MODELING AND ESTIMATION OF EPIDEMIC SPREAD IN POPULATION NETWORKS USING COMPARTMENTAL ODES AND COUPLED HIDDEN MARKOV MODELS*

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

Abstract. Accurate modeling for predicting the long-term effects of epidemic spread is an important problem in the field of control engineering, and the recent global pandemic of the novel SARS-2 coronavirus (SARS-CoV-2) has renewed interest worldwide. Many existing models suffer from a lack of heterogeneity; predictions for large populations are achieved by abstracting away important interaction dynamics while detailed individual-level approaches can only handle small populations. This paper presents a novel multiscale model that combines both large-scale population mitigation and small-scale monitoring at the individual-level by taking advantage of repetitive community structures present in the network of populations. A multi-group compartmental model with ordinary differential equation (ODE) dynamics is used to emulate large-scale disease spread across each community, and a coupled Hidden Markov Model (CHMM) is used to emulate the evolution of an individual's health status. Data such as infected/death counts and contact-tracing are used in combination to estimate the parameters of the model: transition rates, probabilities, and edge weights in the community network. Based on worldwide observations of SARS-CoV-2 since 2020, our technical contributions are two-fold. First, we employ the SEIRD compartmental model (an extension of the more well-known SIR/SIRS models by separating asymptomatic infections from symptomatic infections and including deaths); second, we extend the usual forward-backward algorithm for CHMMs so that multiple simultaneous observation sequences and time-varying parameters are accounted for. 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.

The devastating impact of the COVID-19 disease from the novel SARS-2 coronavirus (SARS-CoV-2) on the economy and health of societies worldwide has lasted for over two years and is arguably ongoing as of June 2022, especially with the chronic emergence of new virus mutations [12] and superspreader phenomena [25]. Even with recent advances in virtual technology, mitigation policies for maximizing public safety is consistently at odds with maintaining the flow of the global economy, and selective re-opening of businesses in the midst of the pandemic remains a subject of debate. For example, draconian strategies like lockdowns, while perhaps the most effective in preventing spread, cannot be enforced long-term due to leaving millions of people without work. Furthermore, factors such as culture, climate, and geography differ by population, which suggests the importance of designing policies which are heterogeneous despite striving towards the same goal of simultaneously minimizing health risk and maximizing economic profit.

Mathematical models (e.g., [19, 33]) have been shown to be quite effective in informing the design of mitigation policies for various past outbreaks such as the influenza pandemics [15, 16] and the SARS epidemic of 2003 [24]. These models may be used to design policies for COVID-19 in similar ways, but with the unique characteristics of SARS-CoV-2 accounted for: the mutations into various strains and

*Submitted to the editors xx x, 2022.

†Department of Computer Science, Columbia University.

‡Division of Engineering and Applied Science, California Institute of Technology. Corresponding author: SooJean Han {soojean@caltech.edu}

resurgent superspreaders caused by its long incubation period. An added challenge is a lack of worldwide consensus surrounding the vaccines presently in circulation, which places high emphasis on non-medical intervention strategies like social distancing and mask-wearing. On the other hand, rapid developments in technology have made large-scale data collection and accessibility possible; as a result, organized active efforts such as frequent contact tracing and close monitoring of individuals are conceivable and have also been proven effective in mitigating the spread [20].

1.1. Related Work. The seminal work [27] in 1927 introduced the original "compartmental model" for emulating epidemic/pandemic dynamics. A compartmental moves the individuals in a population through the different phases of illness (called "compartments") undergone upon exposure to the virus; mathematically, this movement is represented by a set of ordinary differential equations (ODEs). Since then, a variety of extensions have been made to this traditional ODE-based approach, including stochastic factors [31], nonlinear interactions with the infected [13], and temporary immunity [26] 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]). One of the most popular compartmental models for COVID-19 is the SEIR compartmental model [10] which distinguishes between asymptomatic and symptomatic infectious individuals, which is important because individuals may still be able to transmit the virus despite not being ill. However, a well-known disadvantage of compartmental models is an abstraction of heterogeneous population characteristics. In particular, compartmental models do not consider the structural topology of the population network, i.e., individuals in each compartment are assumed to interact uniformly at random with each other. These simplifications yield approximations which may be too coarse, especially for very large populations.

An alternative approach which addresses these pitfalls of compartmental models are coupled hidden Markov models (CHMMs), e.g., [18,29,32], which models the population network at an individual level. The interaction dynamics of the population are incorporated through the chains in the CHMM structure, and since each individual can be made to have different characteristics (e.g., immune response), other forms of heterogeneity can be incorporated. An added benefit to the CHMM architecture is that there are standard algorithms which enable the inference of unknown parameters (e.g., death rate, recovery rate) from observational data such as contact-tracing information; multiple separate observation processes may be incorporated [34] and the parameters to be estimate may be time-varying [7]. However, although finer-grained approximations may be obtained by modeling epidemic spread via CHMMs, they are not scalable to large networks. It is impractical to keep track of each individual's unknown parameters separately; for example, in models such as [23], grid search may be used to estimate the unknown parameters, but this will take time if the initial grid space is too large. Moreover, CHMMs still suffer from the same issue as compartmental models in that repetitive community structures in the topology of the network are not leveraged for efficient prediction and estimation.

The resolution of the model may be traded off between compartmental models and CHMMs by invoking multiscale models, which consider disease propagation at both the population-level and at the individual-level, e.g. [9]. With increasing computational power and more available data over the years, 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 time. Furthermore, the topological structure of the population network is still often not considered in any

of these approaches.

1.2. Our Contributions. Motivated by the above, we propose a novel multiscale model for the propagation of SARS-CoV-2 throughout a population by exploiting community structures in the interaction network and assigning heterogeneous mitigation strategies correspondingly. The multiscale model consists of two *modules*. First, the propagation of COVID-19 across a large-scale population network is modeled as a multi-group extension to the standard *SEIRD compartmental ODE model*, where the groups are referred to as communities in the context of this paper. Second, tracked individuals in each community and their local interaction dynamics are modeled as a *coupled hidden Markov model (CHMM)* where the hidden states correspond to the health statuses of the individual and the observations are their symptoms. Using this model, we address two primary questions:

1. how can we incorporate multiple scales of available data (e.g., large-scale infected/death counts and small-scale contact-tracing) in order to estimate the parameters associated with the model: the transition rates of each compartmental model, the transition probabilities of each CHMM, and the edge weights in the community network?
2. how can we use the trained multiscale model to predict important statistics, e.g., the total number of deaths by a certain time?

One primary novelty of this model is the consideration of the topology of the network and leveraging community structural patterns to achieve both estimation accuracy and population size scalability. With our proposed model, we aim for the development of a more model-based analytical approach, onto which machine learning methods may be added to enhance performance. We use experiments to demonstrate the estimation and prediction performance of our model on real-world COVID-19 data.

1.3. Paper Organization. In Section 2, we describe how to transform a given population into a graph network of communities and individuals for the purposes of our model, and we introduce the SEIRD phases of the COVID-19 disease. In Section 3, we outline the SEIRD compartmental ODE module, the larger scale of out multiscale model, and mathematically define the transition rates and the ODE dynamics. Next, in Section 4, we outline the coupled HMM module, the smaller scale of our multiscale model, and define the transition probabilities per individual. In particular, for estimating the parameters of the HMM module, we emphasize and extension of the Baum-Welch (expectation-maximization) algorithm in two key ways: incorporating multiple observation processes simultaneously, and accounting for time-varying parameters. In Section 5, we discuss how the two levels of available data, infected/death counts and individual-level contact-tracing, are used together to estimate the parameters of both parts of the multiscale model. In Section 6, we demonstrate the application of our multiscale model to multiple experiments, using two datasets which were constructed based on real-world data available online. Finally, the conclusion of the paper is made in Section 7.

