Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

AGENT-BASED SIMULATION FRAMEWORK FOR MULTI-VARIANT SURVEILLANCE

Sifat Afroj Moon[1,3], Jiangzhuo Chen[1], Baltazar Espinoza[1], Bryan Lewis[1], Madhav Marathe[1,2], Joseph Outten[1], Srinivasan Venkatramanan[1], Anil Vullikanti[1,2], and Andrew Warren[1]

[1]Biocomplexity Institute (BII), University of Virginia, Charlottesville, VA, USA
[2]Dept. of Computer Science, University of Virginia, Charlottesville, VA, USA
[3]Computational Sciences and Engineering Division, Oak Ridge National Lab., Oak Ridge, TN, USA

ABSTRACT

Early detection of an emerging VOC (Variant-Of-Concern) is essential for effective preparedness for a disease like COVID-19. The spreading of an emerging VOC not only depends on the disease dynamics of itself but also depends on the state of the circulating variants and the susceptibility of the population. Resources for testing are typically quite limited, and a number of strategies have been considered for deploying them. However, it has been difficult to evaluate the performance of such strategies, especially higher order effects, and inequities, while incorporating constraints on these resources. Here, we develop an agent-based surveillance framework, NetworkDetect, to understand the early warning system of an emerging VOC. Our framework allows us to incorporate various population heterogeneities and resource constraints.

1 INTRODUCTION

The importance of effective surveillance was highlighted during the COVID-19 pandemic. In the initial stages (especially before the vaccine became widely available), an important goal was to determine quickly if any infections have occured within a subpopulation. As the pandemic evolved, and different variants emerged, the goals of surveillance changed. A number of variants, some of which were labeled as "Variants of Concern (VOCs)", such as Delta (B.1.617) and Omicron (B.1.1.529) (Karim and Karim 2021), were associated with increased virulence and transmissibility (Burki 2021). Further, some of the vaccines had reduced efficacy for some variants (Bushman et al. 2021). These highly transmissible variants led to challenging public health questions, e.g., deciding how and whether to relax interventions, the cadence of vaccination, etc. (Skegg et al. 2021). A significant challenge of COVID-19 is that a large fraction of infections were unreported due to various factors such as the lack of testing and asymptomatic infections (Bai et al. 2020; Aguilar et al. 2020; Shaman 2020). Early detection of the emergence of VOCs, and estimation of their prevalence among all infections were crucial in the public health planning process.

Different testing techniques have been used for COVID-19: antigen tests, PCR, and genomic sequencing from saliva, blood, and stool. A significant challenge is the cost associated with testing, both in terms of material and staffing costs. For instance, during the early stages of COVID-19, there was a shortage of testing sites and resources in many regions across the world. Large scale seroprevalence testing for accurate estimation of disease prevalence has only been done in a small number of regions, e.g., (Sood et al. 2020; Zhang et al. 2020), due to its expense. As the pandemic evolved, health departments and other organizations worldwide have explored different strategies for the detection of VOCs; this includes random sampling within the population and within different subpopulations based on their prevalence. Evaluating the effectiveness of such strategies is challenging due to multiple reasons, not the least of which is understanding ground truth for how new variants enter and spread through the population, motivating the need for simulation. An additional complication is that the costs of different kinds of tests vary, and genomic testing, which is generally the only test that can identify a VOC, is most expensive. Prioritization

979-8-3315-3420-2/24/$31.00 ©2024

of testing within subpopulations can lead to fairness and equity issues, especially since the pandemic was associated with a variety of inequitable health outcomes.

Our contributions.

Formalization of the Variant Surveillance Problem: We present a clear and structured definition of the variant surveillance problem, outlining its key parameters, constraints, and evaluation metrics. This formalization facilitates a systematic and rigorous analysis of surveillance strategies.

Flexible Framework for Strategy Evaluation: In conjunction we also present an agent-based framework, NETWORKDETECT (Figure 1), for evaluating a broad class of surveillance strategies to solve VARIANT SURVEILLANCE problem.

Integration of Heterogeneous Factors: Our approach integrates detailed activity-based population models (digital twins) with multi-variant disease dynamics, enabling the evaluation of surveillance strategies under realistic conditions. This includes considering population heterogeneities (contact network characteristics, demographics, disease history), resource limitations, and equity considerations. For concreteness, we focus on two types of testing methods— PCR and genomic testing. A surveillance strategy involves specifying subsets of individuals $S_{pcr}(t), S_g(t)$ to be tested by PCR and genomic tests, respectively, at each time t. Our agent-based model allows us to specify the individuals to be tested using different kinds of characteristics, e.g., contact network characteristics, demographic features of individuals, disease history of patients, community prevalence, along with constraints for surveillance (e.g., budget and access in different regions). Our framework also naturally provides a way to incorporate different metrics of equity and fairness.

Case Study and Insights: We demonstrate the capabilities of NETWORKDETECT through a case study focusing on variation in timing and geographic importation. Our analysis provides insights into the impact of budget allocation, partial immunity, and spatial factors on the timeliness of VOC detection.

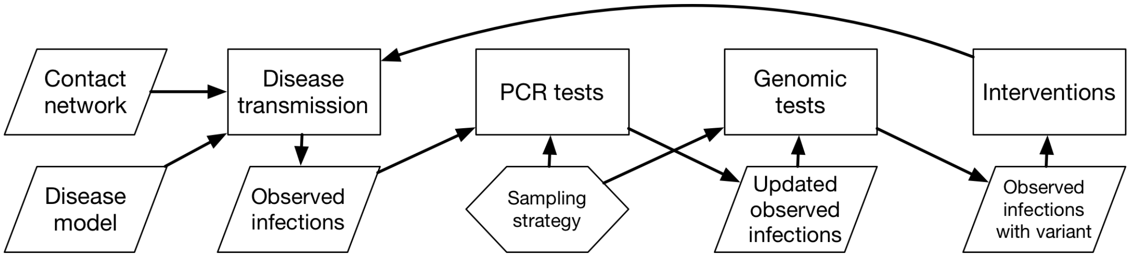


Figure 1: Illustration of the NETWORKDETECT framework.

