at the large scale. After a similar experiment is performed for Dataset 2 and converting the transition probabilities to transition rates, we use the values given in Table 1 and

$$[\beta_{kj}] = \begin{bmatrix} 0.08 & 0.08 & 0.08 & 0.08 \\ 0.04 & 0.04 & 0.64 & 0 \\ 0.02 & 0.02 & 0.02 & 0 \\ 0.01 & 0 & 0 & 0.01 \end{bmatrix}_{(k,j)}$$

The simulation result is shown in Figure 5, from when three people from each group start out exposed. Note that although it is the largest group, Group 4 suffers the least from the infection because it is disconnected from two of the groups. This aligns with our observations from Section 5.3: the number of infections and corresponding deaths are less when people ensure that they meet with others less frequently than average, justifying the effectiveness of the social isolation policies.

5.5. Super-Spreader Effects. Motivated by the ODE simulation from the previous section, we emulate the effect of *super-spreaders* by accounting for high variability in the contact rate of individuals by applying the stochastic model described in 3.2 to Dataset 1. For the sake of graph simplicity, we reduce the original 23 cliques in the dataset down to two groups: one groups corresponding to highly popular locations which contains about 400 members total, and another group with the remaining 915 individuals. Connections within the larger clique and between the two groups are made sparse. The corresponding $[\beta_{kj}]$ matrix, substituting equation (3.2) for the interactions of people are given by:

$$[\beta_{kj}] := \begin{bmatrix} \beta_0 + \sigma W(t) + Y(t) & 0.0021 \\ 0.01 & 0.005 \end{bmatrix}_{(k,j)}$$

where $\beta_0 = 0.03, \sigma = 0.01, \lambda = 0.0143$, and the jump heights of the compound Poisson are distributed as $\xi_i \sim \text{Unif}[0.5, 2]$. The values of α_k, γ_k are determined by what was estimated previously from the HMM experiment of Section 5.2, and the initial number of people in each compartment are chosen to be $I_1(0) = 26, E_1(0) = 30$ for Group 1, and $I_2(0) = 45, E_2(0) = 63$ for Group 2. For both groups, the remaining individuals are all susceptible.

The trajectories of both groups over time are shown in Figure 6. We observe that Group 2 initially enjoys a reduction in the number of infectious individuals for the

