Epidemics 48 (2024) 100779

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Role of heterogeneity: National scale data-driven agent-based modeling for the US COVID-19 Scenario Modeling Hub

Jiangzhuo Chen [a,*], Parantapa Bhattacharya [a], Stefan Hoops [a], Dustin Machi [a], Abhijin Adiga [a], Henning Mortveit [a,b], Srinivasan Venkatramanan [a], Bryan Lewis [a], Madhav Marathe [a,c,*]

[a] Biocomplexity Institute, University of Virginia, Charlottesville, VA, USA
[b] Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA, USA
[c] Department of Computer Science, University of Virginia, Charlottesville, VA, USA

ARTICLE INFO

Keywords:
COVID-19 Scenario Modeling Hub
Agent-based model
Heterogeneity
Complexity
Digital twin

ABSTRACT

UVA-EpiHiper is a national scale agent-based model to support the US COVID-19 Scenario Modeling Hub (SMH). UVA-EpiHiper uses a detailed representation of the underlying social contact network along with data measured during the course of the pandemic to initialize and calibrate the model. In this paper, we study the role of heterogeneity on model complexity and resulting epidemic dynamics using UVA-EpiHiper. We discuss various sources of heterogeneity that we encounter in the use of UVA-EpiHiper to support modeling and analysis of epidemic dynamics under various scenarios. We also discuss how this affects model complexity and computational complexity of the corresponding simulations. Using round 13 of the SMH as an example, we discuss how UVA-EpiHiper was initialized and calibrated. We then discuss how the detailed output produced by UVA-EpiHiper can be analyzed to obtain interesting insights. We find that despite the complexity in the model, the software, and the computation incurred to an agent-based model in scenario modeling, it is capable of capturing various heterogeneities of real-world systems, especially those in networks and behaviors, and enables analyzing heterogeneities in epidemiological outcomes between different demographic, geographic, and social cohorts. In applying UVA-EpiHiper to round 13 scenario modeling, we find that disease outcomes are different between and within states, and between demographic groups, which can be attributed to heterogeneities in population demographics, network structures, and initial immunity.

1. Introduction

The world just witnessed the largest pandemic since 1918. The COVID-19 pandemic led to significant social, economic and health impacts worldwide. Computational models played an important role in supporting the policy makers during the pandemic. The US COVID-19 Scenario Modeling Hub (SMH) (Scenario Modeling Hub, 2023a) was formed in late 2020 to support policy makers. The consortium of modeling teams over the last 3 years considered 18 rounds of different what-if scenarios and created ensemble models that provided senior level policy makers analytical insights. The UVA-EpiHiper team was one of the 15 models that has participated in this community effort. It has contributed projections to the COVID-19 SMH in all rounds from 6 to 13 and most recently round 17. UVA-EpiHiper has also been adapted to participate in the Flu Scenario Modeling Hub (Scenario Modeling Hub, 2023b), the RSV Scenario Modeling Hub (Scenario Modeling Hub, 2023c), and the European COVID-19 Scenario Hub (Scenario Modeling Hub, 2023d). UVA-adaptive was the other team from University of Virginia (UVA) that also supported the SMH. The UVA teams after much deliberations decided to keep these two models active throughout the last 3 years, even though the computational and human costs were significant. Our decision was based on the fact that we wanted to provide results based on two different models; the key difference between them was the level of aggregation. A companion paper on UVA-adaptive (Porebski et al., 2024) describes the other effort. Here we focus on UVA-EpiHiper.

UVA-EpiHiper is a national-scale individual-level agent-based model among the models that participated in the COVID-19 SMH. This was one of its unique features. Its modeling capabilities allow it to incorporate heterogeneities in various surveillance data sets; to implement heterogeneities in the disease model, contact network, and behavior between individuals and among subpopulations; to produce projections that can be stratified to study outcome heterogeneities among geographical, social, or demographical groups. Such capabilities are made computationally efficient by workflows that can handle computational heterogeneities.

* Corresponding authors.
E-mail addresses: chenj@virginia.edu (J. Chen), marathe@virginia.edu (M. Marathe).

https://doi.org/10.1016/j.epidem.2024.100779
Received 15 August 2023; Received in revised form 20 May 2024; Accepted 17 June 2024
Available online 27 June 2024

It is crucial to be able to model the heterogeneities at all scales to better understand an epidemic and better evaluate interventions. This is challenging for the compartmental models or even the metapopulation models. An agent-based networked model is a natural solution. While such a model enables us to model heterogeneous data, policies, individual behavior, and individual disease progression, it poses challenges regarding computations, data needed for initialization, and analyses due to its complexity. This paper describes how the UVA-EpiHiper model handles the heterogeneities with its capabilities and how we address the heterogeneities in the computation and analytics.

Overview of UVA-EpiHiper. UVA-EpiHiper includes a high performance computing oriented pipeline designed for scalable epidemic analytics. The pipeline has five steps. *Step 1:* Build a digital twin of the social contact network. The digital twin is statistically similar to the real-world network but preserves the privacy and confidentiality of individuals. *Step 2:* Initialize the digital twin with surveillance data. *Step 3*: Use a high performance computing oriented simulation (EpiHiper) to calibrate and execute a statistical experiment design to study the specific decision-theoretic questions. *Step 4:* Create aggregate projections from the simulation outputs to be comparable with data from surveillance. *Step 5:* Analyze the simulation generated detailed data to obtain policy insights. Additional details about the pipeline can be found in Bhattacharya et al. (2021, 2023).

1.1. Summary of contributions

In this paper, we focus on three central questions: (*i*) How does UVA-EpiHiper capture various heterogeneities of real-world systems? (*ii*) How do these heterogeneities affect UVA-EpiHiper in terms of model complexity, software complexity, computational complexity, and analytical complexity? (*iii*) How do the heterogeneities impact the resulting epidemic dynamics in various counterfactual scenarios studied as a part of SMH? For the first question, we describe the various forms of heterogeneity that are represented within the UVA-EpiHiper framework. This includes: *network heterogeneity*, *disease model and intervention heterogeneity*, *initialization heterogeneity* in terms of diverse data streams, and *analytical heterogeneity* that stems from the need to analyze large detailed output data. For the second question, we describe how each of these heterogeneities impacts UVA-EpiHiper with respect to three important measures: (*a*) model complexity – the complexity of representing the underlying model, (*b*) software complexity, and (*c*) computational complexity – the computational resources used to complete the needed analysis. For the third question, as a concrete example, we discuss how UVA-EpiHiper was used to support SMH Round 13. Two natural broad questions arise with such detailed models: (*i*) when are such models needed and (*ii*) how does one validate such models. Both are important questions and will be discussed in Section 6.

1.2. Related work

Over the last three decades, agent-based models have become a popular modeling paradigm in epidemiology. Such models include e.g. EpiSimdemics (Barrett et al., 2008), EpiFast (Bisset et al., 2009), FRED (Grefenstette et al., 2013), Indemics (Bisset et al., 2014), EMOD (Bershteyn et al., 2018), and more recent models developed for COVID-19 modeling: Covid-Sim (Ferguson et al., 2020), Covasim (Kerr et al., 2021), OpenABM-Covid19 (Hinch et al., 2021), OpenCOVID (Shattock et al., 2022), and the model in Shoukat et al. (2020). A few agent-based models, including our UVA-EpiHiper have participated in the SMH. The NotreDame-FRED model (Moore et al., 2024) is based on FRED with modifications for COVID-19 and is mainly used for Indiana. The UF-ABM model (Pillai et al., 2023) is an agent-based model developed to study COVID-19 pandemic in Florida. The COVSIM model (Rosenstrom et al., 2024) is a stochastic agent-based COVID-19 simulation model for North Carolina. A few compartmental models have participated in the SMH too, e.g. Porebski et al. (2024), Bouchnita et al. (2024), Srivastava (2023). UVA-EpiHiper is the only agent-based model in SMH that models the whole of USA. These agent-based models usually have an explicit representation of the individuals and the underlying social contact network on which the disease spreads. Comparing with compartmental mass action models, such a representation can capture heterogeneity between individual agents and in the contact structure. The agent-based models allow us to directly capture behaviors and interventions, and to study targeted policies and response strategies. Our UVA-EpiHiper model aims to provide the following features which are crucial in the SMH work: *scalability* in terms of ability to simulate epidemics and interventions on national scale networks (with 100–300 million nodes), *capabilities and expressiveness* in terms of disease and intervention modeling, and *ease of specification* in terms of ability to specify disease models and interventions.

The use of agent-based models for scenario planning and projection also has a rich history. See the paper by Runge et al. (2023) for a more detailed account. Over the past two decades, we have used agent-based models to support scenario projections in epidemic science for various sponsors. Examples of our work include: (*i*) supporting Office of Homeland Security (OHS) and Joint Task Force Civil Support (JTF-CS) on Smallpox (Eubank et al., 2004), (*ii*) targeted layered containment (TLC) study done for the National Security Council (NSC) (Halloran et al., 2008), (*iii*) studies on H1N1 pandemic (Barrett et al., 2011a; Chen et al., 2010, 2018) (*iv*) studies on Ebola epidemic (Rivers et al., 2014; Venkatramanan et al., 2018) (*v*) tabletop exercise of pandemic planning done for the Defense Threat Reduction Agency (DTRA) and senior officials in the USG (Barrett et al., 2011b, 2015) (*vi*) studies done for the Department of Defense (DoD) to support MEDCOM and National Guard (Barrett et al., 2012). During the COVID-19 outbreak, we have continued to support DTRA and Virginia Department of Health (VDH) on various scenario planning exercises, using both metapopulation and agent-based models (Venkatramanan et al., 2019; Chen et al., 2019). The present paper describes our work done in the context of the ongoing Scenario Modeling Hub (SMH) effort (Scenario Modeling Hub, 2023a). This collaborative effort is novel in that it has brought together a *diverse group* of modelers, policy makers, and analysts to design and implement complex scenarios and has used the ensemble results to inform policy makers as they plan and respond during an *ongoing epidemic outbreak*. SMH has repeatedly shown that ensemble of the projections from a diverse set of models provides a more robust set of projections than a single model.

2. Heterogeneities in modeling capabilities of UVA-EpiHiper

UVA-EpiHiper is an individual-based model, in which each individual is explicitly modeled and represented. In this section we describe (*i*) how we model heterogeneities among individuals in terms of their demographic and socio-economic attributes; (*ii*) how we model heterogeneities in the social contact network based on individuals interacting with each other when they have activities at the same locations, instead of a particular structural graph model; (*iii*) how we model heterogeneous disease transmission and progression for different individuals; (*iv*) how we model heterogeneous behavior regarding mitigation measures, including compliance to pharmaceutical interventions (PIs) and non-pharmaceutical interventions (NPIs).

The UVA-EpiHiper model was designed to support a broad range of disease models and a large class of interventions needed to support policy makers in complex scenarios. The formal model is time stepped and captures the following elements of a contagion propagating over a network $G(V, E)$ of vertices V with an associated set $\mathcal{X} = \{X_1, X_2, \ldots, X_m\}$ of health states: (*i*) *transmission* of a contagion from infectious vertices to susceptible vertices, (*ii*) disease *progression* within each vertex that has become infected, and (*iii*) *interventions* which are formal procedures applied to the states of associated set of vertices or edges (intervention target) when certain predicate (trigger condition) is satisfied. We often refer to vertices as people, and although that does not need to be the case, it will be assumed in the following.

2.1. Models of disaggregated populations and social contact networks

The UVA-EpiHiper model uses a digital twin of the population of the study region. Such a digital twin captures the people with demographic attributes, their partition into households with household attributes, an activity sequence for each individual, and a set of residence and activity locations where people conduct their activities. The mapping of activities to locations allows one to infer a contact network which forms the basis for disease transmission in the UVA-EpiHiper model. The construction of the digital twin, which is illustrated in Fig. 1, is carried out so that the synthetic population and network closely resemble their real counterparts on dimensions relevant for epidemic scenarios.

In the constructed digital twin, individual demographic attributes include age, gender, race/ethnicity, employment status, etc. Household attributes include household size, income, and location (latitude/longitude). The construction of households and individuals is based on iterative proportional fitting (IPF) (Beckman et al., 1996) using Public Use Microdata Samples (PUMS) (U.S. Census, 2021b)) and US Census demographic distributions, and is conducted at the resolution of census block groups. A set of geographically embedded synthetic locations is constructed through detailed modeling and data fusion involving PostGIS and machine learning-based techniques, using the Microsoft Building Data (Microsoft, 2020), HERE (HERE, 2020) and BuildingFootprintUSA (BuildingFootprintUSA, 2020) point-of-interest (POI) data, National Center of Education Statistics (NCES) (NCES, 2021) data on school and college locations, land-use classification data (HERE, 2020), and urban/rural classifications (U.S. Census, 2021a).

Each individual is then assigned an *activity sequence* through Classification and Regression Trees (CART) and Finite Volume Method (FVM) (Lum et al., 2016; Breiman, 1984), using harmonized data from the National Household Travel Survey (NHTS) (U.S. Department of Transportation, Federal Highway Administration, 2020) and the American Time Use Survey (ATUS) (U.S. Department of Labor, Bureau of Labor Statistics, 2020). Each activity in an individual's activity sequence includes a type (e.g. home, work, school, college, shopping, religion, or other), a start time, duration in seconds, and a location. The location of each activity is assigned with a set of rules, using the American Community Survey (ACS) commute flow data (U.S. Census, 2020a) and the LEHD Origin–Destination Employment Statistics (LODES) (U.S. Census, 2020b). The location assignment of activities can be represented by the people-location network G_{PL} illustrated in the middle panel of Fig. 1.

From G_{PL} we derive a *contact network* G_P in which vertices are the people and edges are people–people contacts, as in the right panel of Fig. 1. We apply an extension of the Erdős–Rényi random graph model (Erdős and Rényi, 1959) to each location to connect people visiting the location simultaneously. The model is calibrated based on SocioPatterns and POLYMOD data (SocioPatterns, 2023; Cattuto et al., 2010; Mossong et al., 2008; Prem et al., 2017). The network G_P forms a baseline for disease transmission in UVA-EpiHiper, and can be changed by interventions. A more detailed overview of the construction methodology and their validation is provided in Mortveit et al. (2020).

Other significant efforts on constructing digital twin populations include (Tatem, 2017; Weber et al., 2021; Socioeconomic Data and Applications Center, 2020) on gridded populations, and Mistry et al. (2021), Wheaton et al. (2009), Gallagher et al. (2018) whose data structure and details are closer to that in our work. Some epidemic simulation models construct networks on the fly (Kerr et al., 2021; Shattock et al., 2022) or obtain scaling (in terms of population size) by allowing each agent to represent a given number k of actual individuals. A unique aspect of our digital twin is the level of detail included and the diverse data sets used to synthesize the twin. As a result, a population and the resulting network constructed using our approach has extensive heterogeneity in all aspects including the individuals and their attributes, in their household structures, in their activity patterns, and in the locations they visit and the contacts they form at these locations. The heterogeneity is reflected spatially, temporally and socially (e.g. mixing patterns).