2. Preliminaries. The overall multiscale model proposed in this paper is summarized in Fig. 1. Before we discuss each part of the model in detail, we set up and introduce two concepts: first, the representation of the population as a network of communities and second, the SEIRD phases of the COVID-19 disease. These model choices are used in the design of both the large-scale compartmental model and the small-scale CHMMs throughout the paper.

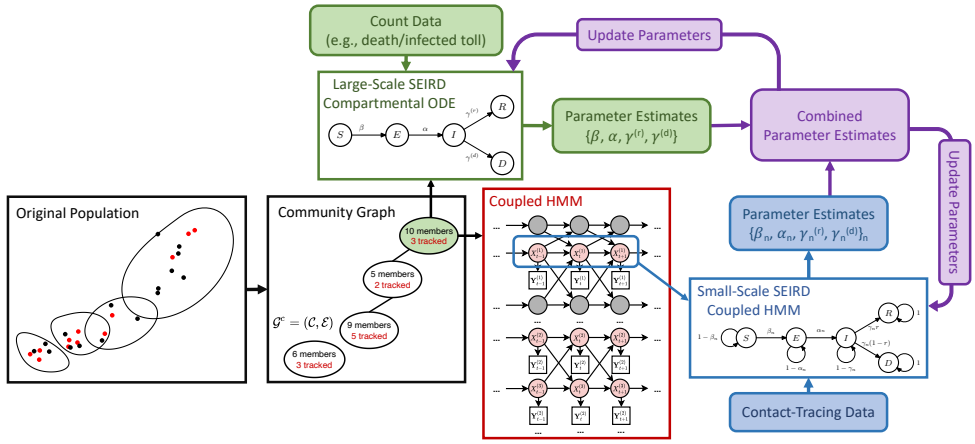

Fig. 1: A flow diagram representation of the multiscale model for parameter estimation of COVID-19. Two entities are identified from the *original population* of individuals (dots): the different communities the individuals belong in (black circles), and the *tracked individuals*, i.e., those whose *contact-tracing data* is available (red dots). The resulting conversion is referred to as the *community graph* (see Section 2.1). Each community in the community graph is modeled according to the large-scale *SERID compartmental ODE model* (in green), and the parameters (i.e., transition rates) of the model is estimated using the available *count data* for the community (see Section 3). The interaction network of all individuals in each community is extracted as a *coupled hidden Markov model* (CHMM, in red), where the hidden state is defined as the individual's health status. Specifically, the SEIRD compartments for the different phases of COVID are embedded into each CHMM, and so each tracked individual is modeled according to a small-scale *SEIRD CHMM* (in blue). Gray nodes represent untracked individuals, whose health status remains unknown, and red nodes represent the health status of a tracked individual, which can be identified through observed symptoms (see Section 4). Contact tracing data is used in each tracked individual's CHMM to estimate his/her unique parameters (i.e., transition probabilities, true health status sequence). The estimates of the model parameters for the overall population network are then combined (in purple) via averaging; both the compartmental ODE and individual CHMM parameters are updated (see Section 5), and the models are then used for predicting quantities such as the future death toll.

2.1. The Community Network Graph. Let $\mathcal{V}$ denote the set of individuals in the population, with a constant cardinality $N := |\mathcal{V}|$. We are primarily interested in the propagation of disease throughout a large-scale network, meaning $|\mathcal{V}|$ is very large. The interaction network of this population is represented as a network of *communities* $\mathcal{G}^c := (\mathscr{C}, \mathcal{E})$, where $\mathscr{C} := \{\mathcal{C}_k\}_{k=1}^K$, where $K \leq |\mathcal{V}|$ is the total number of communities and $\mathcal{E}$ is the set of edges. An edge $e(i,j) \in \mathcal{E}$ is created between two communities C_i and C_j if the two communities share at least one member; the edge is assigned a weight $w(i,j) \in \mathbb{R}^{\geq 0}$ proportional to the number of members they share. Let $N_k := |\mathcal{C}_k|$ be the cardinality of community k, assumed constant, and denote $\mathscr{N}_k := \{\mathcal{C}_j \in \mathscr{C} | e(k,j) \in \mathcal{E}\}$. Every community forms a fully-connected subgraph where the nodes are the individuals. The interpretation of the communities is geographic, in that each community is assigned a certain *type* and each individual may belong to more than one community. For simplicity of presentation, we consider only *static types* in which the membership of individuals is mostly constant over time (e.g., routine

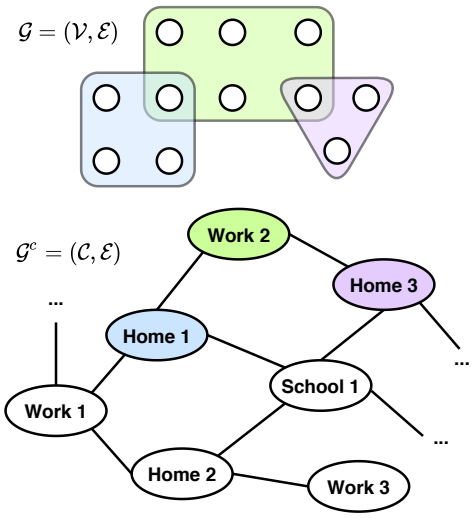

Fig. 2: The full graph network of the individuals and the corresponding communities. Here, there are $M = 3$ types of communities, "work", "school", and "home".

locations such as schools, workplaces, private residences), as opposed to dynamic types where the membership of individuals may vary (stochastically) over time (e.g., special-occasion locations such as restaurants, amusement parks, musical concerts). Let $M \in \mathbb{N}$ be the total number of community types with $\mathscr{C} := \cup_{j=1}^{M} \mathscr{C}^{(j)}$, where $\mathscr{C}^{(j)}$ is the set of all communities of type j. For each individual $n \in \mathcal{V}$, let $\mathscr{V}_n := \{\mathcal{C} \in \mathscr{C} | n \in \mathcal{C}\}$ be the set of communities that n belongs to. A visual representation of the two graphs with $M = 3$ types of communities is shown in Fig. 2.

Assumption 2.1. For all $k \in \{1, \cdots, K\}$, the members of community k remain as members of community k, with no new members joining and no present members departing.

2.2. SEIRD: The Phases of COVID-19. The evolution of COVID-19 spread within each community in $\mathcal{G}^c$ is modeled according to five main phases $\mathcal{X} := \{\mathtt{S}, \mathtt{E}, \mathtt{I}, \mathtt{R}, \mathtt{D}\}$. 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 (i.e., asymptomatic infected); they are carriers of the virus without any noticeable symptoms of the disease, but they are still capable of infecting susceptible individuals. After a certain period of time, individuals would either begin 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 now considered invincible to the virus.

The SEIRD model generalizes the simpler but more standard SIRS model [27] by 1) distinguishing between recovered and deceased individuals, and 2) distinguishing two types of infectious individuals separately as exposed (i.e., asymptomatic infected) and ill (i.e., symptomatic infected) individuals. Note that both exposed and ill individuals are capable of spreading the virus to susceptible individuals; in the case of

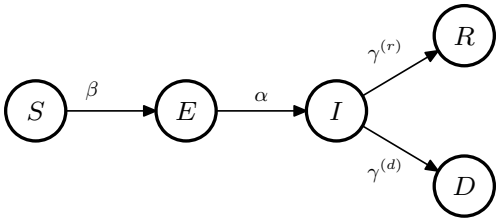

Fig. 3: SEIRD for the ODE module.