2 BACKGROUND

2.1 Synthetic Population

For this study, we need detailed representations of individuals, their demographic characteristics and activities which have a bearing on infection spread. Detailed datasets on populations and contact networks are not available due to privacy and sensitivity constraints. Even in the few cases where small datasets related to mobility and activities are available (e.g. Google COVID-19 Community Mobility Reports (Google 2022)), these are snapshots, and do not contain enough information about individuals and their contacts that can support large scale infectious disease analyses. Here, we use a network model developed using a “first principles” approach (also referred to as a digital twin) (Chen et al. 2021; Eubank et al. 2010; Eubank et al. 2004; Barrett et al. 2009) for our study, which is constructed by integrating a large number of datasets from commercial and public sources, such as Census, NAVTEQ, American Time Use Survey data (ATUS), National Household Travel Survey Data. We summarize this approach below, but refer to (Chen et al. 2021; Eubank et al. 2010; Eubank et al. 2004; Barrett et al. 2009) for more details.

The synthetic population (denoted by set V) represents all individuals (referred to as a “node” from now on) within a geographical region (Virginia, USA, in our case), along with respective households. Each node is endowed with a rich set of demographic features, e.g., age, gender, household size, household location, household income. Each synthetic individual is placed in a household with other synthetic people, and each household is located geographically in such a way that a census of the synthetic population yields results statistically indistinguishable from the original census data, if they are both aggregated to the block group level.

Synthetic individuals are assigned daily activities, which specify the sequence and duration of activities (e.g., home, work, shopping) they perform. This is done using time-use surveys (American Time Use Survey data (ATUS 2021), National Household Travel Survey Data (NHTS 2021) and Multinational Time Use Study). A geolocation is assigned for each activity. These locations are based on data from different sources, such as HERE/NAVTEQ (HERE 2021), National Center for Education Statistics (NCES 2021), LandScan (LandScan 2021), and OpenStreetMap (OpenStreetMap 2021), and the assignment of specific locations to activities done by individuals is done based on statistical models of distance traveled by individuals.

A contact network $G(V,E)$ is constructed through co-location— an edge $(u,v) \in E$ is assigned between a pair of nodes $u,v \in V$ which are at the same location at the same time. Each edge $e \in E$ is associated with a duration and other attributes, which specify the nature of contact. The co-location based social contact network is used for modeling disease transmission in the region. The social contact network of the Virginia population consists of more than 7.6 million nodes and 371.9 million edges.

2.2 Agent-Based Disease Simulation

We develop an agent-based model framework to simulate and explore the VARIANT SURVEILLANCE problem. Disease simulation is a component of the NETWORKDETECT framework (Figure 1). To simulate disease dynamics at the individual level, we have used EpiHiper, a high-performance computing (HPC) agent-based epidemic platform (Hoops et al. 2021; Machi et al. 2021). In this framework each agent represents a person from the locality embodied by the digital twin and each infected agent can spread the disease according to its connectivity in G.

To model the dynamics of SARS-CoV-2 variants, we use a two-variants network contagion model (Pastor-Satorras et al. 2015; Chen et al. 2021; Moon and Scoglio 2021); where the first variant represents existing variants and the second variant represents the new emerging VOC (Figure 2). In this work, we use the term 'VOC' to denote 'variant of concern' in a broad sense, encompassing any variant under investigation, rather than strictly adhering to the specific variants designated by the World Health Organization (WHO). Here, we extend Susceptible-Exposed-Infected-Recovered (SEIR) compartmental epidemic model to incorporate partial immunity and waning immunity. We model that starting point of the variant-1 is day 0 and variant-2 imports in the system at day D_2.

Each individual or agent has nine possible disease states: susceptible (S), exposed by variant-1 (E_1), infectious with variant-1 but asymptomatic (I_{as1}), infectious with variant-1 but symptomatic (I_{s1}), recovered from variant-1 (R_1), exposed by variant-2 (E_2), infectious with variant-2 but asymptomatic (I_{as2}), infectious with variant-2 but symptomatic (I_{s2}), and recovered from variant-2 (R_2). In Figure 2, β_1 and β_2 are the transmissibility of the variant-1 and variant-2; μ is the incubation rate; δ is the recovery rate; δ_{w1} and δ_{w2} are the weaning rates; and α is the partial immunity parameter. In the agent-based model, each individual has its own disease dynamic. The transition of an agent from one state to another state is functionally dependent on agent attributes, exposure through neighbor in G, and other interventions or protective behaviors.