The heterogeneity is manifested in the structural and dynamical properties of the resulting social contact network G_P. For example, the ratio of the number of people to the number of activity locations will clearly influence the number of simultaneous visits and thus the potential for interactions (edges). Similarly, the number of activities and their duration will shape densities and properties (labels) of edges: more activities will typically lead to more interactions, albeit of shorter duration. The network G_P may thus get a larger average degree $\bar{d}$ as people have more activities, but durations of contact will diminish. Flow data such as the American Community Survey (ACS) commute data (U.S. Census, 2020a) will influence the spatial embedding of G_P: as the average commute distance increases, one would expect the network to have "longer" edges (connecting people residing farther away from each other) which in turn could make an epidemic outbreak spread faster throughout a region and thus be harder to contain.

To illustrate the resulting heterogeneity, we computed a few structural measures for the contact networks G_P generated across the set of US states. A more detailed account will be discussed in a separate manuscript that focuses on the construction of the digital twin. In our analysis, we found that some measures are sensitive to details whereas others are not. For example, the average degree $\bar{d}$ varies across the range $27.59 \leq \bar{d} \leq 47.43$ for the US states while the relative size r of a giant component is quite stable satisfying $0.97 \leq r \leq 1.00$. A giant component of a network is a connected component that contains a significant fraction of all the nodes and the fraction is called its relative size. Focusing on the populations and networks for the states of Massachusetts (MA) and Michigan (MI), their average degrees are $\bar{d}_{\mathrm{MA}} = 30.96$ and $\bar{d}_{\mathrm{MI}} = 39.57$. There is, however, virtually no difference between MA and MI in the average contact duration (total hours per person): $\bar{T}_{\mathrm{MA}} = 102$ and $\bar{T}_{\mathrm{MI}} = 103$. Their degree distributions and core number distributions are shown in Fig. 2, further illustrating differences in structural characteristics. While such differences may not impact certain kinds of dynamics over the population networks, it can still offer valuable insight when determining efficient interventions. This structural insight obtained through the constructive model for the populations and arising as emergent properties thereof, could be challenging to capture in agent-based models that do not consider this level of detail in their design approach.

2.2. Models of within-host disease progression and between-host disease transmission

The disease model in UVA-EpiHiper consists of within-host disease progression and between-host disease transmission. The former refers to an individual transitioning from one health state to another health state independent of other individuals. The latter refers to the disease being transmitted from an infectious individual A to a susceptible individual B, causing B to transition to an infected state. The *disease progression* is represented using a probabilistic timed transition system (PTTS) (Bisset et al., 2014; Barrett et al., 2008) over the set of health states $\mathcal{X}$. PTTS extend the classical finite state machine by allowing one to represent probabilistic and timed state transitions as specified by per-edge dwell time distributions. For *disease transmission*, consider a susceptible person P in health states X_k who is in contact with an infectious persons P' in state X_i. We combine the *state infectivity* $\iota(X_i)$ and *state susceptibility* $\sigma(X_k)$ of their health states with the *infectivity scaling factor* $\beta_\iota(P')$ of P' and the *susceptibility scaling factor* $\beta_\sigma(P)$ of P to form the *propensity* ρ associated with the *contact configuration* $T_{i,j,k} = T(X_i, X_j, X_k)$ for the potential transition of the health state of person P to X_j as:

$$\rho(P, P', T_{i,j,k}, e) = \left[T \cdot \tau\right] \times w_e \times \alpha_e \times \left[\beta_\sigma(P) \cdot \sigma(X_i)\right] \times \left[\beta_\iota(P') \cdot \iota(X_k)\right] \times \omega(T_{i,j,k}) \quad (1)$$

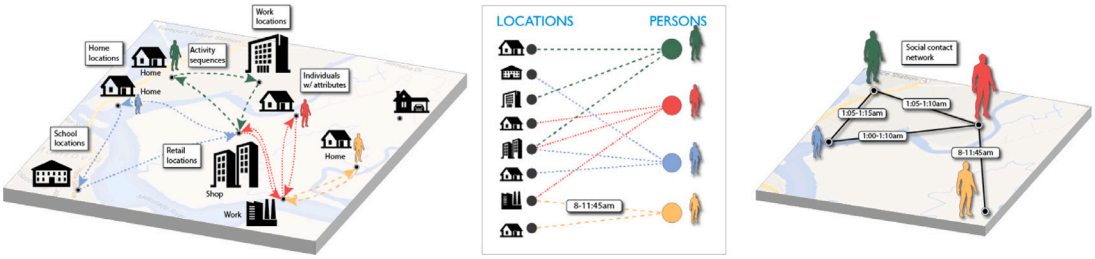


Fig. 1. An illustration of the population and network components used in UVA-EpiHiper. On the left, synthetic people with demographic attributes and household structure are illustrated along with their assignment to residence locations. Each person is assigned an activity sequence consisting of activities such as work and shopping. These activities are mapped to appropriate activity locations (e.g., a government worker goes to work at a location with a government classification) as illustrated by the dashed paths. This complete assignment of activities to locations is formally represented as the people-location network G_{PL} illustrated in the middle panel, which in turn gives rise to a social contact network G_P as shown on the right. The latter captures person–person contacts, including their duration and location.

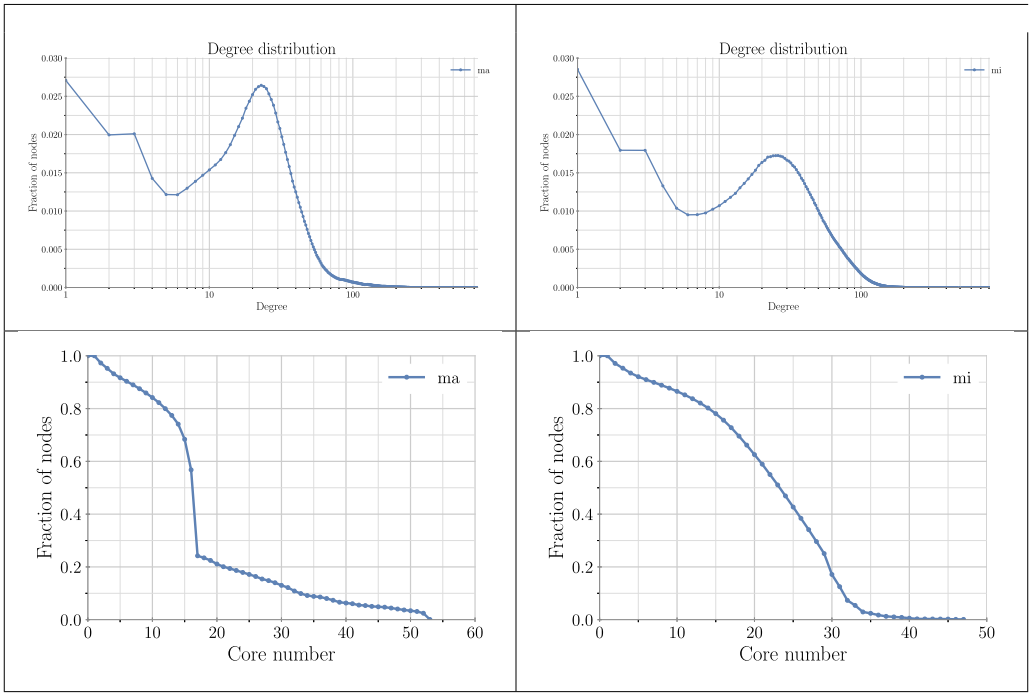


Fig. 2. Heterogeneities between states in terms of structural properties of their contact networks G_P. The top row shows (undirected) degree distributions and the bottom row shows the core number distributions for Massachusetts (left) and Michigan (right).

In Eq. (1), T is the contact duration of the edge $e = (P', P, w_e, \alpha_e, T)$, w_e is the edge weight, and α_e is an indicator variable encoding whether or not the edge is active (e.g., disabled because of an ongoing school closure). Finally, $\omega(T_{i,j,k})$ is the *transmission weight* of the contact configuration $T_{i,j,k}$, and τ is the *transmissibility*. Propensities are determined at each time step and for each person P using Eq. (1) for all edges e and contact configurations $T_{i,j,k}$, generating a sequence ρ_P. A Gillespie process (Gillespie, 1976, 1977) is used to determine if P becomes infected. Also, the person P', to whom one attributes P becoming infected, is chosen randomly with probabilities weighted by the corresponding propensities.

In Fig. 3, we show a simplified version of COVID-19 disease model implemented by UVA-EpiHiper in the SMH work. It illustrates the PTTS and the associated probability p and dwell time d of each state transition. Each state in the diagram is also associated with susceptibility σ and infectivity ι, which can be regulated by different susceptibility factor β_σ and infectivity factor β_ι for different individuals. In UVA-EpiHiper, heterogeneities resulting from various diseases and its manifestation can be represented by (*i*) a different PTTS structure for every individual; (*ii*) the same PTTS structure across all individuals, but with different parameterizations. The former can arise between different types of hosts, e.g., human and vector. The latter is more common, where theoretically we can assign to each individual unique values for the parameters (p_i, d_i) depending on the individual's attributes such as their age, their immune profile and medical history. In actual studies, we often partition the population by a set of variables, e.g. by age into age groups. People in the same group have the same disease model parameterizations. This substantially reduces the complexity of representing and implementing the disease model in UVA-EpiHiper, and improves the computational efficiency of the resulting simulations.

The disease models listed in Table 1 are variations of the COVID-19 disease model, which is an extension of the classical SEIR model, with extra features added over the rounds of scenario modeling to handle various scenarios. Their complexity can be represented by the number of states and the number of state transitions.

Our scenario modeling work began long before SMH. In January 2020, we started with COVID v1 model to study asymptomatic ratio and various outcomes including hospitalization, ventilation, and death. In March 2020, we augmented our disease model to COVID v2 to implement age stratification. We partition the population into five age groups: p (preschooler, 0–4), s (school age, 5–17), a (adult, 18–49), o (older adult, 50–64), and g (golden age, 65+), and assign

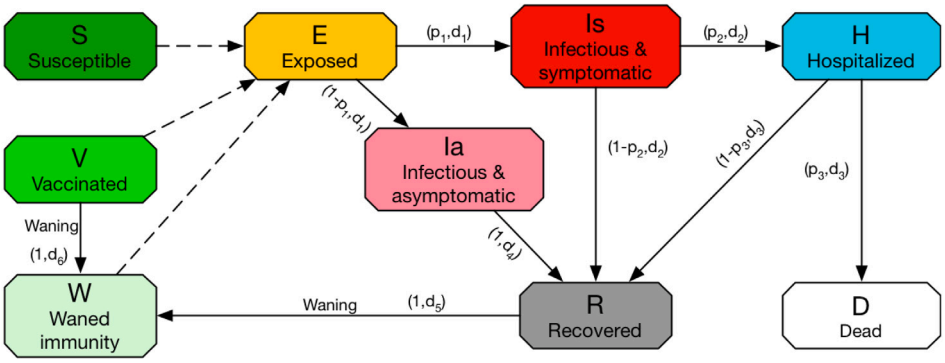


Fig. 3. An illustration of disease models that can be potentially implemented in UVA-EpiHiper. The subgraph connected by solid arrows forms a PTTS for within-host progression. Each arrow is associated with a state transition probability p and a dwell time d which may be a random variable. The dashed arrows are associated with disease transmission, for which other nodes in the network are involved.

Table 1
Variations of COVID-19 disease model implemented by UVA-EpiHiper with different features and complexity.

Disease	Features	Implementation	Complexity		
			States	*Transmissions*	*Progressions*
COVID v1	asymptomatic state; severe outcomes	add states to a basic SEIR model	13	6	16
COVID v2	v1 + age stratification	states/transitions for each age group; transmissions across age groups	90	225	100
COVID v3	v2 + vaccines	vaccinated states with different transitions	105	300	120
COVID v4	v3 + multivariant	variant-specific infectious states	140	600	185
COVID v5	v4 + immune waning/escape	transition from R to W; transmission across variants	170	975	250

different probabilities of transition to H (hospitalized) and D (dead) states, to model e.g. higher likelihood of severe outcomes in senior people if they are infected with COVID-19. At the end of 2020, when vaccines began to be administered, we expanded our model to COVID v3 with vaccinated states. Along the rounds of SMH, we extended the disease model for various doses of vaccines, initially V_1 for being vaccinated with one mRNA dose (either Pfizer or Moderna), V_2 for being vaccinated with two doses, and V_{jj} for being vaccinated with Johnson&Johnson vaccine (from round 6 to round 9). When boosters started to be administered, we added V_{b1} for the first booster (from round 10 to round 13), then V_{b2} for the second booster (round 17). These different vaccinated states differ by associated susceptibility, to represent different efficacies of corresponding vaccines/doses. From SMH round 6, to model multiple variants, we upgraded our disease model to COVID v4 by creating infectious states for each variant and expanded disease transmissions between each combination of infectious state and susceptible state. From SMH round 11, we updated our disease model to COVID v5, in which we implemented asymmetric immune escape, where a node with immunity to an older variant (e.g. Delta), obtained either by infection or by vaccination, can be infected by a newer variant (e.g. Omicron), but not the opposite. In UVA-EpiHiper, immunity waning is implemented as a discrete process by state transition from a state with obtained immunity (natural or vaccinal) to a state with waned immunity.

2.3. Models of pharmaceutical and non-pharmaceutical interventions

UVA-EpiHiper interventions describe both pharmaceutical (PIs) and non-pharmaceutical interventions (NPIs). An UVA-EpiHiper intervention consists of a *trigger condition* C, an *intervention target* T, and a collection of operations that are applied against the variables associated with the elements of the target, or against variables not attached to target entities (through the **once** construct).

The trigger condition C is a Boolean expression formed using time, sizes of sets, and values of variables. The trigger sets as well as the target sets consist of vertices or edges and can be formed using UVA-EpiHiper internal attributes of individuals (nodes) or contacts (edges). These attributes can be augmented through an external trait database defining properties of individuals or contact locations. The target set may be sampled and different operations can be specified for the sampled and non-sampled subsets.

Operations are *ordered*, first by execution time and second by priority. Sub-sequences of operations of the same priority are processed in random order. The operations are organized into the control blocks specified in Listing 1. More details about the semantics of each block can be found in Appendix A.2. In each operation, a variable is assigned the value of an expression, but the assignment is scheduled for execution after a *delay* $d \geq 0$ relative to the current time step. One may additionally assign it an integer *priority* (default value 0) and a *condition* which is a Boolean expression that must hold at execution time to avoid the operation being canceled.