Singapore, as of March 17, 2020, it is reported that roughly 48% of individuals who are ill with COVID-19 had been infected from exposed people alone [17]. However, in the case of COVID-19, it is important to maintain (E) and (I) as separate phases because of the time delay taken for symptoms to show; COVID-19 symptoms typically present within 2-14 days upon exposure [4]. For this reason, SEIRD captures a more accurate description of the COVID-19 disease compared to other types of compartment models [22].

DEFINITION 2.2 (Infectious). *An **infectious** individual is one in which any other susceptible individual 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 exposed and ill individuals are considered infectious compartments.*

3. Compartmental ODE Module: Large-Scale Disease Propagation. We begin by designing the ODE module, which is used to emulate virus spread throughout the community network $\mathcal{G}^c$. Traditional compartmental models assume random uniform interactions among individuals in a population, and do not consider groups, communities, or other heterogeneous interaction dynamics. This may lead to coarse approximations of quantities such as the predicted death toll, especially when a population is very large. In contrast to the traditional approach, we partitioned the population into a network of strongly-connected communities; this allows us to introduce basic heterogeneity into the model, distinguishing inter-community rate of spread from the intra-community rate of spread by changing the edge weights of $\mathcal{G}^c$.

For the purposes of modeling the COVID-19 pandemic, we choose to employ the five-compartment *SEIRD compartmental model*, visualized in Fig. 3, where $\mathcal{X} := \{S, E, I, R, D\}$ are the phases described in Section 2.2. Mathematically, the ODE dynamics for a SEIRD model is governed by the following set of differential equations:

$$\frac{dS(t)}{dt} = -\beta(I(t) + E(t))S(t) \tag{3.1a}$$

$$\frac{dE(t)}{dt} = \beta(I(t) + E(t))S(t) - \alpha E(t) \tag{3.1b}$$

$$\frac{dI(t)}{dt} = \alpha E(t) - (\gamma^{(r)} + \gamma^{(d)})I(t) \tag{3.1c}$$

$$\frac{dR(t)}{dt} = \gamma^{(r)} I(t) \tag{3.1d}$$

$$\frac{dD(t)}{dt} = \gamma^{(d)} I(t) \tag{3.1e}$$

Here, the parameters are the *transition rates*, given by $\boldsymbol{\theta} := [\beta, \alpha, \gamma^{(r)}, \gamma^{(d)}]^\top$. We

define β to be the contact rate between exposed and ill individuals with susceptible individuals in the population, α is the rate at which exposed individuals fall ill, and $\gamma_k^{(z)}$ is the rate that ill individuals of community k leave their ill status, becoming recovered if $z = r$ or deceased if $z = d$.

Assumption 3.1. 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. We also consider the case where the network is closed, in that no interaction with any individuals outside of the population captured by the community network is allowed.

The SEIRD model for multiple communities (and multiple community types) is a simple extension of the original compartmental model (3.1). The compartmental model for the entire community network $\mathcal{G}^c$ can be visualized as K identical copies of Fig. 3, connected to each other according to the topology imposed by $\mathcal{G}^c$. Towards this end, we define the parameter vector for community k as $\boldsymbol{\theta}_k := [\{\beta_{kj}\}_{j\in\{1,\cdots,K\}}, \alpha_k, \gamma_k^{(r)}, \gamma_k^{(d)}]^\top$. Here, $[\beta_{kj}] \in \mathbb{R}^{K\times K}$ (with $\{\beta_{kj}\}_{j\in\{1,\cdots,K\}}$ its vectorized form for each $k \in \{1, \cdots, K\}$) defines the contact rate between each community of the network. Namely, β_{kk} denotes the rate of contact among susceptible and infectious members which belong in the same community k, while β_{kj} denotes the average rate of contact among susceptible individuals in community k and infectious individuals in community j for all other communities $j \neq k$. The intuition for this definition is that a susceptible individual in some community k has at most K different ways of becoming infected (i.e., by coming into contact with a individual of any other community), depending on the topology of the community network.

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

$$\frac{dS_k(t)}{dt} = -\sum_{j=1}^{K} \frac{\beta_{kj}(I_j(t) + E_j(t))}{N_j} S_k(t) \tag{3.2a}$$

$$\frac{dE_k(t)}{dt} = \sum_{j=1}^{K} \frac{\beta_{kj}(I_j(t) + E_j(t))}{N_j} S_k(t) - \alpha_k E_k(t) \tag{3.2b}$$

$$\frac{dI_k(t)}{dt} = \alpha_k E_k(t) - (\gamma_k^{(r)} + \gamma_k^{(d)}) I_k(t) \tag{3.2c}$$

$$\frac{dR_k(t)}{dt} = \gamma_k^{(r)} I_k(t) \tag{3.2d}$$

$$\frac{dD_k(t)}{dt} = \gamma_k^{(d)} I_k(t) \tag{3.2e}$$

where the parameters are defined previously.

Assumption 3.2. Each community has at least one tracked individual.

Remark 3.3. The contact parameter β_{kj} between communities k and j is proportional to the edge weight $w(i, j)$, which describes the strength of the connection between the two communities. The weight $w(i, j)$ increases when the two communities interact more frequently with each other.

Now, we describe a method for estimating the transition rate parameters in $\boldsymbol{\theta}_k$ for each community k based on count data. First, the estimates $\{\tilde{\gamma}_k^{(r)}, \tilde{\gamma}_k^{(d)}\}$ of community k's true rates of recovery and death $\{\gamma_k^{(r)}, \gamma_k^{(d)}\}$ can be determined by observing

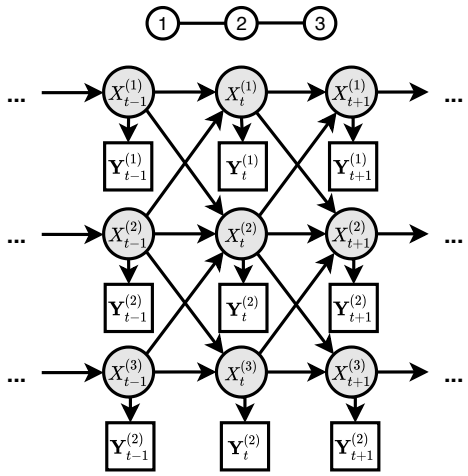

Fig. 4: A sample coupled hidden Markov model relating the health statuses of three individuals in a community, along with the observed symptoms.

time-series data of its recovered and death tolls over time, which are some of the most well-reported statistics pertaining to COVID-19. Second, estimating α_k is difficult from count data alone since, unlike the recovery and death tolls, $E_k(t)$ is difficult to obtain due to the lack of visible symptoms in exposed individuals. We thus rely entirely on contact-tracing data from individuals in $\overline{\mathcal{C}}_k$ to estimate $alpha_k$; this estimation procedure is detailed further in Section 5. Finally, estimating the rates $[\beta_{kj}]_{k,j\in\{1,\cdots,K\}}$ in which susceptible individuals of each community transition to becoming exposed (asymptomatic infected) is tricky due to their dependence on other communities, and we also rely on individual contact-tracing data. Thus, we also defer discussion of their estimation to Section 5.