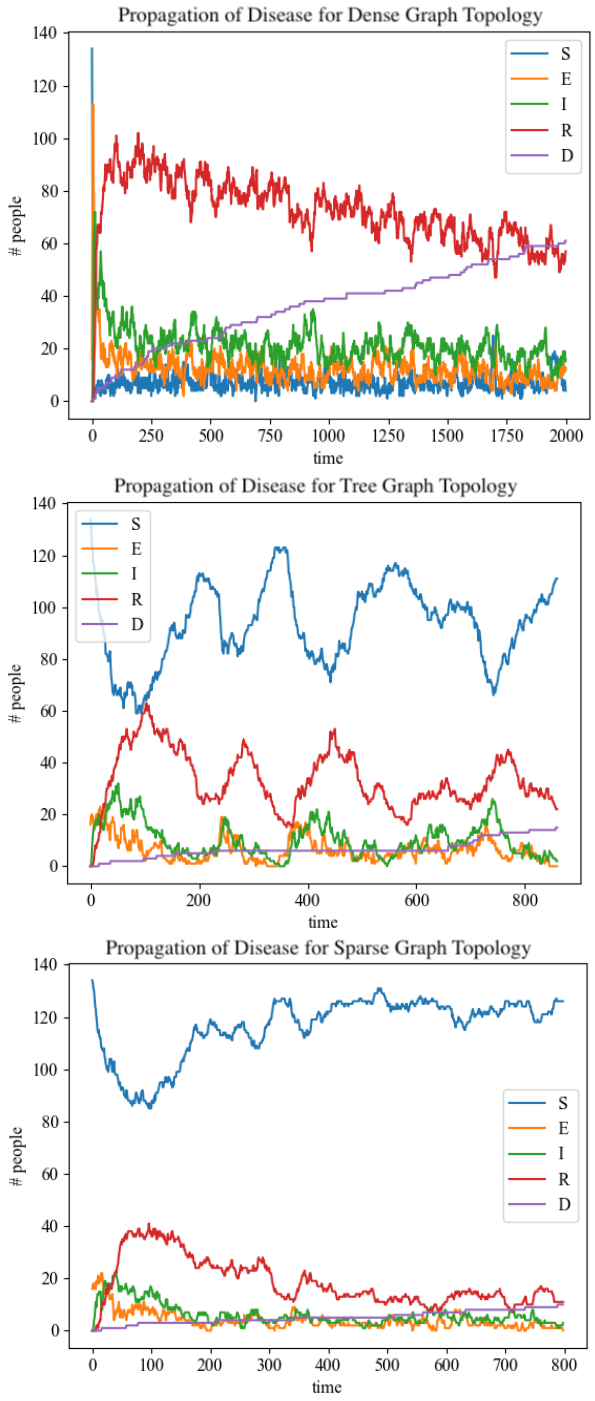

Fig. 4: Emulation via HMM module of the dense, tree, and sparse graphs.

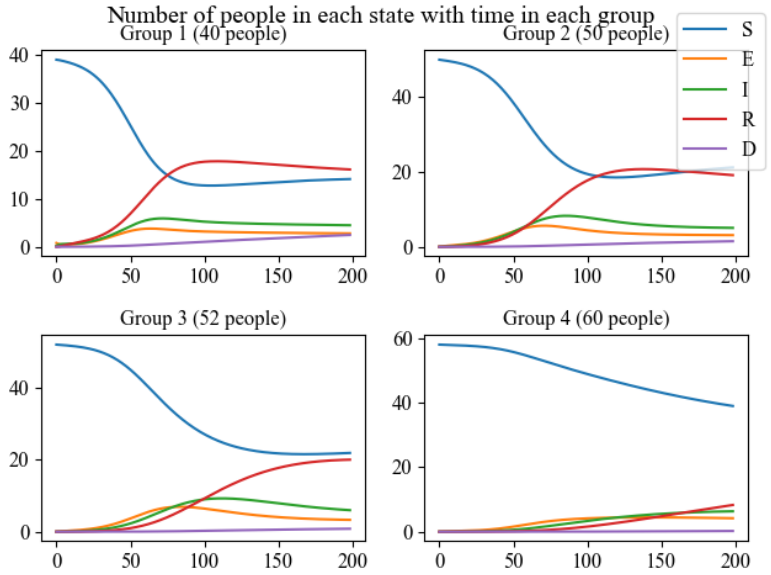

Fig. 5: Emulation via ODE module for each group of Dataset 2.

two months, but increases as a result of coming into contact with members of Group 1, despite the fact that interactions outside of Group 1 are kept at far less than the interactions within Group 1. This experiment verifies an intuitive hypothesis: as a result of the fragile stability, the pandemic ends more slowly than the time it would have taken without the emergence of super-spreaders.

6. Conclusion. In summary, this paper addressed the problem of modeling epidemic spread for control by proposing a novel two-part model consisting of a multi-group ODE-based compartmental model for simulating disease spread across large-scale populations, and an HMM-based compartmental model used to estimate parameters at a small-scale individual level, which are then aggregated and used for the large-scale ODE model. The specific algorithm for parameter estimation invokes a form of forward-backward and Baum-Welch algorithms which account for multiple simultaneous observation processes and time-varying parameters. Additionally, extensions to the stochastic setting have been considered by incorporating Poisson jump noise to account for the effect of possible super-spreaders. The implementation of the model is demonstrated on two datasets which were constructed based on real data from three countries: USA, China, and South Korea. We argue that viewing COVID-19 as a collective case study for control theory leads to a multitude of interesting implications in the general field itself, especially on comparisons between centralized and localized controller designs, as well as carefully accounting for various types of uncertainties in the structure of the system (e.g., topology of a network of subsystems).