Intervention complexity. As a result of this formal set-based structure UVA-EpiHiper is able to implement a rich class of interventions without any coding. That is, modelers have full control over the interventions without the help of programmers. Table 2 describes some common interventions implemented in UVA-EpiHiper for various studies. Table 4 details their representational complexity. The column "Traits" refers to the usage of custom, time-varying attributes of nodes, whereas the column "Demographics" gives the number of fields accessed in the UVA-EpiHiper person trait database. The demographic information in these experiments is only used during initialization. Therefore it has limited influence on the running time.

We investigate impacts of the disease model and intervention complexity with a performance study on the synthetic population of Virginia using COVID v2 model in Table 1. The study included eight

Table 2
Interventions and their descriptions.

Intervention	Description
Voluntary home isolation	Individuals who notice symptoms comply voluntarily (sampling with a compliance rate) with the recommendation to stay at home. All non-home edges of the compliant individuals are deactivated for 15 days.
School closure	All students who go to school or college stay at home. This is implemented by deactivating edges for which either source or target activity is either *school* or *college*. Note that teachers or other school employees are still in contact with each other.
Stay at home orders	These include date dependent stay at home orders (SH), reversal of such orders (RO), as well as an automated policy where the order is triggered by more than 2000 hospitalizations and reversed once the number drops below 2000. Prior to the simulation the individuals who will comply with these orders are selected (sampling with compliance rate) to have the trait `complyWhenOrdered`. Non-home edges of either the target or the source node who is complying are deactivated and activated either at predetermined dates (SH, OR) or according to the prescribed threshold (PS).
Test and isolate asymptotic cases	Individuals, up to a number determined by testing capacity, who do not show symptoms are tested. Voluntary home isolation is applied to those tested with positive results.
Contact tracing distance 1	Close contacts reported by confirmed cases are recommended to quarantine themselves at home. If they comply (sampling with a compliance rate), their non-home edges are deactivated for 15 days.
Contact tracing distance 2	Close contacts reported by confirmed cases and the close contacts of these contacts are recommended to quarantine themselves at home. If they comply (sampling with a compliance rate), their non-home edges are deactivated for 15 days.

Table 3
Differences in the HPC clusters used for executing EpiHiper simulation workflows.

	Rivanna	**Bridges 2**	**Anvil**
# Total nodes	115	488	1000
# CPU cores per node	40	128	128
RAM per node (GB)	384	256	256
CPU make	Intel Xeon Gold 6148	AMD EPYC 7742	AMD EPYC 7763
Network adapters	Mellanox ConnectX-5	Mellanox ConnectX-6	Mellanox ConnectX-6
OS	CentOS Linux 7	CentOS Linux 8	Rocky Linux 8.7
Filesystem	GPFS	Lustre	GPFS

computational experiments (labeled I through VIII in Table 5) with increasing complexity. For reference, each experiment was conducted on compute nodes with dual CPUs having 20 cores each and 375 GB total memory with its specified collection of interventions for 15 replicates.

Fig. 8 shows the clear impact of intervention complexity on running time. For each experiment we show the run time contribution from each of the main simulation tasks (intervention, transmission, update, synchronization, output, and initialization). The height of the ■-colored bar segment shows the time spent executing the interventions across experiments I through VIII. We find that contact tracing interventions (CTD1, CTD2) have higher computational complexity than other interventions, as indicated by Table 4.

3. Heterogeneities in the input data

UVA-EpiHiper model is data-driven, i.e., it takes disease surveillance data and vaccination data to initialize the state of the system including the initial health state of each individual, and to schedule interventions for each individual. In this section, we describe (*i*) how we take age stratified, county level daily confirmed case data to assign individuals to different initial health states and immunity classes, including naively susceptible with no immunity, partially susceptible with waned immunity, and non-susceptible with full immunity; (*ii*) how we take state level daily vaccine administration data stratified by age and dose to assign individuals to various vaccinated states at the beginning of the simulation.

3.1. Initializations of UVA-EpiHiper

The heterogeneities in UVA-EpiHiper not only originate from its models for disease spread and social contact networks, but also come from the data used to initialize the models and to drive the simulations. In this section, we describe how UVA-EpiHiper takes surveillance data sets of various resolutions and uses them to initialize the health state of the population at individual level. We argue that UVA-EpiHiper is able to leverage the details available in these data sets and that the heterogeneities in the input are reflected in the output, which in scenario modeling is the projections of epidemic outcomes.

In UVA-EpiHiper, we do not always model the whole history of a pandemic from its very beginning. Specifically, in scenario modeling of the COVID-19 pandemic which started in the U.S. from early 2020, we initialize the system from time t_0, where $t_0 + \delta$ is the beginning of the scenarios. We choose δ such that UVA-EpiHiper has sufficient "ramp-up" time to catch up with the most recent development of the epidemic. Based on experimenting, δ is about 1–2 months. For example, for scenario modeling in March 2022, we initialize our model from February 2022. We use the history before t_0 to derive the distribution of immunity level in the population and initialize each individual's health state at t_0. This way, we do not need to calibrate and compute the whole trajectory of the system from $t = 0$ to $t = t_0$. We note that this approach is common in e.g. Influenza modeling, where it is impossible to go back to the first cases.

In the COVID-19 model of UVA-EpiHiper, there are four immunity classes: full susceptibility (S), natural immunity (R), vaccinal immunity (V), partial immunity (W). The nodes with vaccinal immunity are partially protected from infection and severe outcomes. The nodes with natural immunity are fully protected from infection (in the absence of immune escape). The nodes with partial immunity are partially protected from infection. We initialize each node to one of the health states corresponding to these immunity classes in the following way. Data from The New York Times (The New York Times, 2023), based on reports from state and local health agencies, provides daily cumulative number of confirmed cases for each county c: $X_{c,t}$ from Jan. 21st, 2020. From CDC website, we have the cumulative number of cases in each age group at national level, from which we compute a distribution D of cases among age groups. We apply D to $X_{c,t}$ to obtain age stratified cumulative confirmed cases $X_{c,a,t}$ for each county c, each age group a, and each day t. With a per-age group case ascertainment rate α_a we

Table 4
UVA-EpiHiper interventions and factors influencing their representational complexity.

Intervention	Details	Node sets	Edge sets	Set operations	Traits	Demographics
VHI	Voluntary home isolation	1	2	2	1	2
SC	School closure	0	1	0	1	2
SH, RO, PS	order, reverse, alternate stay at home order	1	1	1	1	2
TA	Test and isolation of asymptomatic cases	1	5	3	2	2
CTD1	Contact tracing distance 1	4	5	5	3	2
CTD2	Contact tracing distance 2	7	7	8	3	2

Table 5
The list of interventions used in experiments I through VIII.

Experiment	Interventions
I	VHI, SC, SH
II	VHI, SC, SH, RO, TA
III	VHI, SC, SH, TA
IV	VHI, SC, SH, RO, PS
V	VHI, SC, SH, RO
VI	VHI, SC, SH, RO, CTD1
VII	VHI, SC, SH, RO, CTD1, PS
VIII	VHI, SC, SH, RO, CTD2

scale $X_{c,a,t}$ to estimate cumulative number of infections $Y_{c,a,t} = X_{c,a,t}/\alpha_a$. These infections include those reported and those not reported. Now we need to compute among these infections, how many still have full natural immunity (in R state) and how many have waned immunity (in W state) at time t_0. To this end, we consider that nodes can be infected multiple times ($S \to R \to W \to R \to \dots$). So we estimate the probability of future reinfection for each infection in $Y_{c,a,t}$ and consider only the last infections $Y'_{c,a,t}$. For each such infection in $Y'_{c,a,t}$, we apply the waning process, which is based on an exponential distribution of time to wane, to compute the probability that the natural immunity on the individual has waned by time t_0. This way we get an estimate of total number of nodes W_{c,a,t_0} among all $Y'_{c,a,t}$ that we will set to W state at time t_0 and the remaining number $R_{c,a,t_0} = Y'_{c,a,t_0} - W_{c,a,t_0}$ of nodes that we will set to R state at t_0, for each county each age group.

From CDC COVID Data Tracker (Centers for Disease Control and Prevention, 2023), we obtain state level weekly cumulative vaccine administration data by age group by dose: $Z_{s,v,a,t}$ as number of people of age group a in state s vaccinated with dose v by time t, where $t \le t_0$. For each vaccinated individual, we consider the last dose received, and apply the waning process, which is based on an exponential distribution of time to wane, to compute the probability that the vaccinal immunity on the individual has waned by time t_0. In the end, we get an estimate of total number of nodes W'_{s,v,a,t_0} among all $Z_{s,v,a,t}$ that we will set to W state at time t_0 and the remaining number $V_{s,v,a,t_0} = Z_{s,v,a,t_0} - W'_{s,v,a,t_0}$ of nodes that we will set to V state at time t_0, for each state each dose each age group.

We initialize the individuals in a state population as follows. All individuals are set to S state by default. For each age group a of each county c, we randomly choose W_{c,a,t_0} nodes and set them to W state; in the remaining people, we randomly choose R_{c,a,t_0} nodes and set them to R state. Then for each age group a of the state population, we randomly choose $W'(s, v, a, t_0)$ nodes for each dose v and set them to W state; in the remaining people, we randomly choose V_{s,v,a,t_0} nodes and set them to V_v state.

Seeding and calibration. We take the county level confirmed cases $X_{c,t}$ of $t \in (t_0, t_0 + \delta)$ and split the data into two series: $(t_0, t_0 + \delta_1]$ and $(t_0 + \delta_1, t_0 + \delta)$. We apply age stratification and scaling to the former to get daily number of new infections in each age group of each county. We use this to seed the simulation by randomly select nodes according to this time series and set them to E state. The time series $X_{c,t}$, $t \in (t_0 + \delta_1, t_0 + \delta)$ is scaled and aggregated to state level daily number of new infections. It is taken as the target of calibration: we calibrate transmissibility in the disease model to fit the simulated daily number of new infections to this target.

The seeding period (about 2–4 weeks) is chosen so that in the simulation the initial disease spread is stable and consistent with the recent data. The calibration period (about 2–4 weeks) is chosen so that we have enough data points in the target. We do not explicitly model the trajectory before t_0; instead we aggregate the infection and vaccination history to get a snapshot of the immunity distribution in the population at the beginning of our simulation. The parameter that we calibrate combines the intrinsic transmissibility of the disease and factors that affect the population universally. It changed over the course of the pandemic as different variants emerged and social distancing level changed. Our calibration is to identify its most recent value that can be assumed to remain constant (except being modulated by seasonality) for the projection period.

Note that through the modeling capabilities of UVA-EpiHiper, the leverage of county level data or age distribution of cases is generalizable. That is, if data is available at a different resolution or with a different stratification, e.g. if cases by race/ethnicity are reported in some state, then it is straightforward to use such data to initialize UVA-EpiHiper for this particular population, since our synthetic population has race/ethnicity attributes for each individual.

4. Heterogeneities in computation

One of our major goals is to aid policy makers with *real-time*, actionable insights for their decision-making processes. This presents several challenges due to the short-lived relevance of key questions and tight deadlines, typically between three days and two weeks, as observed during the COVID-19 pandemic. Within this time frame, we must design and run simulations, gather and calibrate data, produce statistically significant results, and analyze these results for presentation to policy makers. The process may need repetition if errors occur at any stage.

UVA-EpiHiper uses agent-based fine-grained simulations, the benefits and importance of which have been discussed earlier. These simulations require significant compute resources, much beyond what is available at our university's local cluster, necessitating the use of external, large-scale high performance computing (HPC) clusters. However, the demand for HPC systems is much greater than supply, which leads to significant waiting time for compute jobs submitted to them. Additionally, these systems, being complex, require regular maintenance (planned or otherwise) which can often disrupt our schedules. To mitigate these risks and ensure timely completion of jobs, we typically utilize 2–3 HPC clusters. This approach speeds up the computationally intensive parts of our workflow and allows us to deliver results on time even if one of the clusters is unavailable due to maintenance.

Using multiple HPC clusters, however, introduces a number of technical challenges stemming from the compute heterogeneities.

Table 3 shows the overall differences in the three compute clusters that we have used to generate projections for SMH. The following list summarizes some of the compute heterogeneities and complexities arising from them.

Fig. 4. Projected cumulative number of infections (per 100 K population size) over time after March 13, 2022, across different US states in different scenarios. Each curve is the average over 50 simulation replicates. Between states, the curves show various differences in magnitude (normalized by population size), trends, and impact of waning and new variant.

