Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo, and M. Loper, eds.

HIGH PERFORMANCE AGENT-BASED MODELING TO STUDY REALISTIC CONTACT TRACING PROTOCOLS

Stefan Hoops
Jiangzhuo Chen
Abhijin Adiga
Bryan Lewis
Henning Mortveit
Hannah Baek
Mandy Wilson
Dawen Xie
Samarth Swarup
Srinivasan Venkatramanan

Biocomplexity Institute and Initiative
University of Virginia
Charlottesville, VA 22904, USA

Justin Crow
Elena Diskin
Seth Levine
Helen Tazelaar
Brooke Rossheim
Chris Ghaemmaghami
Rebecca Early

Virginia Department of Health
109 Gov. Street
Richmond, VA 23219, USA

Chris Barrett
Madhav V. Marathe

Biocomplexity Institute and Initiative
and Computer Science Department
University of Virginia
Charlottesville, VA 22904, USA

Carter Price

RAND Corporation
1200 South Hayes Street
Arlington, VA 22202, USA

ABSTRACT

Contact tracing (CT) is an important and effective intervention strategy for controlling an epidemic. Its role becomes critical when pharmaceutical interventions are unavailable. CT is resource intensive, and multiple protocols are possible, therefore the ability to evaluate strategies is important. We describe a high-performance, agent-based simulation model for studying CT during an ongoing pandemic. This work was motivated by the COVID-19 pandemic, however framework and design are generic and can be applied in other settings.

This work extends our HPC-oriented ABM framework EpiHiper to efficiently represent contact tracing. The main contributions are: (*i*) Extension of EpiHiper to represent realistic CT processes. (*ii*) Realistic case study using the VA network motivated by our collaboration with the Virginia Department of Health.

1 INTRODUCTION

We study the problem of contact tracing individuals as an effective public health intervention during an epidemic. Contact tracing is a well-studied and effective method for controlling epidemics caused by infectious diseases (Eames and Keeling 2003). Contact tracing as a mitigation strategy was adopted at an unprecedented scale early on in the COVID-19 pandemic, and all US states have spent considerable

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

resources toward supporting it. However, contact tracing at such a large scale is novel and had not been carried out prior to this pandemic. Implementation of contact tracing at this massive scale is expensive and poses significant logistical challenges, and the efficacy of contact tracing protocols is poorly understood at this scale. Even the effects of basic parameters, such as the rate of processing contacts, is not fully understood, especially in the context of other non-pharmaceutical interventions like social and physical distancing and mask-wearing. To address these issues, we use a parallel agent-based simulation model and highly detailed digital twin of the population of the US state of Virginia to assess contact tracing protocols.

Informally, *case investigation and contact tracing* followed by *case isolation and self-quarantining of contacts* (Kretzschmar et al. 2020) is an intervention strategy wherein an index case (a person) with a confirmed infection is requested to provide information about individuals with whom they have come in contact, a process referred to as *contact elicitation*. What constitutes a contact, and the nature of that contact, will, of course, depend on the particular disease. *Contact tracers* reach out to the reported contacts, inform them about their possible exposure, and request that they self-quarantine for a certain number of days, depending on the incubation period of the disease. This can effectively break the chain of disease transmission, and prevent further spread of the virus in the community (See https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/overview.html for details). The Commonwealth of Virginia implemented a contact tracing protocol early in the pandemic and continues to effectively contact trace to reduce disease incidence.

Our contributions. We use our recently developed ABM platform for epidemiological analysis, called EpiHiper (Chen et al. 2019), together with a detailed model of a social contact network for the Commonwealth of Virginia (called VA-SN), to study the effectiveness of contact tracing protocols. The ABM simulates disease propagation and complex sets of interventions, including various non-pharmaceutical interventions and contact tracing protocols. Our main contributions include the following:

(*i*) We extend the EpiHiper modeling environment to represent the contact tracing process; the implementation is done so that EpiHiper can continue to run efficiently on parallel machines. The implementation of contact tracing protocols in EpiHiper is generic and flexible, allowing nodes to be specified for tracing and quarantine in various ways, such as using demographic information and processing rates to identify candidate nodes.

(*ii*) We carry out extensive performance analyses of the realistic contact tracing protocols, focusing on the high performance computational resources needed using VA-SN. We focus specifically on how the computational costs vary when certain real-world constraints are applied, such as compliance, number of contact tracers, and transmissibility of the disease.

(*iii*) We use the ABM to carry out a realistic case study for the state of Virginia using data and parameters from the ongoing COVID-19 pandemic. This case study is based on the ongoing joint collaboration between the Virginia Department of Health and the University of Virginia. We consider a scenario that is currently unfolding, and study how contact tracing and vaccinations together can *bend* the epidemic curve. Our results point out the importance of continued contact tracing to reach low daily incidence numbers.

(*iv*) Finally, in order to obtain insights into the impact of various contact tracing strategies, we carry out detailed network theoretical analyses of the temporal contact network that results as contact tracing methods are implemented. The analyses reveal how the temporal network evolves as a result of contact tracing, and provides real-world policy insights.

Analytics for Public Health policy. The work reported here represents a six month collaborative effort between the University of Virginia, RAND, and the Virginia Department of Health. It aims to provide analytical support to the department as health officials respond to the COVID-19 pandemic; the questions and the models were informed through this process. It builds on the team's ongoing work in developing scalable agent-based models to support pandemic planning.