2.3 Testing Methods

A variety of tests are used for identifying cases (World Health Organization). A test on an individual i indicates if the node i is infected or not. In most cases, this does not give information on the specific variant; genomic sequencing is needed to determine if the individual is infected by the new VOC, i.e. variant-2. In the case of SARS-Cov-2, infected individuals may remain asymptomatic for part or all of their infectious

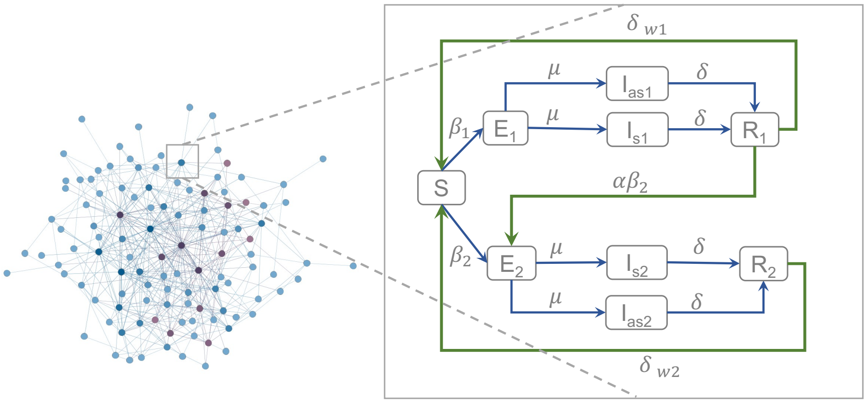


Figure 2: Agent-based 2-variant epidemic model with partial and waning immunity. The zoom-in window presents the state transition diagram of the epidemic model. Here, node represents an individual person or agent.

period or forgo testing which can lead to under-reporting and hindering accurate modelling of virus and variant distribution. In this work we assume complete fidelity in a test's ability to determine infection and use the terms both testing and PCR interchangeably for this presumed assay.

3 VARIANT SURVEILLANCE PROBLEM MODELING

VARIANT SURVEILLANCE enables tracking and characterization of new variants and is best served by choosing individuals for PCR and genomic sequencing efficiently. We use $S_{pcr}(t)$ and $S_g(t)$ to denote the subsets of nodes chosen for PCR and sequencing at time t, respectively; we use $\mathbf{S_{pcr}}$ and $\mathbf{S_g}$ to denote the vectors specifying the subsets over time. Our framework specifies a budget for running tests to conform to real world lab constraints); let B_{pcr} and B_g denote the number of PCR and genomic tests that can be run at each time, respectively. A surveillance strategy $(\mathbf{S_{pcr}}, \mathbf{S_g})$ is feasible if $|S_{pcr}(t)| \leq B_{pcr}$ and $|S_g(t)| \leq B_g$ for time t. Usually B_g is much smaller than B_{pcr}.

There are several potential metrics to evaluate a surveillance strategy $(\mathbf{S_{pcr}}, \mathbf{S_g})$. The probability of detection of a VOC or the minimum time so that the probability of detection of the VOC is at least $\alpha\%$. Here, we will formalize them using the stochastic agent-based disease transmission model. We denote the set of nodes in state X at time t by $X(t)$. For example, $I_{as1}(t)$ denotes the nodes in I_{as1} state at time t. We use $I_i(t)$ to denote all nodes infectious with variant-i at t, e.g. $I_1(t) = I_{as1}(t) \cup I_{s1}(t)$. Let $I(t) = I_1(t) \cup I_2(t)$ denote all infectious nodes at t.

For any $v \in S_{pcr}(t)$, if $v \in I(t)$ (i.e., if the node v is PCR positive), we assume that v is infectious, but do not know with which variant. If $v \in S_g(t) \cap I_1(t)$, we assume that we can infer that v is infected with variant-1; similarly, if $v \in S_g(t) \cap I_2(t)$, we assume that it is infected with variant-2. We say that variant 2 is detected at time t if $S_g \cap I_2(t) \neq \emptyset$; the associated probability of this quantity, Pr[variant 2 is detected at time t], can be estimated using the agent-based simulation model. Examples of metrics we consider are

- Probability of detecting an emerging VOC on or before day $D_2 + d$, denoted by $P(D_2 + d, \mathbf{S_{pcr}}, \mathbf{S_g})$. We have

$$P(D_2 + d, \mathbf{S_{pcr}}, \mathbf{S_g}) = 1 - \prod_{t=1}^{D_2+d} (1 - \Pr[\text{variant 2 is detected at time (or day) } t]). \qquad (1)$$

Given a fixed rate and sampling budget this calculation assumes homogeneous sampling, e.g., (Wohl et al. 2022). This calculation serves as a useful comparison for heterogeneous sampling strategies.

- Using the probability equation we can determine the minimum time d so that the probability of detection of variant 2 is at least $\alpha\%$, denoted by $T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g})$. We have

$$T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g}) = \min\{d : P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g}) \geq \alpha\%\} \tag{2}$$

The metrics $P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g})$ and $T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g})$ both depend on the sampled variant prevalence which in turn depend on the contact network, disease model and initial conditions; however, we drop them from the notation for simplicity. T_{95} is the earliest day d when the cumulative probability $P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g})$ reaches 0.95.

The VARIANT SURVEILLANCE problem. This involves evaluating surveillance strategies $(\mathbf{S_{pcr}},\mathbf{S_g})$ based on

- Metrics $P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g})$ and $T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g})$
- The strategy $(\mathbf{S_{pcr}},\mathbf{S_g})$ is feasible with respect to the budgets B_{pcr} and B_g.

We note that individuals need not fully comply with surveillance strategies recommended by health departments (unless an explicit public health directive is in place, which was the case during some period of the COVID-19 pandemic). Individuals might not comply with such testing strategies if their interests are not fully aligned with the objectives of health departments. Often, equity is an important issue in this kind of alignment, and we also consider this kind of metric.

VARIANT SURVEILLANCE with equity constraints. We consider the metrics $P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g})$ and $T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g})$ in terms of detection of the VOC within specific subpopulations (e.g., based on age, race, ethnicity), prevalence of the disease within these subpopulations when the VOC is detected, and how much of the testing budget is spent within the subpopulations. The goal here is to design testing strategies $(\mathbf{S_{pcr}},\mathbf{S_g})$ which ensure fairness with respect to all these criteria.