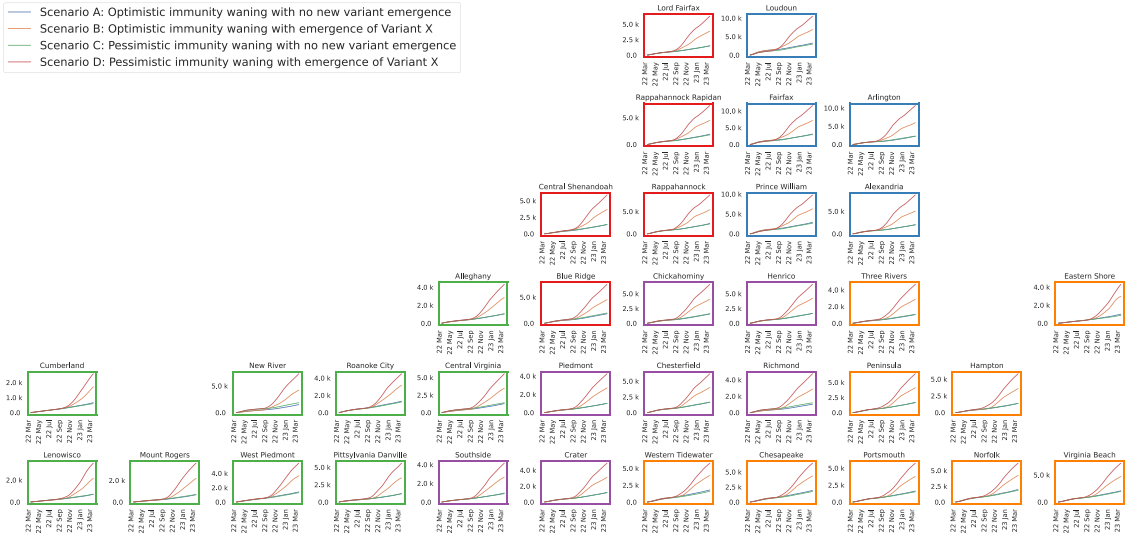
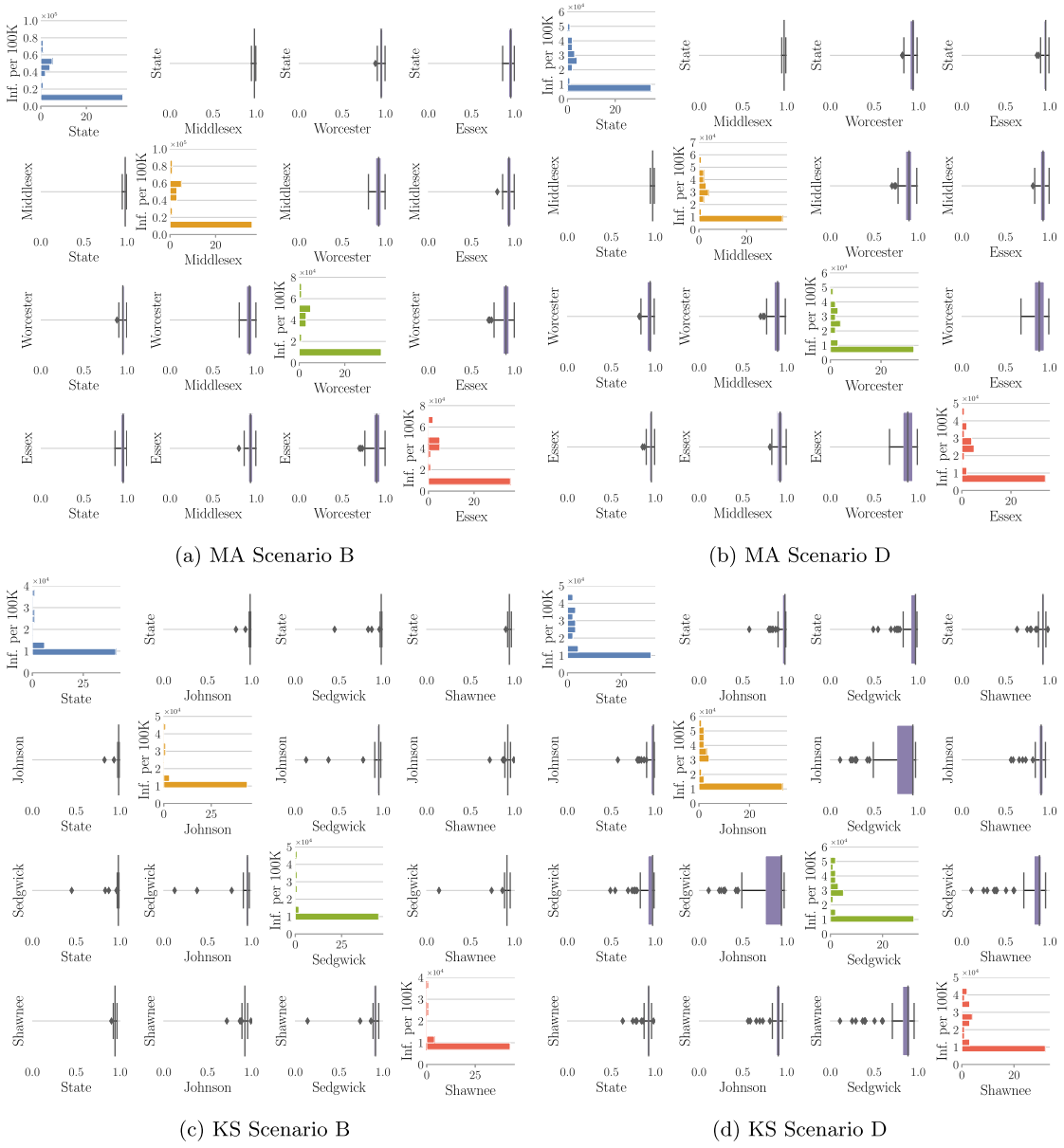


Fig. 5. Projected cumulative number of infections (per 100 K population size) over time after March 13, 2022, across different health districts in the state of Virginia, in different scenarios. Each curve is the average over 50 simulation replicates. Between health districts, the curves show similar trends and impact of waning and new variant, but different magnitude (even after normalization by population size).

- **Difference in compute node configurations** requires separate optimization for different compute configurations, otherwise causes underuse of compute on systems with low memory/CPU ratios.
- **Difference in filesystem availability and limits** requires customization of workflows to account for the different setups, and hinders debugging on systems with more expensive filesystems.
- **Compute reservations vs. custom QoS** (quality of service): This difference requires extensive planning on systems that only support reservations to make large amounts of compute available, otherwise we risk resource waste in case things do not go as planned.
- **Difference in access control requirements** makes development of automate workflows for multi-cluster systems difficult.

Based on our application requirements we concluded that a multi-cluster workflow pipeline should ideally be able to satisfy the following requirements:

- Be able to efficiently execute multi-node distributed memory MPI (Message Passing Interface) applications. In particular be able to leverage existing HPC schedulers (such as Slurm), and use Process Management Interface (PMI) for quick startup of multi-node MPI applications.



(a) MA Scenario B (b) MA Scenario D

(c) KS Scenario B (d) KS Scenario D

Fig. 6. Spatial heterogeneity — correlation plots: Comparing the overall spread in a state with the spread in top three counties by population. Two scenarios (B and D) and two states (MA and KS) are considered for comparison. In each case, 50 replicates are considered. The plots are ordered as follows: state followed by top three counties ordered by decreasing population. The diagonal plots show the histogram of infections per 100 K individuals over the simulation. The off-diagonal plots show the distribution of Pearson's correlation coefficient of the infection time series (epicurves) corresponding to each pair of regions. Plots for scenarios A and C are in the appendix.

- Be able to run on modern secure clusters where login and services like ssh are secured using single sign-on with central authentication, two-factor authentication, networks isolated with VPNs, etc.
- Be able to support complex task dependencies via task-dependency graphs. Additionally, they must support dynamic on-the-fly task creation, which is important for settings such as calibration.
- Be able to do task semantic-aware fault detection and recovery.
- Be able to support disparate HPC site-specific configurations.
- Provide a simple centralized interface to submit tasks that can be run on available HPC clusters.

Wormulon. To be able to answer policy driven questions related to epidemiological modeling in real-time we developed a custom workflow pipeline called Wormulon (Bhattacharya et al., 2023). Wormulon's design was driven by practical considerations and was guided heavily by the mult-cluster system development issues described above. While many systems existed that addressed some of the above issues — such as: HPC cluster schedulers like Slurm (Yoo et al., 2003) and PBS (Feng et al., 2007), pilot-based systems such as Radical Pilot (Merzky et al., 2021), multi-cluster schedulers like Argo and Balsam (Childers et al., 2017; Salim et al., 2019) and Leiden Grid Infrastructure (Somers, 2019), and modern big data and machine learning-oriented schedulers like Mesos (Hindman et al., 2011), Yarn (Vavilapalli et al., 2013), Dask (Rocklin, 2015) and Ray (Moritz et al., 2018), — none of these systems satisfied all of the requirements as stated above. Wormulon was developed to satisfy our task-specific needs. This workflow pipeline helps UVA-EpiHiper to handle computational heterogeneities and provide real-time epidemiological modeling capability.

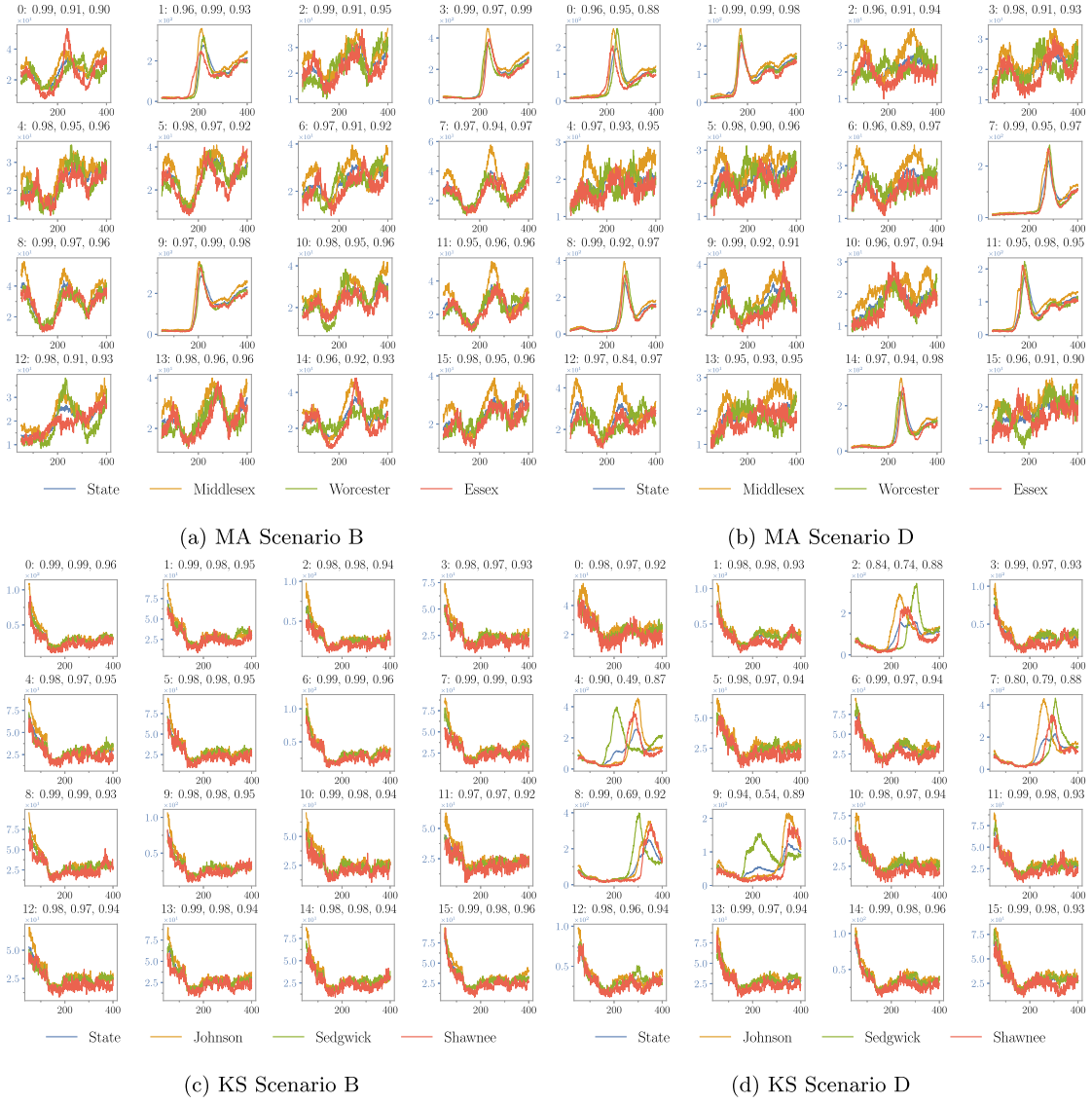


Fig. 7. Spatial heterogeneity — epicurves: Continuing from Fig. 6, we compare disease spread over time in a state with that in top three counties by population size. Plots are for a subset of cascades. The x-axis of each plot shows days from simulation starting date (2022-02-13); the y-axis shows new infection counts per 100 K individuals. The title of each plot contains the cascade number followed by correlation coefficients between the state infection counts and the county infection counts in the descending order of populations. Plots for scenarios A and C are in the appendix.

5. Example: Scenario modeling hub round 13

Round 13 of the Scenario Modeling Hub (SMH) focused on modeling the epidemiological implications of waning immunity from previous infections and vaccinations, and emergence of a new significant SARS-CoV-2 variant. It was performed in March 2022 and aimed to project epidemic outcomes from mid-March 2022 to mid-March 2023.

Overall Round 13 models four scenarios: (*i*) **Scenario A**: optimistic immunity waning with no new variant emergence; (*ii*) **Scenario B**: optimistic immunity waning with emergence of variant X; (*iii*) **Scenario C**: pessimistic immunity waning with no new variant emergence; and (*iv*) **Scenario D**: pessimistic immunity waning with emergence of variant X. Immunity includes protections gained from infection and vaccination. Waning is modeled as a transition from an immune state to a *partially* immune state. In the optimistic (or pessimistic) scenarios, the transition takes place after a median time of 10 (or 4) months with 40% (or 60%) reduction in protection compared to the baseline immunity. Variant X was assumed to have the same transmissibility and severity as existing variants. But an individual with immunity against existing variants has a 30% larger risk to be infected by X. A detailed description of the scenarios can be found at Scenario Modeling Hub (2022). Calibration results are shown in Fig. 9.

The spread pattern of a contagion is a result of the disease characteristics (such as virulence and variants) and the dynamics of the population affected by it (immunity levels, nature of interactions, epidemic response, etc.). To better understand the complex phase space of the agent-based model, we conducted several fine-grained analyses at different scales (node- and node-subset-level) across different Round 13 scenarios and different networks (state- and county-level). To this end, we apply the graphical viewpoint of simulation outputs; each instance of the simulation output can be viewed as *cascade graph ensembles*, where each *cascade* (Newman, 2003) is a highly structured who-infected-who graph with rich structural (network related) and dynamical (disease and disease response related) attributes. We study

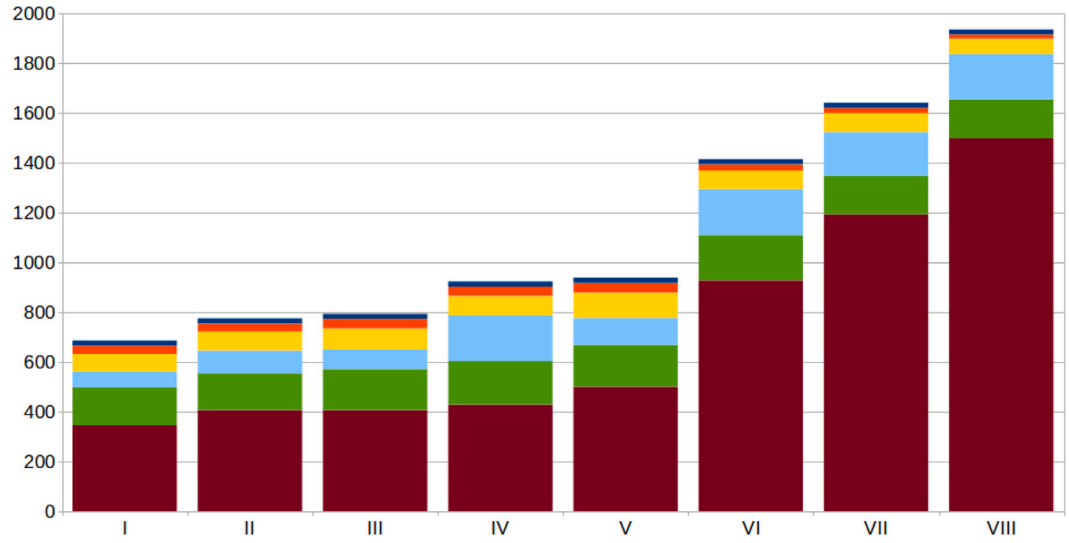


Fig. 8. Impact of intervention complexity on computation time in experiments I–VIII (see text) factored by the main simulation tasks which include intervention, transmission, update, synchronization, output, and initialization. The unit on the *y*-axis is seconds. Clearly time spent on intervention increases with the complexity of interventions involved in the experiment. Specifically, contact tracing interventions (CTD1, CTD2) significantly increase computational complexity of simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