4. CHMM Module: Parameter Estimation from Individuals. While the SEIRD compartmental model is used to propagate the health status of each community, *coupled hidden Markov models (CHMMs)* [11] are used to represent the health status of certain tracked individuals. Detailed contact-tracing data from a single individual offers further insight into the parameters of the dynamics of the disease spread within smaller subsets of each community. Let $X_n(t)\in\mathcal{X}$ be the hidden state of the HMM for individual $n\in\mathcal{V}$ for all time $t\in\mathbb{N}$, with $\mathcal{X}$ the set of possible phases defined in Section 2.2; in the context of epidemic modeling, $X_n(t)$ is the health status of individual $n\in\mathcal{V}$ at each time $t\in\mathbb{N}$. A *chain* $n\in\mathcal{V}$ of the CHMM is a sequence of hidden states $\{X_n(t), t\geq 0\}$ over time for the single individual v, and they are connected to the chains corresponding to the other individuals v interacts with. We 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 $0<t_1<t_2$ to be the sequence of health statuses for individual $n\in\mathcal{V}$ between times t_1 and t_2, and the vector of deterministic values it takes as lowercase $\mathbf{x}_n(t_1:t_2):=(x_n(t_1), x_n(t_1+1),\cdots, x_n(t_2))$.

Although the true health status of any individual may be unknown (especially when (s)he is in the (E) phase), each individual releases a set of observable symptoms (e.g., fever, sore throat, etc.) which can be used to determine whether or not (s)he is infectious (exposed or ill). Suppose we are interested in keeping track of $B\in\mathbb{N}$ specific symptoms of COVID-19; for simplicity, we assume that they oc-

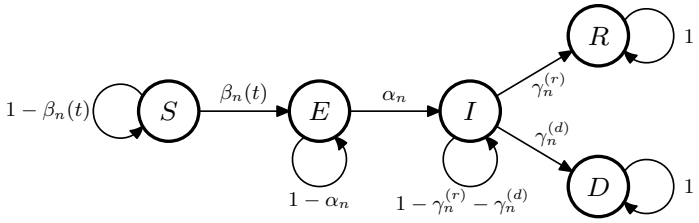

Fig. 5: The underlying Markov chain for a single chain of the CHMM module, using transition probabilities as parameters.

cur independently of each other. Let $Y_n(t) := [Y_{n,1}(t), \cdots, Y_{n,B}(t)]^\top \in \{0,1\}^B$ represent the observed symptoms vector of individual $n \in \mathcal{V}$ at time t, where $Y_{n,j} = 0$ if the individual does not exhibit symptom $j \in \{1, \cdots, B\}$, and 1 otherwise. Further define the *observation probability matrix (OPM)* $O_n(Y_{n,\cdot}(t)|\cdot) \in \mathbb{R}^{B \times |\mathcal{X}|}$ for individual $n \in \mathcal{V}$, with entries $O_n(Y_{n,j}(t)|X_n(t))$ denoting the probability of observing the positive presence symptom j in individual v at time t, i.e., $Y_{n,j}(t) = 1$. The boldfaced notation for $\mathbf{X}$ and $\mathbf{x}$ also extends to $\mathbf{Y}_n, \mathbf{y}_n$ in the same way, as $\{\mathbf{Y}_n(0:t) = \mathbf{y}_n(0:t)\} \equiv \{\mathbf{Y}_{n,1}(0:t) = \mathbf{y}_{n,1}(0:t), \cdots, \mathbf{Y}_{n,B}(0:t) = \mathbf{y}_{n,B}(0:t)\}$.

Assumption 4.1. For each $n \in \mathcal{V}$, the OPM O_n is given and known.

A sample CHMM of a specific community with three individuals, not strongly-connected, is visualized in Fig. 4. Because contact-tracing data only provides us information about the evolution of the observed symptoms over time for a subset of tracked individuals, the transition probabilities among the different phases in $\mathcal{X}$ are unknown. Each individual is assigned a vector of unknown parameters similar to $\boldsymbol{\theta}_k$ for the compartmental model of each community $k \in \{1, \cdots, K\}$. For individual $n \in \mathcal{V}$, the full vector of transition probability parameters is given by $\boldsymbol{\eta}_n(t) := [\beta_n(t), \alpha_n, \gamma_n^{(r)}, \gamma_n^{(d)}]$, and the sparsity pattern of the *transition probability matrix (TPM)* corresponding to the chain of v is given by

$$P_n(t) := \begin{bmatrix} 1-\beta_n(t) & \beta_n(t) & 0 & 0 & 0 \\ 0 & 1-\alpha_n & \alpha_n & 0 & 0 \\ 0 & 0 & 1-\gamma_n^{(r)}-\gamma_n^{(d)} & \gamma_n^{(r)} & \gamma_n^{(d)} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.1}$$

Note that the probability for transitioning from `S` to `E` is time-varying because it is dependent on the time-varying health statuses of his/her immediate neighbors.

We address two specific questions for each individual $n \in \mathcal{V}$: given a complete sequence $\{\mathbf{y}_n(0 : T_{\text{sim}})\}$ of observed symptoms over some time duration $T_{\text{sim}} > 0$

1. how can we estimate the values of the TPM $P_n(t)$ in the CHMM?
2. given the TPM estimates $\hat{P}_n(t)$ for all $t \in [0, T_{\text{sim}}]$, how can we estimate the true health status $x_n(0 : t)$?

For Question 1, the *forward-backward algorithm* (e.g., [28]) and *Baum-Welch (expectation-maximization)* (e.g., [8]) are standard procedures in the HMM literature which can and estimate the transition and observation probabilities in P_n and O_n. A brief review of both methods is provided in Appendix A; for the purposes of this paper, we make two simultaneous extensions: 1) multiple different time series of observations can be incorporated at once, and 2) the unknown parameters are assumed to be time-varying.

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:

$$g_{n,j}(t,x) := \mathbb{P}(X_n(t) = x | \mathbf{Y}_{n,j}(0:T_{\text{sim}}) = \mathbf{y}_{n,j}(0:T_{\text{sim}})) \tag{4.2a}$$

$$h_{n,j}(t,x,z) := \mathbb{P}(X_n(t) = x, X_n(t+1) = z | \mathbf{Y}_{n,j}(0:T_{\text{sim}}) = \mathbf{y}_{n,j}(0:T_{\text{sim}})) \tag{4.2b}$$

The values in (4.2) can be computed using the standard forward-backward recursion described in Appendix A.2. In order to account for time-varying parameters, 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. To aggregate multiple observations into a single definitive estimate, define $\mathbf{w} \in \mathbb{R}^B$ to be weights such that $\sum w_j = 1$.

$$\hat{P}_n(t,x,z) = \left(\sum_{j=1}^{B} w_j \sum_{s=0}^{t-1} a^{t-s} h_{n,j}(s,x,z) \right) \left(\sum_{j=1}^{B} w_j \sum_{s=0}^{t} a^{t-s} g_{n,j}(s,x) \right)^{-1} \tag{4.3}$$

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.

Question 2 can be addressed by applying the standard *Viterbi algorithm* (e.g., [21], see Appendix B for a brief review) to each separate observation sequence, then aggregating them in the following way. Let $\hat{x}_{n,j}(t) \in \mathcal{X}$ be the most likely health status of individual $n \in \mathcal{V}$ at time $t \in [0, T_{\text{sim}}]$ given observation process $j \in \{1, \cdots, B\}$. The health status $\hat{x}_n(t)$ determined by considering all observation processes simultaneously is then given by whichever phase in $\mathcal{X}$ occurs most often in the aggregate set $\{\hat{x}_{n,1}(t), \cdots, \hat{x}_{n,B}(t)\}$. 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.