REFERENCES

[1] *Data science for COVID-19 in South Korea.* https://www.kaggle.com/kimjihoo/

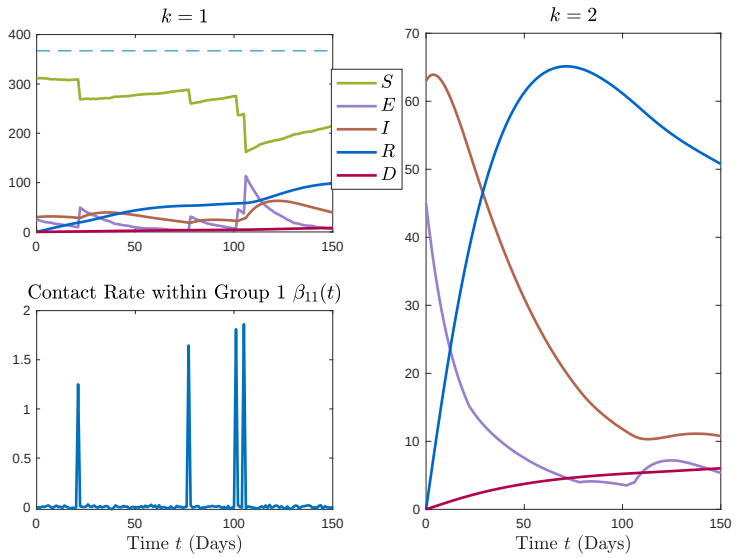

Fig. 6: The spread of virus visualized over the five compartments for two disjoint groups. The rate of interactions within the Group 1 is subjected to stochastic noise.

coronavirusdataset?select=PatientInfo.csv.

[2] *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand.* https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf. Accessed: 2020-03-18.

[3] *Kudos to DXY.cn last update: 03/13/2020, 8:00 PM (EST).* https://threadreaderapp.com/thread/1219661541613531136.html.

[4] *Symptoms of coronavirus*, Center for Disease Control and Prevention (CDC), https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

[5] A. Abou-Ismail, *Compartmental models of the COVID-19 pandemic for physicians and physician-scientists*, S.N. Compr. Clin. Med, (2020), pp. 1–7, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7270519/.

[6] Y. Abu-Mostafa and et.al., *Caltech's CS156 model of COVID-19 trajectory.* http://cs156.caltech.edu/.

[7] C. Anton-Haro, J. A. R. Fonollosa, C. Fauli, and J. R. Fonollosa, *On the inclusion of channel's time dependence in a hidden Markov model for blind channel estimation*, IEEE Transactions on Vehicular Technology, 50 (2001), pp. 867–873.

[8] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, *A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains*, Ann. Math. Statist., 41 (1970), pp. 164–171.

[9] Y. Cai, J. Jiao, Z. Gui, Y. Liu, and W. Wang, *Environmental variability in a stochastic epidemic model*, Appl. Math. Comput., 329 (2018), p. 210–226.

[10] Y. Cai, Y. Kang, M. Banerjee, and W. Wang, *A stochastic sirs epidemic model with infectious force under intervention strategies*, Journal of Differential Equations, 259 (2015), p. 7463–7502.

[11] V. Capasso and G. Serio, *A generalization of the Kermack-McKendrick deterministic epidemic model*, Mathematical Biosciences, 42 (1978), pp. 43 – 61.

[12] Centers for Disease Control and Prevention, COVID-19 Response, *COVID-19 case surveillance public data access, summary, and limitations.* version date: June 27, 2020.

[13] B. J. Coburn, B. G. Wagner, and S. M. Blower, *Modeling influenza epidemics and pandemics: insights into the future of swine flu (H1N1)*, BMC Med., 7 (2009).

[14] V. Colizza, A. Barrat, M. Barthelemy, A. Valleron, and A. Vespignani, *Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions.*, PLoS Med., 4 (2007).