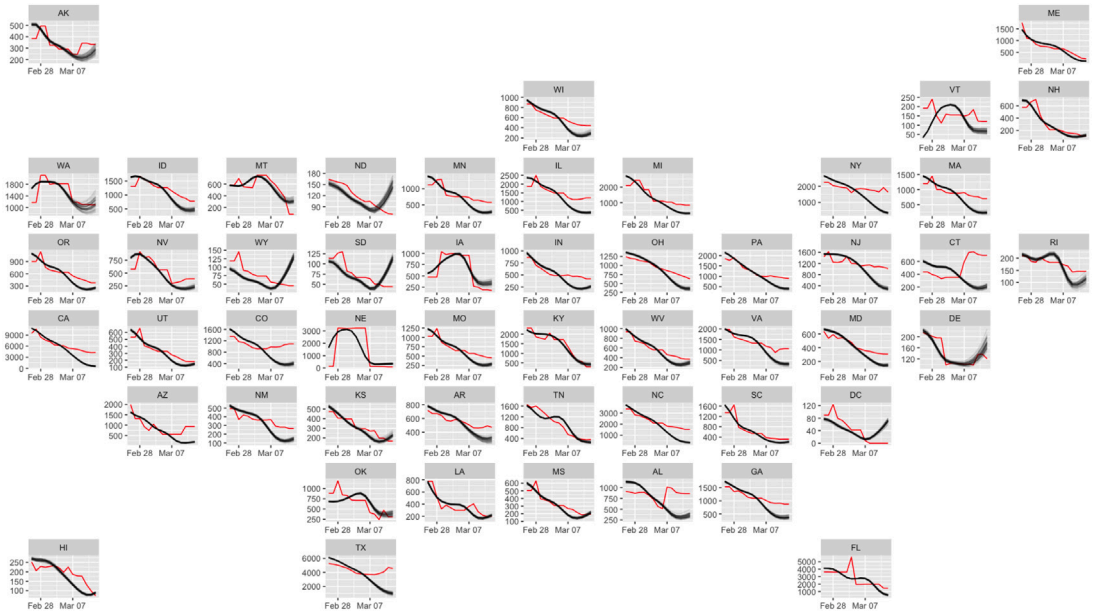


Fig. 9. Simulated data (black) vs. target data (red) in the calibration period for each state. Figure shows how well calibration result fits the time series of case counts from surveillance (The New York Times, 2023). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

how heterogeneity in space, age, and socioeconomic status manifest in the spread patterns.

Summary of findings. By analyzing the projected epidemic outcomes in different scenarios of Round 13, as well as detailed simulated cascades, we find that heterogeneities in our populations and networks, disease models, interventions, and initialization data do lead to heterogeneities in the outcomes: 5.1 aggregate disease outcomes are different between states or at sub-state level in terms of magnitude, trends, or impact of scenario axes considered by the SMH, and differences in network structures between states may have contributed to the differences in disease outcomes; 5.2 disease transmission is dominated by interactions between children which mostly occur in schools. We argue that even when an agent-based model like UVA-EpiHiper produces results at aggregate levels similar to a compartmental or metapopulation model, it provides the possible trajectories, with individual level details, among all trajectories that lead to the same aggregate outcomes. The cascade data generated by UVA-EpiHiper simulations is an example. This enables one to analyze not only what happens but also how it happens, and to obtain unique insight regarding potential interventions to mitigate disease spread.

5.1. Heterogeneities between and within states

Epidemic outcomes over time. We first analyze how heterogeneity between and within different state networks impact the projected epidemic outcomes. In Fig. 4 we show projected cumulative number of infections over time for all scenarios in each US state. We observe different states have different trends (e.g. between California and Oregon), different magnitude (e.g. between California and Washington) even after normalization to counts per 100 K population, different impact

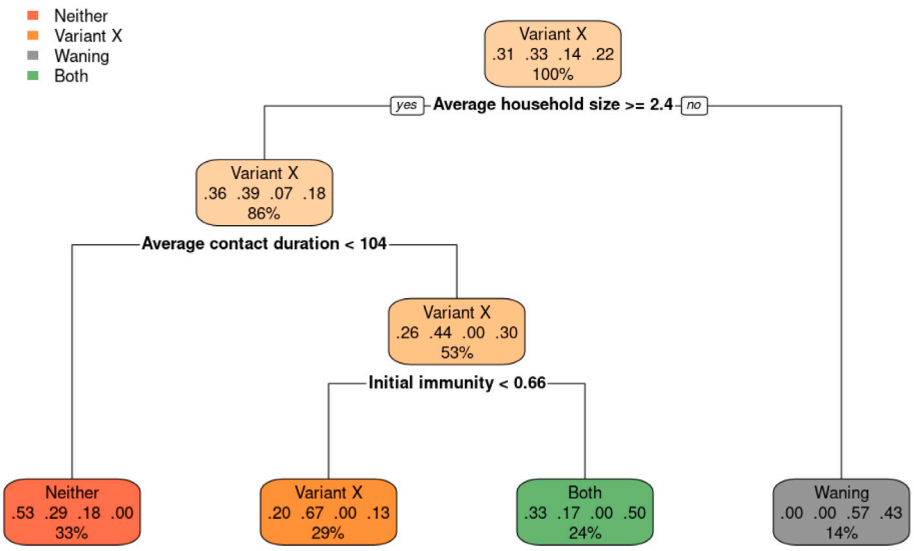


Fig. 10. Decision tree for impacts of immunity waning and emergence of variant X on state level cumulative attack rate. The model includes features on demographics, network structure, initial immunity level, and vaccine coverage, and is fitted using `rpart` (Therneau and Atkinson, 2023). Figure (generated by `rpart.plot` (Milborrow, 2024)) shows that between-state heterogeneity can be partly explained by average household size, average contact duration (the total hours each individual has with all their contacts, averaged over all individuals), and initial immunity level (the immunity each individual has at the beginning of projection period averaged over all individuals).

of waning (e.g. minimal gap between scenarios A and C in Maryland but significant difference between A and C in Vermont), or different impact of new variant (e.g. minimal gap between scenarios A and B in Vermont but significant difference between A and B in Maryland).

Based on whether waning and/or new variant X have significant impacts on the cumulative infections (normalized by population size), we categorize the states into four classes: **neither** has a significant impact, only **variant X** does, only **waning** does, and **both** have significant impacts. We fit a decision tree using features on demographics (average age, average household size), network structure (average contact duration), initial immunity level, and vaccine uptake. We find that the most important predictors are average household size, average contact duration, and initial immunity — see Fig. 10.

On the other hand, in Fig. 5 we observe that different health districts in Virginia seem to have similar trends of cumulative infections over time, as well as similar impact of waning and new variant.[1] But we do observe different magnitude in infection numbers in different health districts even after normalization to counts per 100 K people — higher in Northern Virginia but much lower in the southwest districts.

Detailed comparison between two states. The heterogeneity in population distributions across the study region induces subgraphs of varying density in the contact network. How does the spread in dense subgraphs compare with the spread in the whole network? In this analysis, we consider Massachusetts (MA) and Kansas (KS) which, comparatively have very different county population distributions. MA has one large county, with population around double that of the next largest county, while the top two counties of KS are comparable in size. We analyze the evolution of the infection count over time in the state and in each of the top three counties, and compare them with each other. For scenarios B and D, both of which correspond to emergence of a new variant, we show within-state correlations in Fig. 6 and projected epidemic curves in Fig. 7. The plots for scenarios A and C are in Figs. 11 and 12 in the appendix.

In Fig. 6 we observe that, in both networks and for both scenarios, the state counts are generally more correlated with the top county than with others. For the MA network, we generally observe that the infection curves in the top three counties are representative of the trend for the entire state. One reason for this could be that all these counties are geographically very close to each other (around the Boston area), and therefore, there is a lot of mixing between the populations of these areas. Fig. 13 (in the appendix) shows high edge density between the top county (Middlesex) and the other two counties.

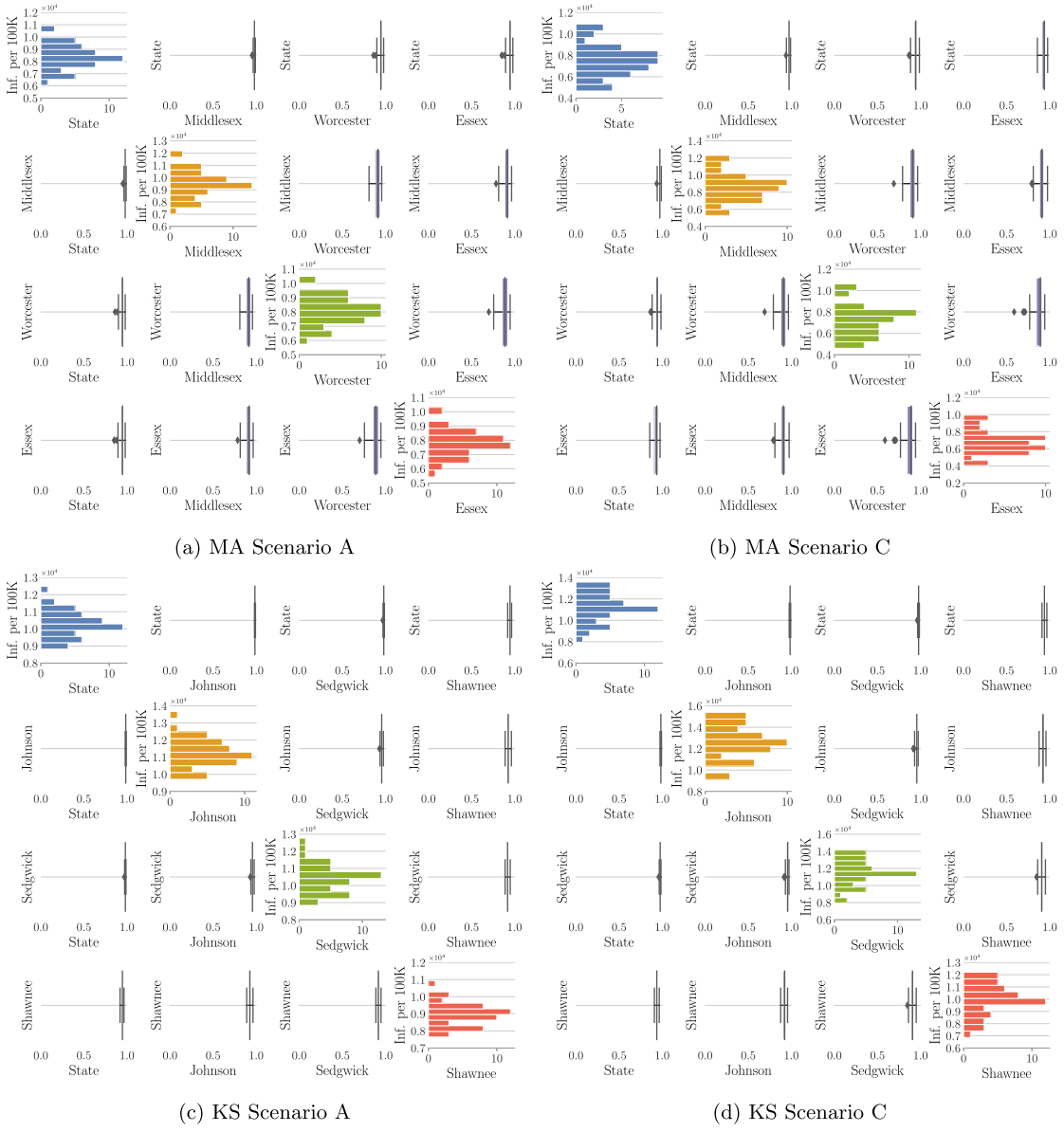
In the case of KS, we observe a similar trend, but there are instances of widely differing infection curves. See for example cascades 2, 4, 7, 8, and 9 in Fig. 7(d). Unlike MA, the top counties of KS are geographically far. Fig. 13 shows relatively low edge density across counties when compared with MA. In Fig. 6, many outliers can be observed for KS in both Scenarios B and D. This can be attributed to the spatial uncertainty in the weekly importation rates of the new variant in the case of Scenarios B and D. In the case of Scenarios A and C, we generally observe very high correlation values.

5.2. Heterogeneities between demographic groups

Here we study the transmissions between and within subpopulations induced by different age groups. Again we consider the KS and MA networks. We partition each state population into three subpopulations: children c (0–17), adults a (18–64), and elderly g (65+). Fig. 14 (in the appendix) shows the number of infections per cascade by age group relative to the subpopulation sizes across scenarios. We first note that in both the KS and MA networks, approximately 60% of nodes are adults and 25% are children. However, the number of infection events (including reinfections) do not reflect the subpopulation sizes. The share of infections for the children is around 40% while this subpopulation only constitutes around 25% of the total population. The higher attack rate in the children, comparing with the other groups, can be explained by the initial immunity level and node degrees in the contact network. Children have higher attack rates than adults due to lower initial immunity levels in children (see Fig. 15(a)). On the other hand, children have higher attack rates than elderly due to higher network degrees of children nodes (see Fig. 15(b)).

We use labeled path motif counts to characterize and quantify the transmissions. For example, a child-infecting-adult event is denoted by $c \to a$. We also consider longer chains of transmission, such as paths of length two. The results of the KS network are in Fig. 16 in the appendix. We observe that transmissions within and across

[1] It is to be noted that while the team had access to heterogeneous vaccine uptake data within Virginia, these were not used for the SMH models, to ensure uniform approach across states. This could partially explain the apparent similarity across Virginia health districts.



(a) MA Scenario A (b) MA Scenario C

(c) KS Scenario A (d) KS Scenario C

Fig. 11. Continued from Fig. 6: Spatial heterogeneity — correlation plots: Comparing the overall spread in a state with the spread in top three counties by population. Two scenarios (A and C) and two states (MA and KS) are considered for comparison. In each case, 50 replicates are considered. The plots are ordered as follows: state followed by top three counties ordered by decreasing population size. The diagonal plots show the histogram of number of infections per 100 K individuals over the simulation. The off-diagonal plots show the distribution of Pearson's correlation coefficient of the infection time series (epicurves) corresponding to each pair of regions.

subpopulations of children and adults dominate the counts. These correspond to mainly school, workplace, and household transmissions. On average, the number of $c \to c$ counts (and $c \to c \to c$ counts among paths of length 2) are the highest, mainly corresponding to school and household transmissions. Many of the transmissions to the elderly are through household contacts. Note that the number of $g \to a \to g$ counts is much higher than $g \to c \to g$. The former count corresponds to transmissions in households as well as in assisted living facilities, while the latter corresponds to only household transmissions. We have similar observations for the MA network in Fig. 17 in the appendix.

6. Discussion

In Section 1, we raised two broad questions that naturally arise when developing and using ABMs such as UVA-EpiHiper: (*i*) when are such models needed and (*ii*) how does one validate such models. Both are important questions, we discuss them briefly below.