5. Parameter Estimation with the Two Modules. The relationship between the small-scale CHMM model (Section 4) and the large-scale compartmental model (Section 3) can be illustrated as follows. An individual in community k who has health status `S` at time t is counted in the number $S_k(t)$. When (s)he transitions to health status `E` at time $t+1$, we decrement $S_k(t+1) = S_k(t) - 1$ and increment $E_k(t+1) = E_k(t) + 1$. This relationship suggests that both count data (e.g., the number of total infections/deaths) and the individual contact-tracing data can be used to estimate the transition rates of the compartmental model.

Define $\tilde{\theta}_k(t) := [\{\tilde{\beta}_{kj}(t)\}_{j \in \{1,\cdots,K\}}, \tilde{\alpha}_k(t), \tilde{\gamma}_k^{(r)}(t), \tilde{\gamma}_k^{(d)}(t)]^\top$ to be the estimate of parameter vector $\boldsymbol{\theta}_k$ for the specific community C_k, $k \in \{1, \cdots, K\}$, at time $t \in (0, T_{\text{sim}}]$. Similarly, define $\hat{\boldsymbol{\eta}}_n(t) := [\hat{\beta}_n(t), \hat{\alpha}_n(t), \hat{\gamma}_n^{(r)}(t), \hat{\gamma}_n^{(d)}(t)]^\top$ to be the estimate of the parameter vector $\boldsymbol{\eta}_n(t)$ for the tracked subset $\overline{\mathcal{V}} \subseteq \mathcal{V}$ of the total population. Note that even for original parameters which are constant (e.g., α_k, α_n, etc.), we consider a time-varying estimate due to the time-varying nature of the estimated TPM $\hat{P}_n(t)$. Moreover, the transition probability estimates in the small-scale CHMM model and the transition rate estimates in the large-scale compartmental model need

to be reconciled. The conversion from any probability p to any rate r for is done through the standard formula $p := 1 - e^{-r}$.

First, by the structure of the estimated TPM $\hat{P}_n(t)$, we can estimate some parameters by matching the corresponding entries: $\hat{\beta}_n(t) = \hat{P}_n(t, \mathtt{S}, \mathtt{E})$, $\hat{\alpha}_n(t) = \hat{P}_n(t, \mathtt{E}, \mathtt{I})$, $\hat{\gamma}_n^{(r)}(t) = \hat{P}_n(t, \mathtt{I}, \mathtt{R})$, $\hat{\gamma}_n^{(d)}(t) = \hat{P}_n(t, \mathtt{I}, \mathtt{D})$. To estimate the $\{\tilde{\beta}_{kj}(t)\}_j$ parameters, first define $\kappa_n(t)$ to be the number of infectious tracked neighbors of $n \in \overline{C}_k$ in community C_k. To get an estimate $\hat{\kappa}_n(t)$ of $\kappa_n(t)$, individual n exchanges Viterbi algorithm results with his/her tracked neighbors $\{\hat{x}_{m,j}(t), m \in \mathcal{N}(n)\}$, then simply counts the number of people inferred to be infectious.

Note the probability $\beta_n(t)$ ($\hat{\beta}_n(t)$) for individual n to transfer from $\mathtt{S}$ to $\mathtt{E}$ is dependent on $\kappa_n(t)$ ($\hat{\kappa}_n(t)$). A sensible choice for $\beta_n(t)$ should meet two criteria: 1) it should be impossible for n to become exposed if $\kappa_n(t) = 0$, and 2) it should be a function which increases with $\kappa_n(t)$. One such choice, from [18] is a modified Beta distribution; in this paper, we choose a straightforward probabilistic relationship:

$$\hat{P}(t, \mathtt{S}, \mathtt{E}) = \hat{\beta}_n(t) = 1 - \left(1 - \hat{\beta}_n^{(1)}(t)\right)^{\hat{\kappa}_n(t)} \tag{5.1}$$

Here $\beta_n^{(1)} \in [0, 1]$ is the constant probability of a susceptible person becoming exposed upon contact with an infectious individual, and $\hat{\beta}_n^{(1)}(t)$ is is time-varying estimate. We use a superscript 1 to emphasize that the probability corresponds to a *single* susceptible individual and a *single* infectious individual. Because $\kappa_n(t) = 0$ implies $\beta_n(t) = 0$, we specifically consider the case where $\kappa_n(t) > 0$, from which $\hat{\beta}_n^{(1)}(t)$ is determined by algebraic rearrangement of (5.1). Once $\hat{\beta}_n^{(1)}(t)$ is determined for all tracked n, we estimate the $\{\tilde{\beta}_{kj}(t)\}_j$ for each k via

$$\tilde{\beta}_{kk}(t) = -\sum_{n \in \overline{C}_k} \ln(1 - \hat{\beta}_n^{(1)}(t)), \quad \tilde{\beta}_{kj}(t) = -\sum_{n \in \overline{C}_k \cap \overline{C}_j} \ln(1 - \hat{\beta}_n^{(1)}(t)) \tag{5.2}$$

where $k, j \in \{1, \cdots, K\}$ and $\overline{C}_k$ is the set of tracked individuals in community C_k. Similar to the $\hat{\beta}$'s, we rely entirely on contact-tracing data from individuals in $\overline{C}_k$ to estimate $\tilde{\alpha}_k(t)$:

$$\tilde{\alpha}_k(t) := -\sum_{n \in \overline{C}_k} \ln(1 - \hat{\alpha}_n(t)) = -\sum_{n \in \overline{C}_k} \ln(1 - \hat{P}_n(t, \mathtt{E}, \mathtt{I})) \tag{5.3}$$

Note that both the $\{\tilde{\beta}_{kj}(t)\}_j$ and $\tilde{\alpha}_k(t)$ were estimated using only individual contact-tracing data because the lack of visible symptoms in exposed individuals makes it difficult to obtain the counts $E_k(t)$ for each k. In contrast, the ill, recovered, and death tolls have been some of the most frequently reported statistics throughout the worst of the COVID-19 pandemic, and so we can obtain additional estimates from the large-scale count data: The time cycle period for each rate in the compartmental model is one day.

$$\tilde{\gamma}_k^{(r)}(t) := \frac{1}{2}\left(\frac{1}{I_k(t)}(R_k(t+1) - R_k(t)) - \sum_{n \in \overline{C}_k} \ln(1 - \hat{\gamma}_n^{(r)}(t))\right)$$

$$\tilde{\gamma}_k^{(d)}(t) := \frac{1}{2}\left(\frac{1}{I_k(t)}(D_k(t+1) - D_k(t)) - \sum_{n \in \overline{C}_k} \ln(1 - \hat{\gamma}_n^{(d)}(t))\right)$$

for time $t \in [0, T_{\text{sim}})$.