4 NETWORKDETECT FRAMEWORK

We describe here our agent-based framework, NETWORKDETECT, for modeling a rich class of surveillance strategies for two variant spread. We illustrate our framework in Figure 1 and describe the steps as follows.

- We construct a network model $G(V,E)$ for a given region (as described in Section 2.1).
- The ABM is initialized with a set $E_1(0)$ of variant-1 infections. Similarly, a set $E_2(D_2)$ of 2nd variant sources is chosen at time D_2, from susceptible nodes S.
- At each time t, the disease transmission and surveillance strategy involves the following steps
  - Update the disease states of all nodes from the ABM disease simulation.
  - Pick a subset $S_{pcr}(t)$ from the whole population V based on a sampling strategy (which will be specified later) with $|S_{pcr}(t)| \leq B_{pcr}$. PCR tests are done on all nodes in $S_{pcr}(t)$. Let $S^+_{pcr}(t)$ be the subset testing positive.
  - Pick a subset $S_g(t)$ based on a sampling strategy, which can take into account $I_{s1}(t), I_{s2}(t), S^+_{pcr}(t)$ (specified later), with $|S_g(t)| \leq B_g$, for whom genomic tests will be done. Let $S^+_g(t)$ denote the subset of nodes which test positive for the second variant.
  - The nodes in $S^+_{pcr}(t)$ and $S^+_g(t)$ implement local interventions in the model.
- Compute the metrics $P(D_2+d,\mathbf{S_{pcr}},\mathbf{S_g})$ and $T_\alpha(\mathbf{S_{pcr}},\mathbf{S_g})$.

The disease transmission computation has a time complexity of $O(|V|+|E|)$ and produces $O(|V|)$ state transition data for each time step t and the computation is distributed over multiple processes with a parallel algorithm.

Memory-based realistic surveillance: The NETWORKDETECT allows us to implement memory in the sampling process. We have implemented one memory constraint for the PCR budget B_{PCR} and two memory constraints for the genomic budget B_G. For the PCR budget memory constraint, we apply that if a node is

PCR positive on time t, the framework will not use the B_{PCR} budget again on that node until it recovers. For the genomic sequence budget memory constraints, we consider that: 1) if a node is selected for $S_g(t)$, our sampling strategy will not use B_g budget until it recovers, and 2) if a node is a genomic test positive (= detected by variant-2), our sampling strategy will never use the B_g budget on that node again. Therefore, the sampling framework keeps the memory of three types of nodes: 1) PCR test positive, 2) genomic tested, and 3) genomic test positive. Let M_{pcr} be the set of nodes that are PCR tested but not recovered yet, and M_g is the set of nodes that are genomic tested or genomic test positive but not recovered yet.

- Pick a subset $S_{pcr}(t) \subset V - I_{s1}(t) \cup I_{s2}(t) - M_{pcr}$ based on a sampling strategy (which will be specified later) with $|S_{pcr}(t)| \leq B_{pcr}$. PCR tests are done on all nodes in $S_{pcr}(t)$.
- Let $U(t) = I_{s1}(t) \cup I_{s2}(t) \cup S^{+}_{pcr}(t) - M_g$.
- Pick a subset $S_g(t) \subset U(t)$ based on a sampling strategy (speficied later), with $|S_g(t)| \leq B_g$, for whom genomic tests will be done. Let $S^{+}_{g}(t)$ denote the subset of nodes which test positive for the second variant.

We describe some of the possible strategies for choosing $S_{pcr}(t), S_g(t)$ below.

Strategies for choosing $S_{pcr}(t)$ and $S_g(t)$

- ***Random sampling***: In this strategy, we randomly select $S_{pcr}(t)$ from $V - I_s(t)$ with the PCR budget memory constraint. Here, $I_s(t)$ denotes the set of symptomatic people on time t. Then, we randomly select $S_g(t)$ from the symptomatic PCR test positive and asymptomatic PCR test positive nodes with the two genomic budget memory constraints.
- ***Network centrality sampling***: The NETWORKDETECT allows us to evaluate the sampling of different central nodes in the contact network. For a specific centrality measure, such as degree, betweenness, or eigenvector, we first order $V - I_s(t)$ nodes according to that particular centrality. For example, for degree centrality, we first order nodes according to their degree and then use two strategies. In the first strategy, we have used the following steps:
  - Pick first B_{pcr} nodes from the ordered $V - I_s(t)$ list for the PCR sampling according to the memory constraint.
  - Form a candidate pool for the genomic test from symptomatic nodes and asymptomatic PCR test-positive nodes. Order the nodes in the candidate pool according to their degree.
  - Select top nodes from the ordered candidate pool for the genomic test set $S_g(t)$ with the memory constraint.

  In the second strategy, we use the following steps:
  - Randomly select $S_{pcr}(t)$ from the top quartile of the ordered $V - I_s(t)$ with the PCR budget memory constraint
  - Follow the second step of the first strategy.
  - Randomly select $S_g(t)$ from the top quartile of the ordered candidate pool with the genomic budget memory constraints.
- ***Surveillance with fairness constraints***: Many notions of fairness can be used. Here, we describe surveillance strategies which ensure demographic parity. We assume the population V is partitioned into groups $V = V_1 \cup \ldots \cup V_k$; for instance each V_i could be the population within a county or some other geographical region. A simplest fairness criterion is to ensure that S_{pcr} and S_g are split as uniform as possible across each V_i. A different notion is to ensure that the probability of detection within each V_i is similar. Our framework allows us to evaluate the fairness of such strategies; designing fair strategies is a challenging open question.