Role of ABMs in epidemic scenario modeling. Modeling environments such as UVA-EpiHiper (also see other papers in this issue on use of agent-based models (Moore et al., 2024; Pillai et al., 2023; Rosenstrom et al., 2024)) provide diversity to the pool of models used in the SMH. They are often computationally resource intensive and challenging from the standpoint of software maintenance and model enhancements. On the other hand, such models allow one to: (*i*) get a more detailed picture of the epidemic spread and incorporate the diverse data sets that are often available; (*ii*) look at the model output in ways that are not typically possible in compartmental and statistical models, e.g. looking at the transmission trees to understand chains of transmission; and (*iii*) study the impact of the epidemic and intervention on any desired subgroup so far as it is represented in the digital twin. The last part is important — the basic premise is that models do not have to be made for every question that might arise but a more general model can be used. For instance, a simple compartmental model might not have age-stratified population. Making inferences on age

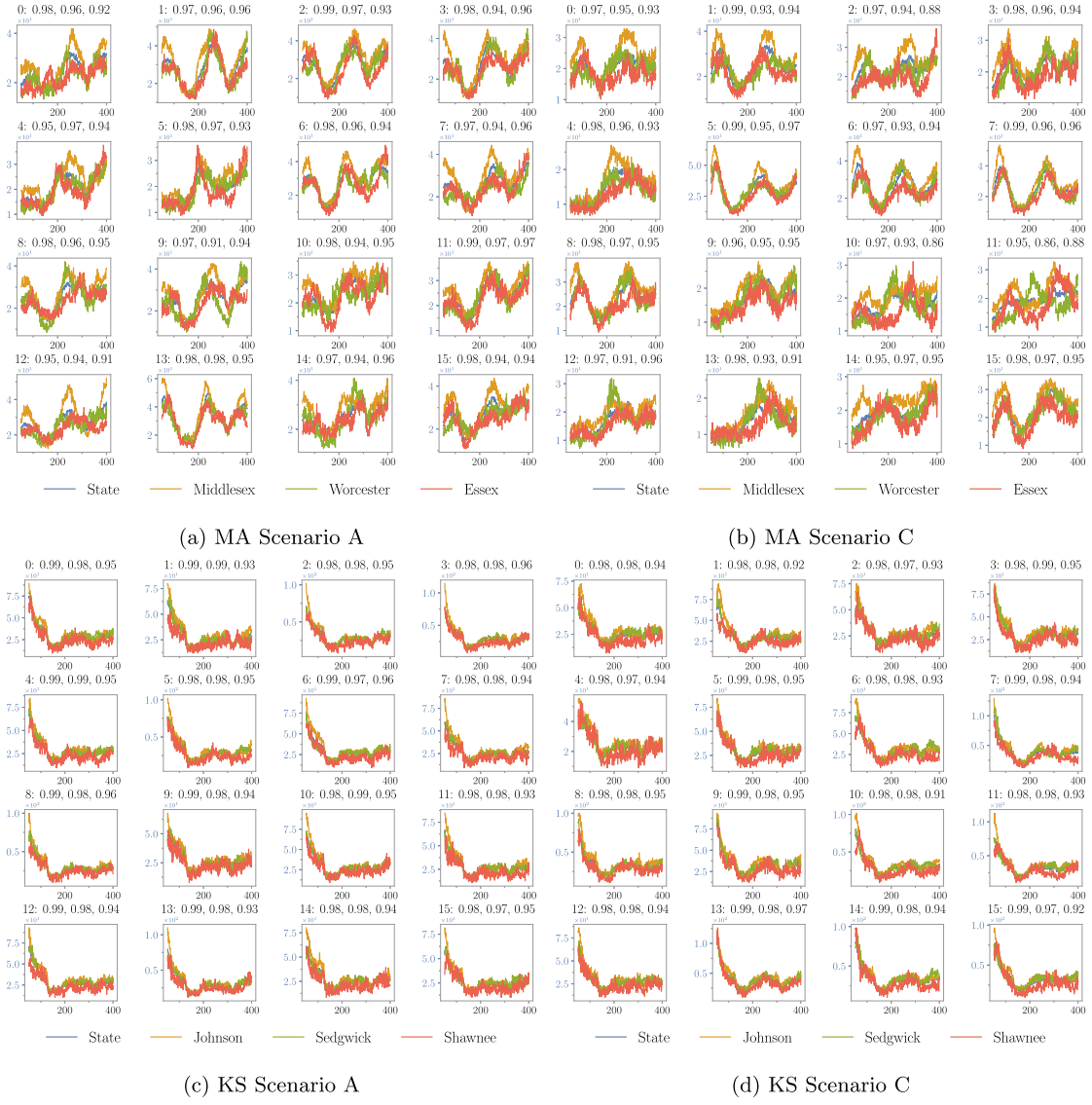


Fig. 12. Spatial heterogeneity — epicurves: continued from Fig. 7. We compare disease spread over time in a state with that in top three counties by population size. The x-axis of each plot shows days from simulation starting date (2022-02-13); the y-axis shows new infection counts per 100 K individuals. Plots are for a subset of cascades. The title of each plot contains the cascade number followed by correlation coefficients between the state infection counts and the county infection counts in the descending order of population size. Plots for Scenarios B and D are in the main text.

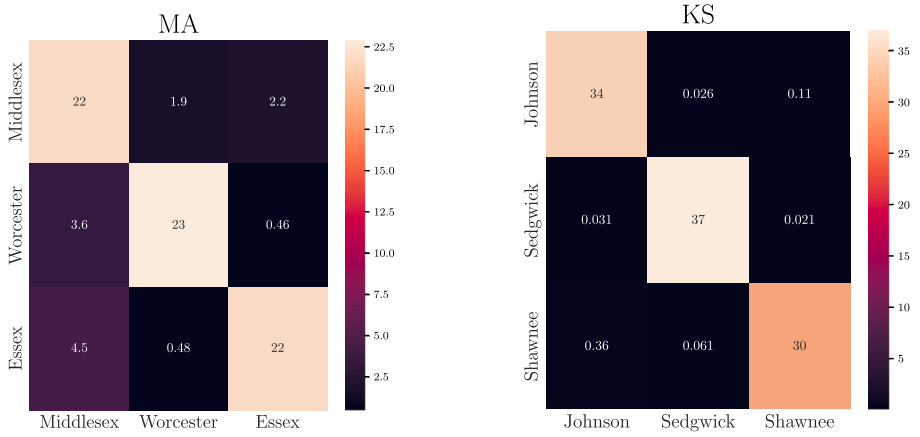


Fig. 13. Edge densities within and across counties. For the non-diagonal entries, the plot shows the ratio of number of edges between two counties to the population of the county in the row. For the diagonal entries, it is the ratio of the number of edges in the county to the population of the county.

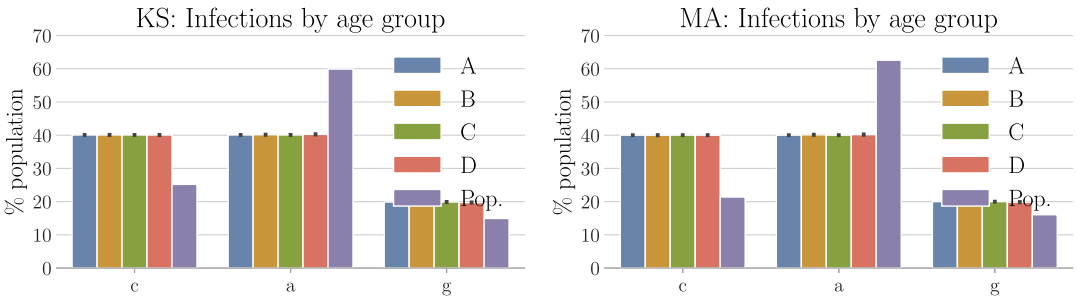
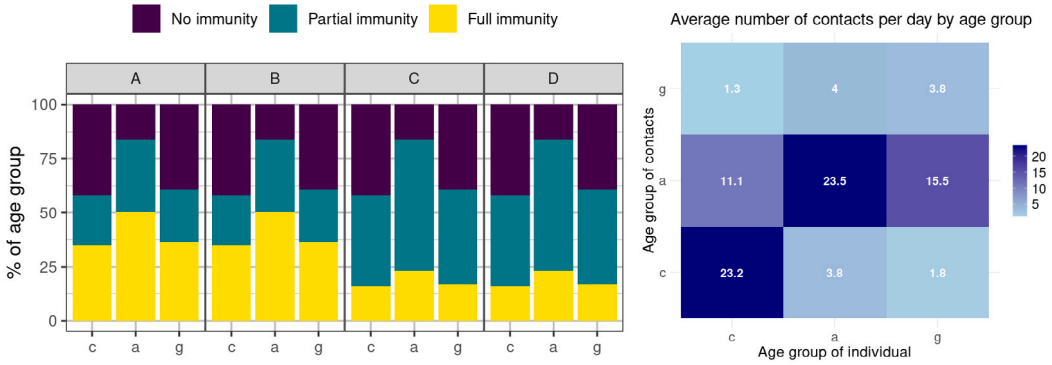


Fig. 14. Fraction of infections per age group and its comparison with population sizes. For each age group, we show the number of infections as a fraction of total infections in different scenarios. It seems about 40% of all infections belong to c and a while 20% belong to g. We also show the number of individuals in each age group as a fraction of total population. Clearly adults (a) has a lower attack rate relative to its group size. This is true in both networks and all scenarios.



(a) Initial immunity distribution by age group. (b) Contact matrix between age groups.

Fig. 15. (a) In all scenarios, initial immunity levels in children and elderly are lower than in adults. (b) Children and adults have more contacts than elderly. The higher attack rate in children is due to the combined effect from both (a) and (b). Figure shows analysis results for MA. The results for KS are similar.

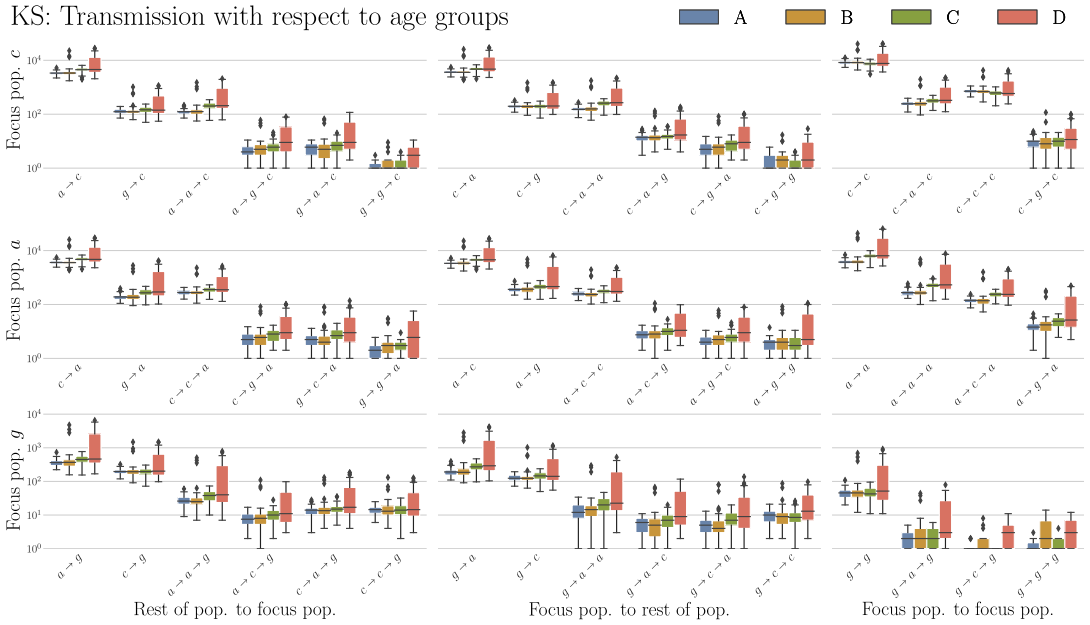


Fig. 16. Transmission motifs for studying the propagation within and across different subpopulations. In this case, we have plotted labeled path motif counts corresponding to adults a, children c and elderly g. We have shown counts of labeled path motifs of lengths 1 and 2 in the transmission cascade. Each row focuses on one subpopulation. The left column corresponds to how the rest of the population affects the focus population. The center column corresponds to how the focus population affects the rest of the population. The right column corresponds to how the transmission occurs from focus population to itself. The results are for the KS network.

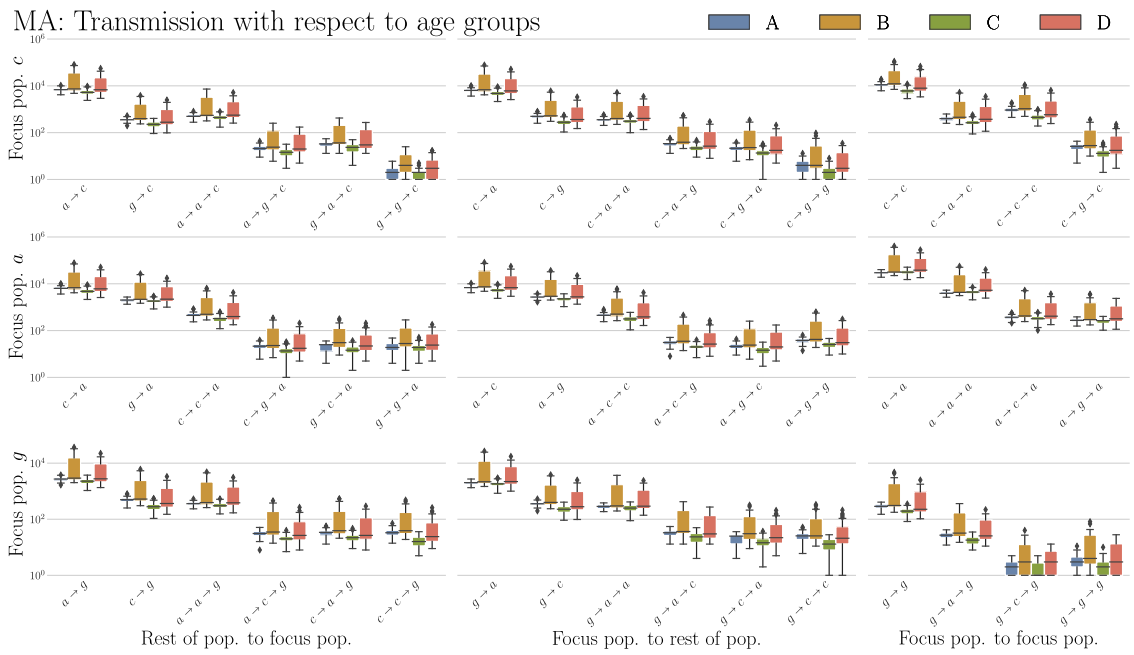


Fig. 17. Continued from Fig. 16: Transmission motifs for studying the propagation within and across different subpopulations. In this case, we have plotted labeled path motif counts corresponding to adults a, children c and elderly g. We have shown counts of labeled path motifs of lengths 1 and 2 in the transmission cascade. Each row focuses on one subpopulation. The left column corresponds to how the rest of the population affects the focus population. The center column corresponds to how the focus population affects the rest of the population. The right column corresponds to how the transmission occurs from focus population to itself. The results are for the MA network.

stratified populations would need compartmental models to include age stratification. But now if one wants to understand the heterogeneity in space, the model will have to add compartments for say each county and each age group. In agent-based models such as UVA-EpiHiper, one gets this for free and it is largely a question of analyzing the output data than constantly changing the model structure. Another use of such models in our opinion is that they provide an impetus to collect highly resolved data sets. The advent of personalized digital devices has already facilitated collection of highly detailed and personalized data sets. We believe these data sets can be used to initialize and calibrate ABMs leading to a more nuanced picture of the epidemic outbreak, see Grekousis and Liu (2021), Aleta et al. (2020), Cencetti et al. (2021), Vogt et al. (2022) for further discussion on this topic.