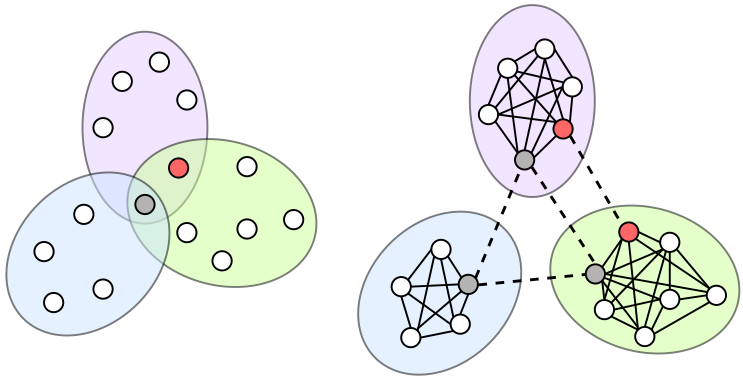

Fig. 6: [Left] Original local community graph for two individuals (gray and red) who are being tracked. [Right] The corresponding graph representation of the local community graph. Communities are strongly-connected, and nodes which are shared among communities are replicated. Edges between replicated nodes (dashed black lines) between them have weights which depend on the mitigation strategy employed.

Remark 5.1. The multiscale model can be reduced to either a full compartmental ODE or a full coupled HMM, models which are prevalent throughout much of the existing literature on epidemic modeling (see Section 1.1 for the review). In particular, the multiscale model can be reduced to a full CHMM model by using the Markov chain dynamics: $X_n(t+1) = x$ with probability $P_n^{(t)}[X_n(t), x]$ for some $x \in \mathcal{X}$; the number of individuals per compartment at each time t can then be obtained by counting the number of individuals which belong in each phase, e.g., $S_k(t) = \sum_{n \in \mathcal{C}_k} \mathbb{1}\{X_n(t) = \mathtt{S}\}$. If the CHMM module replaces the ODE module a means of propagating individuals across each phase of COVID-19, the multiscale model becomes fully stochastic, and further details about intra-community interactions may effectively remove the need to introduce the community structure. However, as mentioned in Section 1, 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. Using coarser community partitions via the ODE module allows us to abstract away an appropriate number of details.

6. Simulation. Premise.

6.1. Experiments Setup.

The numerical experiments throughout this section are performed on two datasets, described below.

Dataset 1: We construct a dataset based off of [1], which is a collection of real-world contact-tracing time series data about the spread of COVID-19 in South Korea during 2020. The communities corresponding to the SEIRD compartmental module from Section 3 are constructed based off of Korea's provinces, so there is only $M = 1$ type of community for this dataset. Individuals within each province are assumed to interact according to a strongly-connected graph, while the edge weights between each pair of provinces are designed based off passenger traffic data collected in 2018 from KOSIS [3]. The total number of individuals considered in this dataset is approximately 5000, and each individual $v \in \mathcal{V}$ is modeled as a CHMM with nominal parameter values $\boldsymbol{\eta}_v(t)$ calculated based on their 'age', 'symptom-onset-date', 'confirmed-date',

and 'released-date' fields from the original contact-tracing dataset.

Dataset 2: We construct a more artificial dataset in the following way. We choose a specific number of communities K, each of some size N_k, $k \in \{1, \cdots, K\}$ which is randomly chosen from some range. From each community $\mathcal{C}_k$, a subset of members belong to more than one community, and another subset $\overline{\mathcal{C}}_k \subseteq \mathcal{C}_k$ of members individual is chosen to be tracked. Some initial distribution of health statuses across the total population is chosen, and the "true" behavior of disease spread throughout the network is emulated by propagating the CHMM of every individual (tracked or not). This generates a sample path of compartment counts over time per community, as well as health status sequences and observed symptoms over time for each tracked individual in the population; this sample path of values is precisely our dataset. In one of our experiments, we consider different parameters (e.g., community structures, proportion of population which is tracked) in order to demonstrate their effects on the forecasting performance of our multiscale model.

Symptoms Observation Matrix: We consider the symptoms of 1) fever/-headache/migraine, 2) difficulty breathing/blockage in lungs, 3) sore throat/scratchy throat/coughing, 4) stomach pain/indigestion/diarrhea,5) sneezing/runny nose/itchy nose, and 6) 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 [14].

$$O_n(y_{n,j} = 1|x) = \begin{bmatrix} 0.1 & 0.1 & 0.9 & 0 & 0 \\ 0.05 & 0.05 & 0.65 & 0.01 & 0 \\ 0.07 & 0.07 & 0.73 & 0.01 & 0 \\ 0 & 0 & 0.03 & 0.01 & 0 \\ 0 & 0.9 & 0.95 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \in \mathbb{R}^{B \times |\mathcal{X}|}$$

That is, the (j, x)th entry of the matrix above defines the probability of observing symptom $j \in \{1, \cdots, B\}$ from individual n (i.e., $y_{n,j} = 1$) given his/her current health status is $x \in \mathcal{X}$.

Performance Metrics: We consider the following metrics to evaluate the performance of our multiscale model on the datasets described above. Specific to the CHMM module, we look at two metrics. First, the *binary correlation between the predicted and true sequence of states*, i.e., the ratio of correctly-inferred states determined by the multi-observation Viterbi algorithm (see Section 4) in comparison to the true health states of the individual.

$$\sum_{t=1}^{T_{\text{sim}}} \mathbb{1}\{\hat{x}_{n,j}(t) = x_{n,j}(t)\} \tag{6.1}$$

Second, we investigate the average entrywise absolute-value difference between true and estimated TPMs for each time $t \in [0, T_{\text{sim}}]$:

$$\frac{1}{|\mathcal{X}|^2} \sum_{x,z} |P_n(t, x, z) - \hat{P}_n^{(t)}(x, z)| \tag{6.2}$$

We choose the metric in this way, as opposed to the standard $\mathcal{L}_1$ or $\mathcal{L}_2$ norms, so that any deviations away from the true matrix caused by a massive error in one single entry do not greatly impact the overall error.

Code Hyperparam	Definition
`K`	Num communities
`community_size_range`	[min, max] number of members per community
`multi_communities_prop`	Prop. of total population in multiple communities
`range_simult_communities`	[min, max] num communities one person can be in
`tracked_prop`	Prop. of people in each community tracked

Table 1: Possible parameters to play around with from the code (artificial dataset).

For parameter estimation of both communities and individuals, we consider mean absolute-value differences:

(6.3)
$$\frac{1}{K}\sum_{k=1}^{K}|\hat{\chi}^{(t)}-\chi|,\ \chi\in\{\alpha_k,\gamma_k^{(r)},\gamma_k^{(d)}\},\quad \frac{1}{\overline{N}}\sum_{n=1}^{\overline{N}}|\hat{\chi}^{(t)}-\chi|,\ \chi\in\{\beta_n(t),\alpha_n,\gamma_n^{(r)},\gamma_n^{(d)}\}$$

where χ is a placeholder variable for the original parameters in $\boldsymbol{\theta}_k$ or $\boldsymbol{\eta}_n$, and $\overline{N}\leq N$ is the number of tracked individuals in the total population.

Prediction and Forecasting: We substitute the estimated parameters $\tilde{\boldsymbol{\theta}}_k(T_{\text{sim}})$ $:=[\{\tilde{\beta}_{kj}(T_{\text{sim}})\}_{j\in\{1,\cdots,K\}},\tilde{\alpha}_k(T_{\text{sim}}),\tilde{\gamma}_k^{(r)}(T_{\text{sim}}),\tilde{\gamma}_k^{(d)}(T_{\text{sim}})]$ obtained from Section 5 into the compartmental module (3.2) in order to forecast the evolution of counts up to some future time $T_{\text{sim}}^{+}>T_{\text{sim}}$. For the purposes of prediction, we keep the rates constant at the final value estimate at time T_{sim}. To determine the accuracy of the prediction, we consider the absolute-value difference between the true and predicted counts $X(t)$ over time $t\in(T_{\text{sim}},T_{\text{sim}}^{+}]$, where X is a placeholder for one of the original compartments $X\in\{I,R,D\}$. Note that the susceptible and exposed individuals are not considered since they are difficult to obtain the true values of in the real-world.