5 RESULTS AND DISCUSSIONS

We present results for $T_\alpha(\mathbf{S_{pcr}}, \mathbf{S_g})$, for random with memory-based surveillance, on a digital twin for Virginia. Here, we consider the simplest disease model with no interventions (Figure 2), i.e., infected nodes are not isolated, and there are no social distancing or vaccination strategies. We study the impact of budgets on the $T_\alpha(\mathbf{S_{pcr}}, \mathbf{S_g})$ under different epidemic scenarios.

5.1 Experimental Setup for the Early Detection

We explore 216 very different outbreak scenarios ($\beta_1 \in \{0.02, 0.03\}$; $\beta_2 \in \{1.2\beta_1, 1.4\beta_1, 1.6\beta_1\}$; partial immunity $\in \{0\%, 25\%, 50\%, 75\%\}$; $D_2 \in \{90, 120, 150$ day$\}$; and importation or seeding location $\in$\{random all over Virginia, an urban county : Fairfax, a rural county: Montgomery\}) for the Virginia contact network ($|V| = 7,688,058, |E| = 371,888,622$). For the disease model calibration, we have used the parameters values form Espinoza et al. (Espinoza et al. 2023). Incubation period is $N(5,1)$ and infectious duration is $exp(1/9day)$. The symptomatic rate is 55% and time to wane is $exp(1/180day)$. For each epidemic scenario, we have proposed 18 budget combinations ($B_{pcr} \in \{10,000, 20,000, 30,000\}$; and $B_g \in \{15, 50, 100, 150, 200, 250\}$). Therefore we have 3888 different budget and epidemic scenarios. For each epidemic scenario, we have 60 stochastic realizations (Bisset et al. 2009; Venkatramanan et al. 2018). Then, NETWORKDETECT ran 20 Monte Carlo simulations on an epidemic realization for each budget scenario. Therefore, all of our results come from the observations of the $4,665,600$ stochastic agent-based simulation of a very large detailed contact network.

5.2 Disease Dynamics and Detection Time

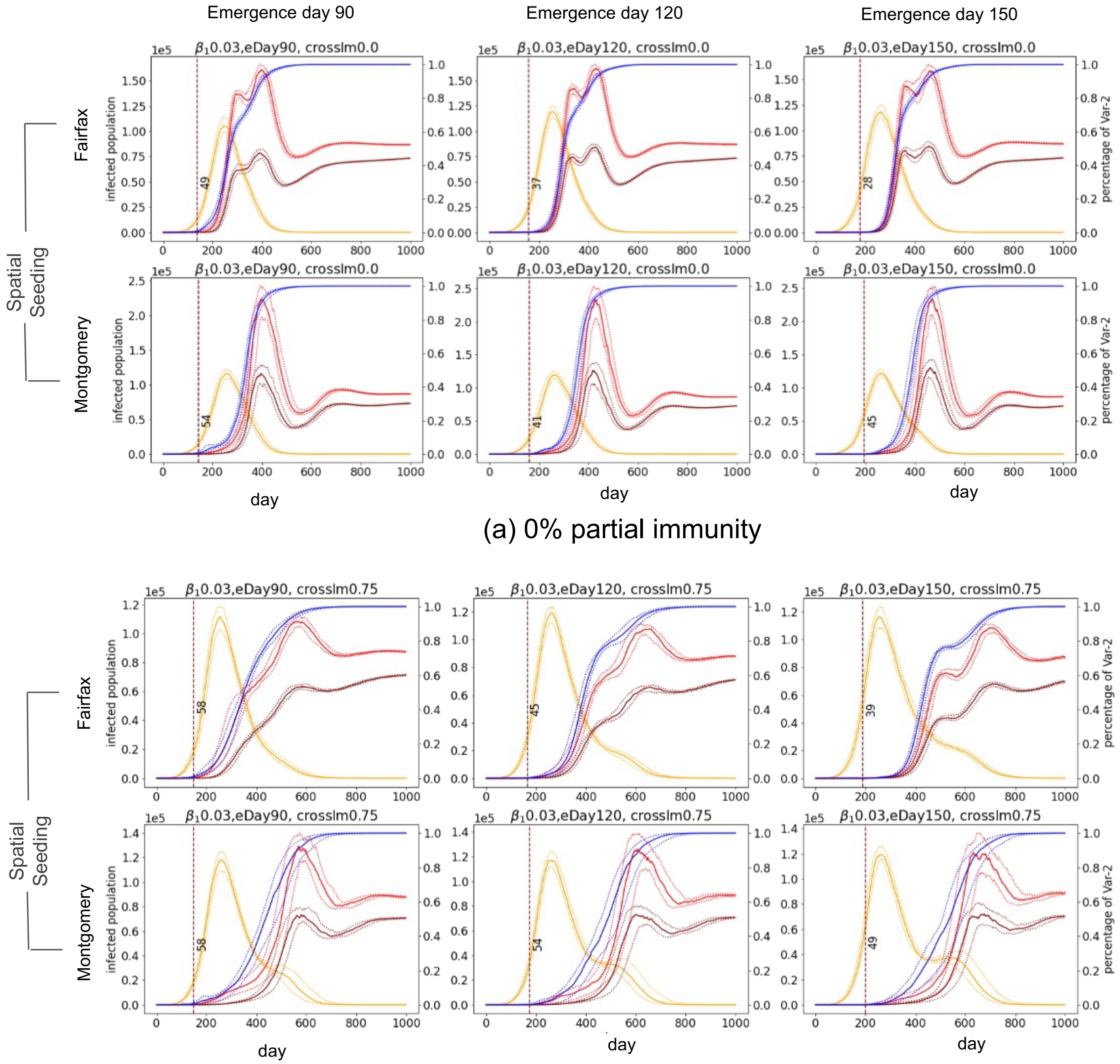
The NETWORKDETECT framework allows us to explore spatio-temporal disease dynamics at any spatial resolution and its corresponding detection time T_{95}. It also allows us to check the impact of reinfection, partial immunity, or waning immunity on the detection time. Figure 3 shows the infected curves for different population and detection time T_{95} day for all over Virginia in different epidemic scenarios. Here, β_1 is 0.03, β_2 is $1.4\beta_1$, B_p is $30,000$ per day, and B_g is 15 per day all over Virginia.