Validating agent-based epidemic models used for scenario modeling. Validation of complex systems and large ABMs is challenging as well (Carley, 2017; Adiga et al., 2019; Popper, 2005; Oreskes, 2003; Forrester, 1971; Senge and Forrester, 1980; Oreskes, 2018). When using such models for scenario modeling, we can consider three components of validation: (*i*) Data (or external) validation: comparing model output data with real life, in-situ, and in-vivo measurements where state–space predictions by the model are matched with measured data; (*ii*) Structural validation: ensure that local functions or rules used to represent agent (component) interaction, behavior, and decision-making are correct and adequate; and (*iii*) Functional validation: the model should reproduce global well-known structural features of the complex system that is being modeled. In general, data validation alone is not adequate for large scale ABMs, where data matching exercises are usually postdictions of historical information such as matching an epidemiological model output to an infection time series of the 1968 flu season. While useful, such examination can also be misleading and inadequate. In general, postdiction is challenging for SMH; see Runge et al. (2023) for further discussions. Nevertheless, one way to do this is to consider synthetic data based scenarios. We have begun initial discussion on such scenarios as a future SMH round. Beyond retroactive and predictive validity, external validity should also reproduce important features of the state–space of the complex system that is being modeled. Over the last two decades, we have developed a formal computational theory of coevolving graphical dynamical systems (CGDS) (Adiga et al., 2019). The theory allows us to address questions related to structural validation; e.g. comparing two simulations, ensuring the networks are synthesized correctly, etc. Nevertheless this remains an active area of research and much needs to be done in terms of developing formal methods.

Limitations. The heterogeneities in an agent-based model are limited by the heterogeneities in the input data. For example, health disparities between race and ethnicity groups can be modeled directly by UVA-EpiHiper, but meaningfully only if we have relevant data, such as surveillance of cases and deaths and vaccine coverage in each racial/ethnic group. We did not model racial/ethnic disparities in SMH rounds until the recent equity round, where we have race/ethnicity specific data for California and North Carolina to calibrate our model. In SMH work, for the calibration of UVA-EpiHiper we have mainly used surveillance data on confirmed cases as the target. As the collection of such data was discontinued by The New York Times (The New York Times, 2023) in March 2023 and other agents (e.g. The Johns Hopkins Coronavirus Resource Center (The Johns Hopkins Coronavirus Resource Center, 2023)), we have to look for other data sources as the calibration target. Wastewater surveillance data is an option. We plan to expand our digital twin with an additional layer of wastewater surveillance and update the UVA-EpiHiper pipeline to integrate wastewater data-based calibration.

7. Concluding remarks

We described UVA-EpiHiper modeling framework that has been used to support the US COVID-19 SMH over the last 2.5 years. This along with UVA-adaptive (Porebski et al., 2024) were two models used by the UVA team throughout the pandemic. The development, extension, and use of the model as the pandemic evolved required significant efforts by the team. New problems arose as pandemic evolved and in general the models had to be updated constantly. The SMH played an important role in guiding the development of the model. In each round

the scenarios were novel and required new capabilities to be added to the modeling environment. The lively discussions that took place on Fridays proved invaluable in this regard.

As we move forward, the UVA-EpiHiper modeling framework will need to be enhanced for new questions that are likely to arise. This includes: (*i*) further improving the performance of the system; (*ii*) new capabilities in terms of modeling multi-network dynamical processes (e.g. modeling mask wearing or hesitancy and its coevolution at individual level), (*iii*) taking new data sources into account to improve model calibration, (*iv*) modeling inter-state disease transmissions, and (*v*) network-aware initializations that take into account different susceptibility levels of nodes due to their network properties (e.g. degree).

CRediT authorship contribution statement

Jiangzhuo Chen: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Parantapa Bhattacharya:** Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Stefan Hoops:** Data curation, Formal analysis, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Dustin Machi:** Funding acquisition, Resources, Software, Writing – review & editing. **Abhijin Adiga:** Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Henning Mortveit:** Data curation, Formal analysis, Validation, Visualization, Writing – original draft, Writing – review & editing. **Srinivasan Venkatramanan:** Data curation, Funding acquisition, Visualization, Writing – review & editing. **Bryan Lewis:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Madhav Marathe:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the members of UVA Biocomplexity Institute and UVA Research computing. We also thank all members of the SMH for lively discussions throughout the last few years; these discussions were helpful in creating many of the extensions reported in the paper. In particular, we thank Cecile Viboud for carefully going through our model during the consortium's own "peer review process". Her questions and suggestions were thoughtful and helped us improve our models a great deal. Finally, we thank staff members at the PSC, Purdue Supercomputing center and UVA Research computing for their exceptional help and support over the past three years to ensure that our studies could be done in a timely manner. This work was supported in part by the following grants: University of Virginia Strategic Investment Fund (Award Number SIF160), National Science Foundation Grants CCF-1918656 (Expeditions), OAC-1916805 (CINES), IIS-1955797, VDH, United States Grant PV-BII VDH COVID-19 Modeling Program VDH-21-501-0135, DTRA subcontract/ARA S-D00189-15-TO-01-UVA, NIH 2R01GM109718-07, and CDC MIND cooperative agreement U01CK000589. This work used resources, services, and support from the COVID-19 HPC Consortium (https://covid19-hpcconsortium.org/), a private–public effort uniting government, industry, and academic leaders who are volunteering free compute time and resources in support of COVID-19 research as well as computing resources provided by the NSF XSEDE program.

Appendix

A.1. EpiHiper: HPC-enabled simulation platform for agent-based epidemic models

The UVA-EpiHiper modeling process was conducted with a clear path towards a highly efficient and scalable implementation. It is based on our simulation engine, EpiHiper, which can routinely handle populations of size $10^8 - 10^9$ and their detailed contact networks. The EpiHiper software architecture is a hybrid MPI/OpenMP design that is implemented in C++ for high performance. The contact networks can be represented either as text or binary files, with the option to perform pre-partitioning for the desired target combinations of compute nodes and cores. The population (vertices) and their contacts (edges) can be equipped with customizable traits which are exposed to EpiHiper through a Postgres database. This database, which can be shared among computational experiments, has been finely tuned to handle a large number of simultaneous queries, particularly as they occur at the initialization stage of large EpiHiper compute jobs. Similarly, EpiHiper supports a location trait database. As detailed in Section 2.1, each edge is associated with a location, and locations can be augmented with attributes and presented to EpiHiper through this database. This provides a flexible approach to modulating transmission by location (or location type) and constructing highly location-specific interventions. The same applies to interventions cast in terms of the person/edge trait database.

One of the key designs that set EpiHiper apart from other epidemic simulation tools (e.g., (Bershteyn et al., 2018; Ferguson et al., 2020; Hinch et al., 2021; Kerr et al., 2021; Shattock et al., 2022)) is that the *disease models and interventions are specified externally*. Many models support configuration of existing models and interventions. New disease models or interventions will require adding new code to the simulation code base. This design decision for EpiHiper was to cleanly disentangle this aspect and, at least in principle, lower the bar for ease-of-use by removing the need for programming skills from the user as well as the need to understand the software design and implementation encountering such cases. Further while initializing the simulation extensive checks are performed in order to assure contact network, the person and location trait database, the contagion model, initialization, and the interventions are consistent and all required operations can be performed during runtime. If validation fails a detailed message pinpointing the problem location in the configuration files is generated to help problem solving.

A.2. Semantics of interventions in UVA-EpiHiper

Semantics of intervention blocks. The operations within the **once** block are executed whenever the trigger condition C holds, even if the target set is empty. It is used to set variables that are not attached to elements of the intervention target (e.g., the number of available vaccines on a given day). All operations within the **foreach** block are applied to the matching variables of the target elements. Aspects such as compliance are handled through the **sampling** block: several sampling methods are supported where operations are applied to the **sampled** and **nonsampled** sets. We note that recursive application of operation ensembles are supported in the **sampling** control structure.

Semantics of operations. The syntax of operations is provided in Listing 1. In an operation, a variable is assigned the value of an expression, the assignment being scheduled for execution using a non-negative offset **delay** relative to the current time step. The assignment may optionally be assigned an integer **priority** (default value 0) and a **condition** (default value True) which is a Boolean expression that must hold at the scheduled execution time for the assignment to be carried out.

Operation execution. All operations enter a priority queue which is sorted first by scheduled execution time and second by priority. Within a time step, all operations scheduled are processed in priority order. Collections of operations of equal priority are processed in random order. Finally, an operation is executed only if its condition is true at the time of processing. Priorities and conditions allow for fine grained conflict resolution. Regarding the processing order of interventions, one needs to pay careful attention when designing interventions where the order of operations may matter. It is the responsibility of the user constructing the (set of) interventions to assign priorities and conditions to ensure interventions are applied in the intended order.

Set construction. UVA-EpiHiper interventions can target any subset of individuals or edges. Set elements can be selected by internal attributes which may be user defined (traits). Furthermore the external PostgreSQL database allows a user to define arbitrary properties of individuals and locations. Association between UVA-EpiHiper and the database are achieved through unique IDs. Thus sets created through internal attributes or external properties can be combined through set operations (intersection and union). This allows the user to construct any subset of individuals or edges which may be used as targets or in triggers.

Listing 1: The UVA-EpiHiper block structure for operations used in interventions expressed in grammar form.

```
operationEnsemble :=
  once <operationList>
  foreach <operationList>
  sampling <samplingSpecification>
    sampled <operationEnsemble>
    nonsampled <operationEnsemble>

samplingSpecification :=
(
  relativeSampling ( individual | group ) <percentage> |
  absoluteSampling <integer>
)

operationList := <operation>+

operation := <variable> <operator> <expression> \
     delay(<integer>)       \
   [ priority(<integer>) ] \
   [ condition(<bool_expression>) ]

operator := ( = | *= | /= | += | -= )
```

References

Adiga, Abhijin, Barrett, Chris, Eubank, Stephen, Kuhlman, Chris J., Marathe, Madhav V., Mortveit, Henning, et al., 2019. Validating agent-based models of large networked systems. In: 2019 Winter Simulation Conference. WSC, IEEE, pp. 2807–2818.

Aleta, Alberto, Martin-Corral, David, Pastore y Piontti, Ana, Ajelli, Marco, Litvinova, Maria, Chinazzi, Matteo, Dean, Natalie E., Halloran, M. Elizabeth, Longini, Jr., Ira M., Merler, Stefano, et al., 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nat. Hum. Behav. 4 (9), 964–971.

Barrett, Chris, Beckman, Richard, Bisset, Keith, Chen, Jiangzhuo, DuBois, Thomas, Eubank, Stephen, Kumar, V.S. Anil, Lewis, Bryan, Marathe, Madhav V., Srinivasan, Aravind, et al., 2012. Optimizing epidemic protection for socially essential workers. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. pp. 31–40.

Barrett, Christopher L., Bisset, Keith R., Eubank, Stephen G., Feng, Xizhou, Marathe, Madhav V., 2008. EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In: SC '08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing. pp. 1–12.

Barrett, Chris, Bisset, Keith, Leidig, Jonathan, Marathe, Achla, Marathe, Madhav, 2011a. Economic and social impact of influenza mitigation strategies by demographic class. Epidemics 3 (1), 19–31.

Barrett, Christopher L., Eubank, Stephen, Marathe, Achla, Marathe, Madhav V., Pan, Zhengzheng, Swarup, Samarth, 2011b. Information integration to support model-based policy informatics. Innov. J. 16 (1).

Barrett, Christopher, Eubank, Stephen, Marathe, Achla, Marathe, Madhav, Swarup, Samarth, 2015. Synthetic information environments for policy informatics: a distributed cognition perspective. In: Governance in the Information Era. Routledge, pp. 285–302.

Beckman, Richard J., Baggerly, Keith A., McKay, Michael D., 1996. Creating synthetic baseline populations. Transp. Res. A 30 (6), 415–429.

Bershteyn, Anna, Gerardin, Jaline, Bridenbecker, Daniel, Lorton, Christopher W., Bloedow, Jonathan, Baker, Robert S., et al., 2018. Implementation and applications of EMOD, an individual-based multi-disease modeling platform. Pathog. Dis. 76 (5).

Bhattacharya, Parantapa, Chen, Jiangzhuo, Hoops, Stefan, Machi, Dustin, Lewis, Bryan, Venkatramanan, Srinivasan, et al., 2023. Data-driven scalable pipeline using national agent-based models for real-time pandemic response and decision support. Int. J. High Perform. Comput. Appl. 37 (1), 4–27.

Bhattacharya, Parantapa, Machi, Dustin, Chen, Jiangzhuo, Hoops, Stefan, Lewis, Bryan, Mortveit, Henning, et al., 2021. AI-driven agent-based models to study the role of vaccine acceptance in controlling COVID-19 spread in the US. In: 2021 IEEE International Conference on Big Data. pp. 1566–1574.

Bisset, Keith R., Chen, Jiangzhuo, Deodhar, Suruchi, Feng, Xizhou, Ma, Yifei, Marathe, Madhav V., 2014. Indemics: An interactive high-performance computing framework for data-intensive epidemic modeling. ACM Trans. Model. Comput. Simul. 24 (1).

Bisset, Keith R., Chen, Jiangzhuo, Feng, Xizhou, Kumar, V.S. Anil, Marathe, Madhav V., 2009. EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In: Proceedings of the 23rd International Conference on Supercomputing. pp. 430–439.

Bouchnita, Anass, Bi, Kaiming, Fox, Spencer J., Meyers, Lauren Ancel, 2024. Projecting Omicron scenarios in the US while tracking population-level immunity. Epidemics 100746.

Breiman, L., 1984. Classification and regression trees. In: Wadsworth statistics/probability series, Wadsworth International Group, New York.

BuildingFootprintUSA, 2020. https://www.buildingfootprintusa.com/. (Last accessed 1 April 2020).

Carley, Kathleen M., 2017. Validating Computational Models. Technical Report CMU-ISR-17-105, Institute for Software Research, Carnegie Mellon University.