6.2. Preliminary Simulations. Network with $K=10$ communities and $N=116$ individuals in the population. Communities are roughly equally-partitioned: [15, 21, 19, 26, 18, 18, 20, 24, 14, 20]. There are 55 total tracked individuals. See Fig. 7 and Fig. 8.

7. Conclusion. In this paper, we addressed the problem of modeling epidemic spread for control by proposing a novel multiscale model consisting of two main parts: 1) compartmental model which separates among population communities and emulates large-scale disease using ODE dynamics, and 2) a coupled HMM model which is embedded within each individual to emulate the evolution of an his/her health status over time. The novelty in our proposed technique is the combination and extensions of two models which are traditionally used separately in the study of epidemics. Furthermore, the combination of two scales of epidemic modeling allows for the incorporation of multiple scales of data, such as large-scale infected/death counts and individual-level contact-tracing. Model parameters, i.e. transition rates, transition probabilities, and edge weights, are estimated using extensions of standard methods of parameter estimation (e.g., Viterbi's algorithm, Baum-Welch), an the trained model is then used to forecast predictions on the spread of SARS-CoV-2. 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. Additional extensions to the

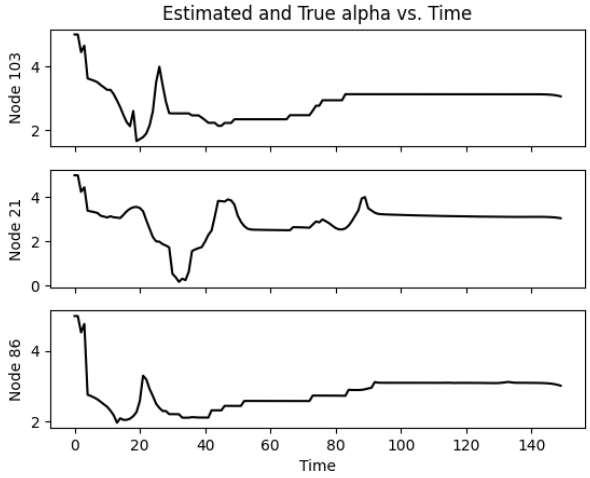

Fig. 7: Absolute-value entrywise difference between TPM estimate $\hat{P}_n(t)$ and true vs time (see (6.2)).

stochastic setting have been considered by incorporating Poisson jump noise to account for the effect of possible super-spreaders. 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).

Appendix A. Estimating the TPM of an HMM.

A.1. The Baum-Welch Algorithm. For a single individual $n \in \mathcal{V}$, estimating the TPM $P_n(t)$ based on a single observation sequence $j \in \{1, \cdots, B\}$ can be solved according to the standard Baum-Welch algorithm. Define $\hat{\eta}_{n,j}^{(t)}$ to be the estimate of the true parameter vector $\boldsymbol{\eta}_n$ at time t based on observation sequence j, and define a corresponding auxiliary function as:

$$Q_{n,j}(t) := \mathbb{E}\left[\log\left(p_{n,j}^{(c)}(\mathbf{X}_n(0{:}t), \mathbf{Y}_{n,j}(0{:}t)|\boldsymbol{\eta}_n(0{:}t))\right)\middle|\mathbf{y}_{n,j}(0{:}T), \hat{\eta}_{n,j}^{(0:t)}\right] \tag{A.1}$$

where $p_{n,j}^{(c)}$ denotes the joint probability distribution of observing a complete set of data $\{\mathbf{x}_n(0{:}t), \mathbf{y}_{n,j}(0{:}t)\}$ for individual n:

$$\begin{aligned} &p_{n,j}^{(c)}(\mathbf{x}_n(0{:}t), \mathbf{y}_{n,j}(0{:}t)|\boldsymbol{\eta}_n(0{:}t)) \\ &\qquad = q_n(x_n(0)) \prod_{s=0}^{t-1} P_n(s, x_n(s), x_n(s+1)) \prod_{s=0}^{t} O_n(y_{n,j}(s)|x_n(s)) \end{aligned}$$

We maximize the (A.1) to determine the optimal initial probability distribution $\hat{q}_{n,j}^{(t)}$ and the optimal TPM $\hat{P}_{n,j}^{(t)}$. Note that the maximization must be done subject to the regularity conditions $\sum_{u\in\mathcal{X}} \hat{P}_{n,j}^{(t)}(x,u) = 1$ and $\sum_{x\in\mathcal{X}} \hat{q}_{n,j}^{(t)}(x) = 1$ for all $x \in \mathcal{X}$. The

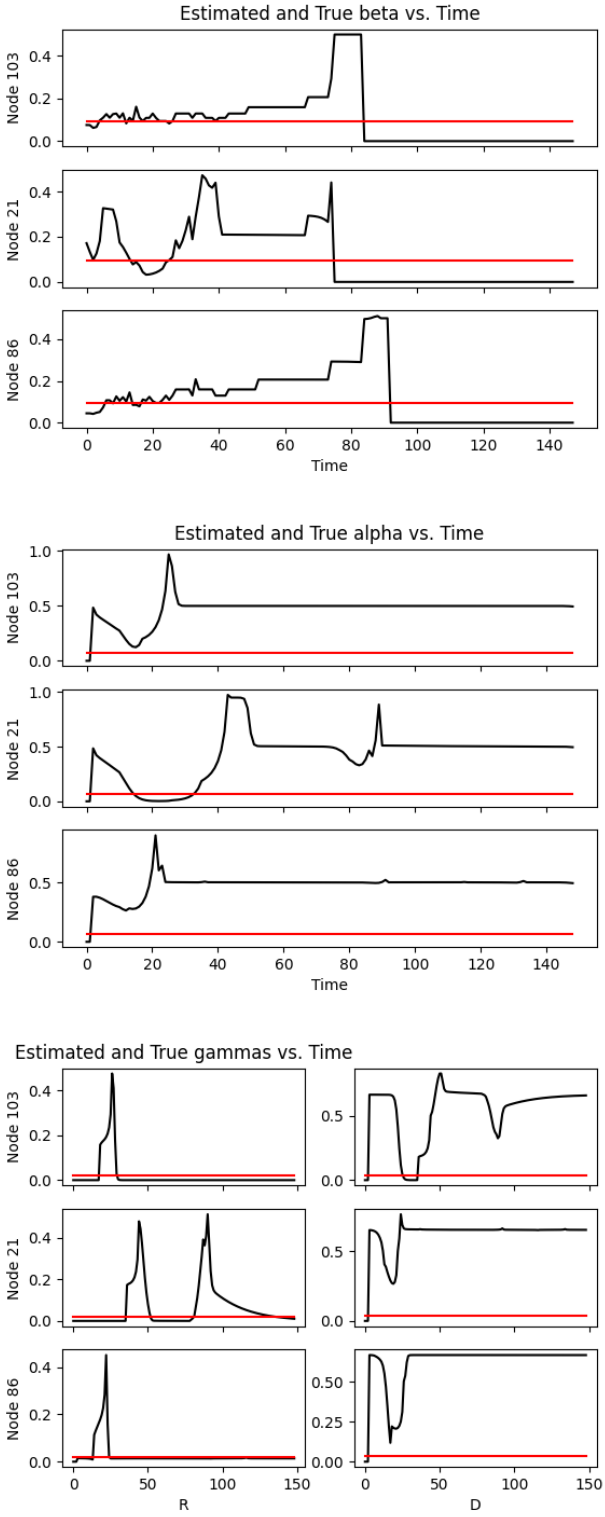

Fig. 8: The estimated parameters $\hat{\beta}_n^{(1)}(t)$, $\hat{\alpha}_n(t)$, $\hat{\gamma}_n^{(r)}(t)$ and $\hat{\gamma}_n^{(d)}(t)$ over time, with constant red line denoting the true value.