5.3 Impact of Partial Immunity on Detection Time

We explore a wide range of partial immunity and its impact on disease dynamics and detection time. Figure 3(a) shows disease dynamics and detection time when there is no partial immunity or 0% immunity from variant-1 to variant-2. On the other hand, Figure 3(b) describes 75% partial immunity or high immunity from variant-1 to variant-2. Comparison between Figure 3(a) and 3(b) allow us to understand the impact of partial immunity on the new VOC disease dynamics and detection time. As expected, we observe later detection times (higher T95 values) in scenarios with higher partial immunity (Figure 3). This is because partial immunity, acquired from prior infection with variant-1, reduces the pool of susceptible individuals for the emerging variant-2. Consequently, the transmission of variant-2 is slowed, delaying its detection even with the same testing strategies and budgets. This highlights the importance of incorporating immunity levels, whether from natural infection or vaccination, when designing surveillance strategies.

5.4 Budget Optimization for PCR and Genomic Testing

The NETWORKDETECT framework allows us to investigate the importance of different budgets. It also guides budget optimization.
Importance of PCR Budget (B_p): Our analysis reveals a limited impact of voluntary asymptomatic PCR testing on early VOC detection. This is evident from the candidate pool analysis (Figure 4), which shows that the contribution of asymptomatic PCR-positive individuals to the genomic testing pool is negligible compared to the symptomatic population. The red curves (symptomatic population size in the candidate pool) and the black curves (candidate pool size) in Figure 4 almost overlap. The contribution of the



(a) 0% partial immunity

(b) 75% partial immunity

Figure 3: Epidemic curves and T_{95} for different epidemic scenarios. The red vertical lines and associated texts represent the median detection day T_{95}. Each subplot corresponds to a specific emergence day and location. The left y-axis in a subplot shows the variant-1 infected (yellow line), variant-2 infected (red line), and variant-2 reinfected curve (maroon line). The right y-axis shows the percentage of variant-2 among all the infected population (blue line). The x-axis is the time in days. Solid lines represent the median, and dotted lines represent the interquartile range.

asymptomatic population in the candidate pool (blue curves) is minimal, less than 4%, compared to the contribution of the symptomatic population. This finding suggests that, under the tested scenarios and given the limited genomic testing capacity (B_g), prioritizing PCR testing solely for symptomatic individuals might be a more efficient strategy for early VOC detection. However, this conclusion should be further evaluated considering other factors like the prevalence of asymptomatic infection, testing costs, and potential benefits of identifying asymptomatic carriers for isolation and contact tracing.

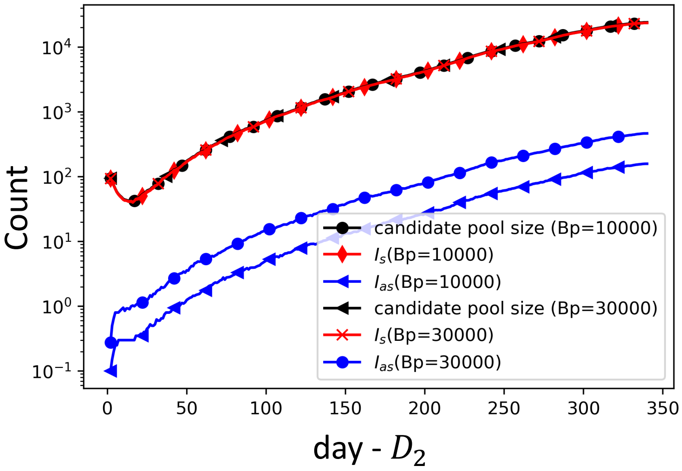
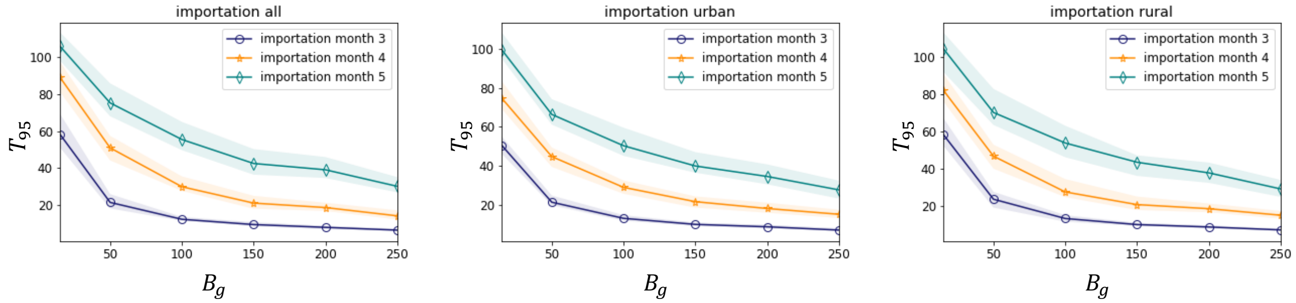


Figure 4: Candidate pool analysis; candidate pool size, symptomatic population (I_s) in the candidate pool, and asymptomatic population (I_{as}) in the candidate pool for two PCR testing budgets, $B_{pcr} \in \{10000\ day^{-1}, 30000\ day^{-1}\}$. Here, $\beta_1 = 0.02$, $D_2 = 90$ days, importation location is all over Virginia, and partial immunity from the previous variant is 50%. The Y-axis is in the log scale.