Cattuto, Ciro, Van den Broeck, Wouter, Barrat, Alain, Colizza, Vittoria, Pinton, Jean-François, Vespignani, Alessandro, 2010. Dynamics of person-to-person interactions from distributed RFID sensor networks. PLoS One 5 (7), e11596.

Cencetti, Giulia, Santin, Gabriele, Longa, Antonio, Pigani, Emanuele, Barrat, Alain, Cattuto, Ciro, Lehmann, Sune, Salathe, Marcel, Lepri, Bruno, 2021. Digital proximity tracing on empirical contact networks for pandemic control. Nat. Commun. 12 (1), 1655.

Centers for Disease Control and Prevention, 2023. COVID data tracker. https://covid.cdc.gov/covid-data-tracker/. (Last accessed 24 March 2023).

Chen, Jiangzhuo, Hoops, Stefan, Lewis, Bryan L., Mortveit, Henning S., Venkatramanan, Srini, Wilson, Amanda, 2019. EpiHiper: Modeling and Implementation. Technical Report 2019-003, NSSAC, University of Virginia.

Chen, Jiangzhuo, Marathe, Achla, Marathe, Madhav, 2010. Coevolution of epidemics, social networks, and individual behavior: a case study. In: Advances in Social Computing: Proceedings of the Third International Conference on Social Computing, Behavioral Modeling, and Prediction. pp. 218–227.

Chen, Jiangzhuo, Marathe, Achla, Marathe, Madhav, 2018. Feedback between behavioral adaptations and disease dynamics. Sci. Rep. 8 (1), 12452.

Childers, J. Taylor, Uram, Thomas D., Benjamin, Doug, LeCompte, Thomas J., Papka, Michael E., 2017. An Edge Service for Managing HPC Workflows. In: Proceedings of the Fourth International Workshop on HPC User Support Tools.

Erdős, P., Rényi, A., 1959. On random graphs I. Publ. Math. Debrecen 6, 290–297.

Eubank, Stephen, Guclu, Hasan, Kumar, V.S. Anil, Marathe, Madhav V., Srinivasan, Aravind, Toroczkai, Zoltan, Wang, Nan, 2004. Modelling disease outbreaks in realistic urban social networks. Nature 429 (6988), 180–184.

Feng, Hanhua, Misra, Vishal, Rubenstein, Dan, 2007. PBS: A unified priority-based scheduler. In: Proceedings of the 2007 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems. pp. 203–214.

Ferguson, Neil, Laydon, Daniel, Nedjati Gilani, Gemma, Imai, Natsuko, Ainslie, Kylie, Baguelin, Marc, et al., 2020. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand.

Forrester, Jay W., 1971. Counterintuitive behavior of social systems. Theory Decis. 2 (2), 109–140.

Gallagher, Shannon, Richardson, Lee F., Ventura, Samuel L., Eddy, William F., 2018. SPEW: Synthetic populations and ecosystems of the world. J. Comput. Graph. Statist. 27 (4), 773–784.

Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. 22, 403–434.

Gillespie, Daniel T., 1977. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. 81 (25), 2340–2361.

Grefenstette, John J., Brown, Shawn T., Rosenfeld, Roni, DePasse, Jay, Stone, Nathan T.B., Cooley, Phillip C., et al., 2013. FRED (A Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. BMC Public Health 13 (1), 940.

Grekousis, George, Liu, Ye, 2021. Digital contact tracing, community uptake, and proximity awareness technology to fight COVID-19: a systematic review. Sustain. Cities Soc. 71, 102995.

Halloran, M. Elizabeth, Ferguson, Neil M., Eubank, Stephen, Longini, Jr., Ira M., Cummings, Derek A.T., Lewis, Bryan, et al., 2008. Modeling targeted layered containment of an influenza pandemic in the United States. Proc. Natl. Acad. Sci. 105 (12), 4639–4644.

HERE, 2020. http://www.here.com. (Accessed April 2020).

Hinch, Robert, Probert, William JM, Nurtay, Anel, Kendall, Michelle, Wymant, Chris, Hall, Matthew, et al., 2021. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. PLoS Comput. Biol. 17 (7), e1009146.

Hindman, Benjamin, Konwinski, Andy, Zaharia, Matei, Ghodsi, Ali, Joseph, Anthony D., Katz, Randy H., Shenker, Scott, Stoica, Ion, 2011. Mesos: A platform for fine-grained resource sharing in the data center. In: Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation. NSDI, pp. 295–308.

Kerr, Cliff C., Stuart, Robyn M., Mistry, Dina, Abeysuriya, Romesh G., Rosenfeld, Katherine, Hart, Gregory R., et al., 2021. Covasim: an agent-based model of COVID-19 dynamics and interventions. PLoS Comput. Biol. 17 (7), e1009149.

Lum, Kristian, Chungbaek, Youngyun, Eubank, Stephen, Marathe, Madhav, 2016. A two-stage, fitted values approach to activity matching. Int. J. Transp. 4, 41–56.

Merzky, André, Turilli, Matteo, Titov, Mikhail, Saadi, Aymen Al, Jha, Shantenu, 2021. Design and performance characterization of RADICAL-pilot on leadership-class platforms, CoRR abs/2103.00091.

Microsoft, 2020. U.S. building footprints. https://github.com/Microsoft/USBuildingFootprints.

Milborrow, Stephen, 2024. Rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'.

Mistry, Dina, Litvinova, Maria, Pastore y Piontti, Ana, Chinazzi, Matteo, Fumanelli, Laura, Gomes, Marcelo F.C., et al., 2021. Inferring high-resolution human mixing patterns for disease modeling. Nature Commun. 12 (1), 323.

Moore, Sean, Cavany, Sean, Perkins, T. Alex, Espana, Guido Felipe Camargo, 2024. Projecting the future impact of emerging SARS-CoV-2 variants under uncertainty: modeling the initial omicron outbreak. Epidemics 47, 100759.

Moritz, Philipp, Nishihara, Robert, Wang, Stephanie, Tumanov, Alexey, Liaw, Richard, Liang, Eric, et al., 2018. Ray: A distributed framework for emerging AI applications. In: 13th USENIX Symposium on Operating Systems Design and Implementation OSDI 18. pp. 561–577.

Mortveit, Henning S., Adiga, Abhijin, Barrett, Chris L., Chen, Jiangzhuo, Chungbaek, Youngyun, Eubank, Stephen, et al., 2020. Synthetic Populations and Interaction Networks for the U.S.. Technical Report 2019-025, NSSAC, University of Virginia.

Mossong, Joël, Hens, Niel, Jit, Mark, Beutels, Philippe, Auranen, Kari, Mikolajczyk, Rafael, et al., 2008. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 5, 1.

NCES, 2021. National Center for Education Statistics. http://nces.ed.gov. (Last accessed December 2021).

Newman, Mark E.J., 2003. The structure and function of complex networks. SIAM Rev. 45 (2), 167–256.

Oreskes, Naomi, 2003. The role of quantitative models in science. Model. Ecosyst. Sci. 13, 27.

Oreskes, Naomi, 2018. Why believe a computer? Models, measures, and meaning in the natural world. In: The Earth Around Us. Routledge, pp. 70–82.

Pillai, Alexander N., Toh, Kok Ben, Perdomo, Dianela, Bhargava, Sanjana, Stoltzfus, Arlin, Longini, Jr., Ira M., Pearson, Carl A.B., Hladish, Thomas J., 2023. Agent-based modeling of the COVID-19 pandemic in Florida. Epidemics 47, 100774.

Popper, Karl, 2005. The Logic of Scientific Discovery. Routledge.

Porebski, Przemyslaw, Venkatramanan, Srinivasan, Adiga, Aniruddha, Klahn, Brian, Hurt, Benjamin, Wilson, Mandy L., Chen, Jiangzhuo, Vullikanti, Anil, Marathe, Madhav, Lewis, Bryan, 2024. Data-driven mechanistic framework with stratified immunity and effective transmissibility for COVID-19 scenario projections. Epidemics 47, 100761.

Prem, Kiesha, Cook, Alex R., Jit, Mark, 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLOS Comput. Biol. 13, e1005697.

Rivers, Caitlin M., Lofgren, Eric T., Marathe, Madhav, Eubank, Stephen, Lewis, Bryan L., 2014. Modeling the impact of interventions on an epidemic of Ebola in sierra leone and liberia. PLoS Curr. 6.

Rocklin, Matthew, 2015. Dask: Parallel computation with blocked algorithms and task scheduling. In: Proceedings of the 14th Python in Science Conference, Vol. 130. Citeseer, p. 136.

Rosenstrom, Erik T., Ivy, Julie S., Mayorga, Maria E., Swann, Julie L., 2024. COVSIM: A stochastic agent-based COVID-19 SIMulation model for North Carolina. Epidemics 46, 100752.

Runge, Michael C., Shea, Katriona, Howerton, Emily, Yan, Katie, Hochheiser, Harry, Rosenstrom, Erik, Probert, William J.M., Borchering, Rebecca, Marathe, Madhav V., Lewis, Bryan, Venkatramanan, Srinivasan, Truelove, Shaun, Lessler, Justin, Viboud, Cécile, 2023. Scenario design for infectious disease projections: Integrating concepts from decision analysis and experimental design. medRxiv.

Salim, Michael A., Uram, Thomas D., Childers, J. Taylor, Balaprakash, Prasanna, Vishwanath, Venkatram, Papka, Michael E., 2019. Balsam: Automated Scheduling and Execution of Dynamic, Data-Intensive HPC Workflows. https://arxiv.org/abs/1909.08704v1.

Scenario Modeling Hub, 2022. COVID-19 Scenario Modeling Hub Round 13 scenarios. https://github.com/midas-network/covid19-scenario-modeling-hub/blob/master/previous-rounds/README_Round13.md.

Scenario Modeling Hub, 2023a. The COVID-19 Scenario Modeling Hub. https://covid19scenariomodelinghub.org/. (Last accessed 31 December 2023).

Scenario Modeling Hub, 2023b. Flu Scenario Modeling Hub. https://fluscenariomodelinghub.org/. (Last accessed 31 December 2023).

Scenario Modeling Hub, 2023c. RSV Scenario Modeling Hub. https://github.com/midas-network/rsv-scenario-modeling-hub. (Last accessed 31 December 2023).

Scenario Modeling Hub, 2023d. The European COVID-19 Scenario Hub. https://covid19scenariohub.eu/. (Last accessed 31 December 2023).

Senge, Peter M., Forrester, Jay W., 1980. Tests for building confidence in system dynamics models. Syst. Dyn. TIMS Stud. Manag. Sci. 14, 209–228.

Shattock, Andrew J., Le Rutte, Epke A., Dünner, Robert P., Sen, Swapnoleena, Kelly, Sherrie L., Chitnis, Nakul, Penny, Melissa A., 2022. Impact of vaccination and non-pharmaceutical interventions on SARS-CoV-2 dynamics in Switzerland. Epidemics 38, 100535.

Shoukat, Affan, Wells, Chad R., Langley, Joanne M., Singer, Burton H., Galvani, Alison P., Moghadas, Seyed M., 2020. Projecting demand for critical care beds during COVID-19 outbreaks in Canada. Can. Med. Assoc. J. 192 (19), E489–E496.

Socioeconomic Data and Applications Center, 2020. Gridded population of the world (GPW), v4. https://sedac.ciesin.columbia.edu/data/collection/gpw-v4. (Last accessed 1 April 2020).

SocioPatterns, 2023. http://www.sociopatterns.org/. (Last accessed 17 February 2023).

Somers, Mark, 2019. Leiden Grid Infrastructure. https://lgi.tc.lic.leidenuniv.nl/LGI/docs/LGI.pdf.

Srivastava, Ajitesh, 2023. The variations of SIkJalpha model for COVID-19 forecasting and scenario projections. Epidemics 45, 100729.

Tatem, Andrew J., 2017. WorldPop, open data for spatial demography. Sci. Data 4.

The Johns Hopkins Coronavirus Resource Center, 2023. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19.

The New York Times, 2023. Coronavirus (Covid-19) data in the United States. https://github.com/nytimes/covid-19-data. (Last accessed 24 March 2023).

Therneau, Terry, Atkinson, Beth, 2023. Rpart: Recursive partitioning and regression trees.

U.S. Census, 2020a. 2011–2015 5-year ACS commuting flows. (Last accessed April 2020).

U.S. Census, 2020b. Longitudinal Employer-Household Dynamics. https://lehd.ces.census.gov/data/. (Last accessed 26 November 2020).

U.S. Census, 2021a. 2010 Census Urban and Rural Classification. (Last accessed June 2021).

U.S. Census, 2021b. Public Use Microdata Sample (PUMS). Last accessed 24 May 2021.

U.S. Department of Labor, Bureau of Labor Statistics, 2020. The American Time Use Survey (ATUS). (Last accessed February 2020).

U.S. Department of Transportation, Federal Highway Administration, 2020. The National Household Travel Survey (NHTS). (Last accessed February 2020).

Vavilapalli, Vinod Kumar, Murthy, Arun C., Douglas, Chris, Agarwal, Sharad, Konar, Mahadev, Evans, Robert, et al., 2013. Apache hadoop yarn: Yet another resource negotiator. In: Proceedings of the 4th Annual Symposium on Cloud Computing. pp. 1–16.

Venkatramanan, Srinivasan, Chen, Jiangzhuo, Fadikar, Arindam, Gupta, Sandeep, Higdon, Dave, Lewis, Bryan, Marathe, Madhav, Mortveit, Henning, Vullikanti, Anil, 2019. Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints. PLoS Comput. Biol. 15 (9), e1007111.

Venkatramanan, Srinivasan, Lewis, Bryan, Chen, Jiangzhuo, Higdon, Dave, Vullikanti, Anil, Marathe, Madhav, 2018. Using data-driven agent-based models for forecasting emerging infectious diseases. Epidemics 22, 43–49.

Vogt, Florian, Haire, Bridget, Selvey, Linda, Katelaris, Anthea L., Kaldor, John, 2022. Effectiveness evaluation of digital contact tracing for COVID-19 in New South Wales, Australia. Lancet Public Health 7 (3), e250–e258.

Weber, Eric, Moehl, Jessica, Weston, Spencer, Rose, Amy, Sims, Kelly, 2021. LandScan USA 2020 [Data set]. Oak Ridge National Laboratory. http://dx.doi.org/10.48690/1523373.

Wheaton, William D., Cajka, James C., Chasteen, Bernadette M., Wagener, Diane K., Cooley, Philip C., Ganapathi, Laxminarayana, Roberts, Douglas J., Allpress, Justine L., 2009. Synthesized population databases: A US geospatial database for agent-based models. Methods Rep. (RTI Press) 2009 (10), 905.

Yoo, Andy B., Jette, Morris A., Grondona, Mark, 2003. Slurm: Simple linux utility for resource management. In: Workshop on Job Scheduling Strategies for Parallel Processing. Springer, pp. 44–60.