optimal point has the following closed-form expression:

$$\hat{q}_{n,j}^{(t)}(x) = g_{n,j}(0,x) \tag{A.2a}$$

$$\hat{P}_{n,j}^{(t)}(x,z) = \left(\sum_{s=0}^{t-1} h_{n,j}(s,x,z)\right)\left(\sum_{s=0}^{t} g_{n,j}(s,x)\right)^{-1} \tag{A.2b}$$

where the $g_{v,j}(t,x)$ and $h_{v,j}(t,x,z)$ are defined the same way as in (4.2). The procedure is repeated for each $t \in [0, T_{\text{sim}}]$ so that we obtain an estimate of $\hat{q}_{n,j}^{(t)}$ and $\hat{P}_{n,j}^{(t)}$ which evolves over time. In the following section, the forward-backward algorithm is discussed as a way to compute the $g_{v,j}(t,x)$ and $h_{v,j}(t,x,z)$.

A.2. The Forward-Backward Algorithm. 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:

$$f_{n,j}(t,x) = \sum_{z\in\mathcal{X}} f_{n,j}(t-1,z) O_n(y_{n,j}(t)|x) \hat{P}_{n,j}^{(t-1)}(z,x), \;\; f_{n,j}(0,x) := q_n(x) O_n(y_{n,j}(0)|x) \tag{A.3a}$$

$$b_{n,j}(t,x) = \sum_{z\in\mathcal{X}} b_{n,j}(t+1,z) \hat{P}_{n,j}^{(t)}(x,z) O_n(y_{n,j}(t+1)|z), \quad b_{n,j}(T,x) = 1 \;\forall x \in \mathcal{X} \tag{A.3b}$$

Now define $g_{n,j}(t,x)$ and $h_{n,j}(t,x,z)$ as in (4.2). The variables defined in (A.3) allow us to simplify the expressions beyond their definitions:

$$g_{n,j}(t,x) = \frac{f_{n,j}(t,x) b_{n,j}(t,x)}{\sum_{z\in\mathcal{X}} f_{n,j}(t,z) b_{n,j}(t,z)} \tag{A.4a}$$

$$h_{n,j}(t,x,z) = \frac{f_{n,j}(t,x) \hat{P}_{n,j}^{(t)}(x,z) O_n(\mathbf{y}_{n,j}(t+1)|z) b_{n,j}(t+1,z)}{\sum_{u,w\in\mathcal{X}} f_{n,j}(t,u) \hat{P}_{n,j}^{(t)}(u,w) O_n(\mathbf{y}_{n,j}(t+1)|w) b_{n,j}(t+1,w)}$$

Note that the expressions for (A.4) are dependent on previous estimates of the TPM $\hat{P}_{n,j}^{(0:t-1)}$ for each time t; essentially, we are recursively building new estimates of $\hat{P}_{n,j}^{(t)}$ based on the previous time's estimates.

Appendix B. Estimating the Hidden State Sequence. Recall $X_n(t) \in \mathcal{X}$ refers to the hidden state of individual $n \in \mathcal{V}$, and suppose we are given a time series of observations $Y_{n,j}(0:t)$ for symptom $j \in \{1, \cdots, B\}$. The estimated time-varying TPM underlying the HMM is given by $\hat{P}_n^{(t)}$ at time t, and the known OPM is given by O_n. The initial state is known and given by $X_n(0) = x_n(0)$. Then the standard *Viterbi algorithm* can be applied to the observation sequence $j \in \{1, \cdots, B\}$ to estimate the sequence $\hat{x}_{n,j}(1:t)$ of likely hidden states over time based on symptom j. The probability of observing some specific sequence of health statuses $x_n(1:t)$ for some $t \le T_{\text{sim}}$ is given by:

$$\mathbb{P}(\{X_n(t) = x_n(t), n \in \mathcal{V}, t \in [0,T]\})$$

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

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 recursively compute:

$$\delta_{n,j}(0,x) = q_n(x)O_n(y_{n,j}(0)|x)$$
$$\delta_{n,j}(t,x) = \max_{z\in\mathcal{X}} \delta_{n,j}(t-1,z)\hat{P}_n^{(t-1)}(z,x)O_n(y_{n,j}(t)|x),\ t\geq 1$$

Then for the specific observation sequence j, the optimal sequence of states is given by $\hat{x}_{n,j}(t) := \text{argmax}_{z\in\mathcal{X}}\delta_{n,j}(t,z)$.

REFERENCES

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

[3] *KOrean Statistical Information Service (KOSIS)*, https://kosis.kr/eng/statisticsList/statisticsListIndex.do?menuId=M_01_01&vwcd=MT_ETITLE&parmTabId=M_01_01#content-group.

[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] N. Bellomo, R. Bingham, M. A. J. Chaplain, G. Dosi, G. Forni, D. A. Knopoff, J. Lowengrub, R. Twarock, and M. E. Virgillito, *A multiscale model of virus pandemic: Heterogeneous interactive entities in a globally connected world*, Mathematical Models and Methods in Applied Sciences, 30 (2020), pp. 1591–1651.

[10] M. H. A. Biswas, L. T. Paiva, and M. De Pinho, *A SEIR model for control of infectious diseases with constraints*, Mathematical Biosciences & Engineering, 11 (2014), p. 761.

[11] M. Brand, N. Oliver, and A. Pentland, *Coupled Hidden Markov Models for complex action recognition*, 2012 IEEE Conference on Computer Vision and Pattern Recognition, (1997).

[12] E. Callaway, *Are COVID surges becoming more predictable? New Omicron variants offer a hint*, https://www.nature.com/articles/d41586-022-01240-x.

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

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

[15] 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).

[16] 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).

[17] 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.

[18] 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.
[19] 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.
[20] 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.
[21] G. D. Forney, *The Viterbi algorithm*, Proceedings of the IEEE, 61 (1973), pp. 268–278.
[22] J. Gallagher, *Covid reinfection: Man gets Covid twice and second hit 'more severe'*, https://www.bbc.com/news/health-54512034.
[23] Y. Gu, *COVID-19 projections using machine leaning*, https://covid19-projections.com/.
[24] 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.
[25] A. Gómez-Carballa, J. Pardo-Seco, X. Bello, F. Martinón-Torres, and A. Salas, *Superspreading in the emergence of covid-19 variants*, Trends in Genetics, 37 (2021), pp. 1069–1080.
[26] H. W. Hethcote and S. A. Levin, *Periodicity in epidemiological models*, Applied Mathematical Ecology, 18 (1989), pp. 193 – 211.
[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] P. van den Dreessche and J. Watmough, *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Mathematical Biosciences, 180 (2002), pp. 29–48.
[34] 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.
[35] R. Yu and Y. Ma, *DeepGLEAM*, https://jacobsschool.ucsd.edu/news/release?id=3129.