B_g budget optimization: In the United States, CDC uses genomic surveillance to track emerging SARS-CoV-2 variants that cause COVID-19. Although, genomic sequencing capacity in the United States has increased in the pandemic time (Lambrou et al. 2022), still effective genomic surveillance strategy is needed to track the Variant of Interest (VOI) in the population. Figure 5 presents the T_{95} day for different B_g budget scenarios in the three seeding locations. The analysis of genomic sequencing budget optimization reveals a consistent pattern across different seeding scenarios. Increasing the genomic sequencing budget (B_g) leads to a sharp decrease in the T95 value (faster detection) up to a certain point, often referred to as the "elbow point." However, beyond this elbow point, further budget increases yield diminishing returns in terms of early detection. This suggests that allocating resources to genomic sequencing beyond the elbow point might not be cost-effective for improving early detection. This finding highlights the importance of identifying this optimal budget point to maximize the impact of limited genomic sequencing resources.

6 CONCLUSION

This research develops a NetworkDetect framework to solve a Variant Surveillance problem in a multi-variant environment. It studies Variant Surveillance problem as a function of different disease characteristics, possible importation scenarios, and limited testing budgets for the state of Virginia. The NetworkDetect framework is an agent-based stochastic detail-oriented framework, which is flexible enough to deal with other contact networks and different epidemic scenarios. The agent-based heterogeneous surveillance framework NetworkDetect allows us more flexibility in an experimental setup compared to the homogeneous frameworks (Wohl et al. 2022; Espinoza et al. 2023). In the homogeneous assumptions for population mixing and testing conditions, the probability of detecting a single positive case depends on



(a) (b) (c)

Figure 5: Impact of genomic sequence budget (B_g) per day on the T_{95} day for three importation locations. Here, $\beta_2 = 1.4\beta_1$, and $\beta_1 = 0.03$.

the sampling size and the disease prevalence level. However, from heterogeneous, more realistic agent-based simulation, we found that seeding (or importation) also plays an essential role in the early detection of new variants. Although our framework is general, the results of our framework are sensitive to the underlying network model, which explicitly depends on the population data. In many cases, population data is messy and insufficient.

The NETWORKDETECT framework allows us to evaluate any surveillance strategy at the individual level with history in a multi-variant environment.

ACKNOWLEDGMENTS

This research is supported by University of Virginia Strategic Investment Fund award number SIF160, Virginia Department of Health grant VDH-21-501-0135-1, CDC Pathogen Genomics Centers of Excellence network (PGCoE) grant 6NU50CK000555-03-01, NSF Grants OAC-1916805 (CINES), CCF-1918656 (Expeditions), IIS-1931628, IIS-1955797, and NIH grant R01GM109718.
This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable,world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

Aguilar, J. B., J. S. Faust, L. M. Westafer, and J. B. Gutierrez. 2020. “Investigating the impact of asymptomatic carriers on COVID-19 transmission”. *MedRxiv*.

ATUS 2021. “American Time Use Survey,U.S. BUREAU OF LABOR STATISTICS”. [Online; accessed 15-Dec-2022].

Bai, Y., L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen *et al*. 2020. “Presumed asymptomatic carrier transmission of COVID-19”. *JAMA* 323(14):1406–1407.

Barrett, C., R. Beckman, M. Khan, V. S. A. Kumar, M. Marathe, P. Stretz, *et al*. 2009. “Generation and Analysis of Large Synthetic Social Contact Networks”. In *Proceedings of the Winter Simulation Conference*.

Bisset, K. R., J. Chen, X. Feng, V. A. Kumar and M. V. Marathe. 2009. “EpiFast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems”. In *Proceedings of the 23rd international conference on Supercomputing*, 430–439.

Burki, T. 2021. “Understanding variants of SARS-CoV-2”. *The Lancet* 397(10273):462.

Bushman, M., R. Kahn, B. P. Taylor, M. Lipsitch and W. P. Hanage. 2021. “Population impact of SARS-CoV-2 variants with enhanced transmissibility and/or partial immune escape”. *medRxiv* https://doi.org/10.1101/2021.08.26.21262579.

Chen, J., S. Hoops, *et al.* 2021. “Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness”. *medRxiv* https://doi.org/10.1101/2021.02.04.21251012.

Espinoza, B., A. Adiga, S. Venkatramanan, A. S. Warren, J. Chen, B. L. Lewis, , , , *et al.* 2023. “Coupled models of genomic surveillance and evolving pandemics with applications for timely public health interventions”. *Proceedings of the National Academy of Sciences* 120(48):e2305227120.

Eubank, S., C. Barrett, R. Beckman, K. Bisset, L. Durbeck, C. Kuhlman, , , *et al.* 2010. “Detail in network models of epidemiology: are we there yet?”. *Journal of biological dynamics* 4(5):446–455.

Eubank, S., H. Guclu, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai *et al.* 2004. “Modelling Disease Outbreaks in Realistic Urban Social Networks”. *Nature* 429(6988):180–184.

Google 2022. “COVID-19 Community Mobility Reports - Google”. [Online; accessed 1-Dec-2022].

HERE 2021. [Online; accessed 15-Dec-2022].