[15] F. Diamond, *Asymptomatic Carriers of COVID-19 Make it Tough to Target*. https://www.infectioncontroltoday.com/covid-19/asymptomatic-carriers-covid-19-make-it-tough-target. Accessed: 2020-03-23.
[16] W. Dong, A. S. Pentland, and K. A. Heller, *Graph-coupled hmms for modeling the spread of infection*, in Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence, UAI'12, AUAI Press, 2012, p. 227–236.
[17] S. Eubank, H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, *Modelling disease outbreaks in realistic urban social networks*, Nature, 429 (2004), pp. 180–184.
[18] M. Fisher and S.-H. Choe, *How South Korea Flattened the Coronavirus Curve*, https://www.nytimes.com/2020/03/23/world/asia/coronavirus-south-korea-flatten-curve.html. Accessed: 2020-03-23.
[19] G. D. Forney, *The Viterbi algorithm*, Proceedings of the IEEE, 61 (1973), pp. 268–278.
[20] J. Gallagher, *Covid reinfection: Man gets Covid twice and second hit 'more severe'*. https://www.bbc.com/news/health-54512034. Accessed: 2020-10-29.
[21] A. Gray, D. Greenhalgh, L. Hu, X. Mao, and J. Pan, *A stochastic differential equation SIS epidemic model*, SIAM Journal on Applied Mathematics, 71 (2011), pp. 876–902.
[22] Y. Gu, *COVID-19 projections using machine leaning*. https://covid19-projections.com/.
[23] A. Gumel, S. Ruan, T. Day, J. Watmough, F. Brauer, P. Driessche, D. Gabrielson, C. Bowman, M. Alexander, S. Ardal, J. Wu, and B. Sahai, *Modeling strategies for controlling SARS outbreak*, Proceedings. Biological sciences / The Royal Society, 271 (2004), pp. 2223–2232.
[24] S. Han and S.-J. Chung, *Incremental nonlinear stability analysis for systems perturbed by Lévy noise*, IEEE Transactions on Automatic Control, (2020, submitted).
[25] H. W. Hethcote and S. A. Levin, *Periodicity in epidemiological models*, Applied Mathematical Ecology, 18 (1989), pp. 193 – 211.
[26] K. Kasulis, *'Patient 31' and South Korea's sudden spike in coronavirus cases*, https://www.aljazeera.com/news/2020/03/31-south-korea-sudden-spike-coronavirus-cases-200303065953841.html. Accessed: 2020-03-23.
[27] W. Kermack and A. McKendrick, *Contributions to the mathematical theory of epidemics—I*, Bulletin of Mathematical Biology, 53 (1991), pp. 33 – 55.
[28] L. R. Rabiner, *A tutorial on Hidden Markov Models and selected applications in speech recognition*, Proceedings of the IEEE, 77 (1989), pp. 257–286.
[29] M. Salathe, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, *A high-resolution human contact network for infectious disease transmission*, in Proceedings of the National Academy of Sciences (PNAS), Dec 2010.
[30] M. Serhani and H. Labbardi, *Mathematical modeling of COVID-19 spreading with asymptomatic infected and interacting peoples*, J. Appl. Math. Comput., (2020), pp. 1–20, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431117/.
[31] H. E. Soper, *The interpretation of periodicity in disease prevalence*, Journal of the Royal Statistical Society, 92 (1929), pp. 34 – 73.
[32] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Regis, J. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems, *Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees*, BMC Medicine, 9 (2011).
[33] Xiaolin Li, M. Parizeau, and R. Plamondon, *Training hidden Markov models with multiple observations-a combinatorial method*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2000), pp. 371–377.
[34] O. Younis and S. Fahmy, *Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach*, in IEEE INFOCOM 2004., vol. 1, 2004.
[35] R. Yu and Y. Ma, *DeepGLEAM*. https://jacobsschool.ucsd.edu/news/release?id=3129.