Hoops, S., J. Chen, A. Adiga, B. Lewis, H. Mortveit, H. Baek, , , , *et al.* 2021. “High performance agent-based modeling to study realistic contact tracing protocols”. In *2021 Winter Simulation Conference (WSC)*, 1–12. IEEE.

Karim, S. S. A. and Q. A. Karim. 2021. “Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic”. *The Lancet* 398(10317):2126–2128.

Lambrou, A. S., P. Shirk, M. K. Steele, P. Paul, C. R. Paden, B. Cadwell, , , , *et al.* 2022. “Genomic surveillance for SARS-CoV-2 variants: predominance of the Delta (B. 1.617. 2) and omicron (B. 1.1. 529) variants—United States, June 2021–January 2022”. *Morbidity and Mortality Weekly Report* 71(6):206.

LandScan,Oak Ridge National Laboratory 2021. [Online; accessed 15-Dec-2022].

Machi, D., P. Bhattacharya, S. Hoops, J. Chen, H. Mortveit, S. Venkatramanan, , , , *et al.* 2021. “Scalable epidemiological workflows to support covid-19 planning and response”. In *2021 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 639–650. IEEE.

Moon, S. A. and C. M. Scoglio. 2021. “Contact tracing evaluation for COVID-19 transmission in the different movement levels of a rural college town in the USA”. *Scientific reports* 11(1):1–12.

NCES 2021. “National Center for Education Statistics”. [Online; accessed 15-Dec-2022].

NHTS 2021. “The National Household Travel Survey (NHTS),Federal Highway Administration U.S. Department of Transportation”. [Online; accessed 15-Dec-2022].

OpenStreetMap 2021. [Online; accessed 15-Dec-2022].

Pastor-Satorras, R., C. Castellano, P. Van Mieghem, and A. Vespignani. 2015. “Epidemic processes in complex networks”. *Reviews of modern physics* 87(3):925.

Shaman, J. 2020. “An estimation of undetected COVID cases in France”. *Nature* 590:38–39.

Skegg, D., P. Gluckman, G. Boulton, H. Hackmann, S. S. A. Karim, P. Piot *et al.* 2021. “Future scenarios for the COVID-19 pandemic”. *The Lancet* 397(10276):777–778.

Sood, N., P. Simon, P. Ebner, D. Eichner, J. Reynolds, E. Bendavid *et al.* 2020. “Seroprevalence of SARS-CoV-2–specific antibodies among adults in Los Angeles County, California, on April 10-11, 2020”. *JAMA* 323(23):2425–2427.

Venkatramanan, S., B. Lewis, J. Chen, D. Higdon, A. Vullikanti and M. Marathe. 2018. “Using data-driven agent-based models for forecasting emerging infectious diseases”. *Epidemics* 22:43–49.

Wohl, S., E. C. Lee, B. L. DiPrete, and J. Lessler. 2022. “Sample size calculations for variant surveillance in the presence of biological and systematic biases”. *medRxiv*:2021–12.

World Health Organization. “Guidance for surveillance of SARS-CoV-2 variants”.

Zhang, W., J. P. Govindavari, B. D. Davis, S. S. Chen, J. T. Kim, J. Song, , *et al.* 2020. “Analysis of genomic characteristics and transmission routes of patients with confirmed SARS-CoV-2 in Southern California during the early stage of the US COVID-19 pandemic”. *JAMA network open* 3(10):e2024191.

AUTHOR BIOGRAPHIES

Sifat Afroj Moon is a Scientist in the Computational Sciences and Engineering Division at the Oak Ridge National Laboratory. Her research interests include network science, agent-based modeling, high-performance computing, discrete algorithms, and artificial intelligence. moons@ornl.gov

Jiangzhuo Chen is a Research Associate Professor in the Biocomplexity Institute (BII) at the University of Virginia. His research interests are in computational epidemiology, agent-based modeling and simulation, and causal machine learning. chenj@virginia.edu

Baltazar Espinoza is a Research Assistant Professor in the Biocomplexity Institute (BII) at the University of Virginia. His research interests lie at the intersection of mathematical and computational models to study infectious diseases and population dynamics. be8dq@virginia.edu

Bryan Lewis is a Research Associate Professor in the Biocomplexity Institute (BII) at the University of Virginia. His research interests are in public health and epidemiology, epidemiologic modeling, social network construction, and graph measures and dynamic networks. brylew@virginia.edu

Madhav Marathe is a Distinguished Professor in the Biocomplexity Institute (BII) at the University of Virginia with interests in network science, computational epidemiology, AI, foundations of computing, socially coupled system science and high performance computing. marathe@virginia.edu

Joseph Outten is a Software Engineer at Metaform. He is a former student and research assistant of the Biocomplexity Institute (BII) at the University of Virginia. outtenjoseph@gmail.com

Srinivasan Venkatramanan is a Research Assistant Professor in the Biocomplexity Institute (BII) at the University of Virginia, with interests in infectious disease modeling, model calibration and forecasting, simulation analytics, network-based causal inference, and game theory. srini@virginia.edu

Anil Vullikanti is a Professor in the Biocomplexity Institute (BII) and the Department of Computer Science at the University of Virginia. His interests are in network science, and the foundations of AI and machine learning. vsakumar@virginia.edu

Andrew Warren is a Research Assistant Professor in the Biocomplexity Institute (BII) at the University of Virginia. His interests lie in developing and applying algorithms for processing biological data for insight and hypothesis testing using comparative genomics, phylodynamics, experimental analysis, machine learning, data mining, and graph modeling. As a scientist at BII, I work with a trans-disciplinary team of experts in modeling different levels of infectious disease spread for public health and defense research. asw3xp@virginia.edu