2 RELATED WORK

Manual contact tracing is a widely used strategy that has been immensely successful in controlling past outbreaks; see (Keeling et al. 2020; Eames and Keeling 2003; Eames 2007; Klinkenberg et al. 2006; Kiss et al. 2005; Kiss et al. 2008) for discussions on contact tracing, its effectiveness, and mathematical models used to study contact tracing in networks. The role and effectiveness of contact tracing during the 2014 Ebola, SARS, and MERS outbreaks have also been investigated; see (Eames and Keeling 2003; Liu et al. 2015; Kwok et al. 2019; Olu et al. 2016; Swanson et al. 2018). Digital contact tracing has emerged as a potentially powerful technology to control outbreaks, especially during the recent COVID-19 pandemic: see (Ferretti et al. 2020; Ahmed et al. 2020; Anglemyer et al. 2020; Salathé et al. 2020; Lorch et al. 2020). Analysis of digital contact tracing has shown that its effectiveness depends on the quality of the device, trust within the population, uptake of the app, and the compliance of its users (Chan et al. 2020; Anglemyer et al. 2020; Giabbanelli and Li 2020; Salathé et al. 2020; Barrat et al. 2020).

Recently there have been a few papers related to the use of agent-based models to analyze contact tracing protocols. Papers most related to our work include: (Ferretti et al. 2020; Abueg et al. 2020; Kerr et al. 2020; Aleta et al. 2020); each of these papers use detailed agent-based models in evaluating various contact tracing policies. Our work differs from these papers in the following ways: (*i*) the agent-based models we have used consider a much more detailed representation of the underlying social contact network; (*ii*) the spatial scale and size reflects an entire state, and therefore allows us to capture the spatial and temporal heterogeneity of the underlying network and infection process, and (*iii*) our representation includes realistic non-pharmaceutical interventions that dynamically modify the network as the epidemic progresses.

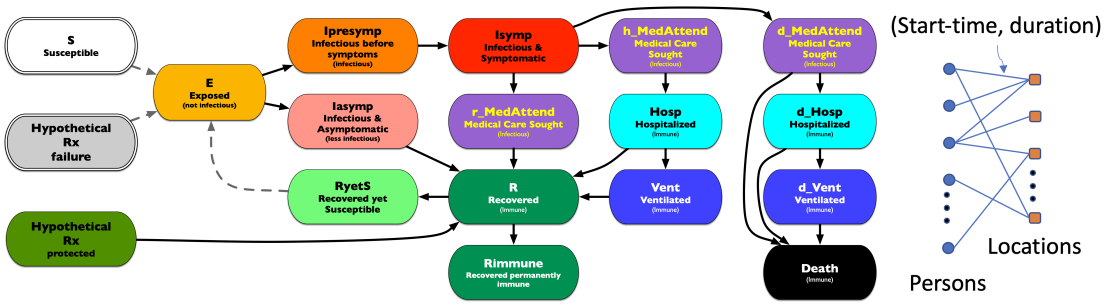
Our work is closest to two earlier papers (Kerr et al. 2020; Abueg et al. 2020). The paper by Kerr et al. (Kerr et al. 2020) uses the *Covasim* ABM platform (Kerr et al. 2020) to simulate the effect of contact tracing in Seattle WA. The work of Abueg et al. (2020) uses their ABM platform called OpenABM to simulate the contact tracing protocols in three counties around Seattle. There are some differences between our work and their work.

First, the social contact network developed by (Kerr et al. 2020; Abueg et al. 2020) is substantially simpler – they use age-mixing matrices to develop social contact networks. This approximation is possibly adequate when studying a large metropolitan area such as Seattle, but it does not capture contacts based on specific locations and contacts based on individual mobility. The result of this approximation for the problem at hand is two-fold: (*i*) social contact networks do not display local to global structures wherein nodes have high contact in and around their home, and progressively fewer contacts with individuals farther off; (*ii*) the network structure impacts contact tracing efficacy as well as the performance of parallel ABMs. When applying methods based purely on age-mixing, one ignores the structure of population spread within a state and its interconnectivity.

Second, the disease transmission models used in EpiHiper are far more detailed. These models are age-stratified, and, hence, the overall impact from the models are likely to be somewhat different. Third, the size and the geographical scale of the network considered in our paper is substantially different than the networks considered in (Kerr et al. 2020; Abueg et al. 2020). Finally, the paper by (Kerr et al. 2020) begins to do network analysis to assess the impact of the contact tracing methods. Our paper is motivated by their work, but extends the analysis substantially.

3 MODELS AND METHODS

The current work builds on two important ideas and data sets. First, we use a digital twin of Virginia, which is described further below. Second, the contact tracing module is built on top of the EpiHiper agent-based simulation platform. The ABM simulates disease propagation and a complex set of interventions, including various non-pharmaceutical interventions and vaccine allocation schemes. Each contact between an infective and a susceptible individual may lead to the transmission of the disease based on contact time and transmissibility. Information on EpiHiper can be found in (Chen et al. 2019; Chen et al. 2021).



(a) Disease Model (b) Contact Network

Figure 1: (a) The COVID-19 disease model. This model is implemented for 5 age groups with transmissions being possible between all ages. (b) An illustration of the contact network.

3.1 The Virginia Social Contact Network Model

The VA-SN model is derived from a digital twin of Virginia, which was previously constructed by synthesizing multiple data sets, including demographic data from the American Community Survey, weekly activity data from the National Household Travel Survey, built environment data from the Microsoft Building Database, HERE, and the public school data from the National Center for Education Statistics. Details of the process can be found in prior work (Adiga et al. 2015; Chen et al. 2021).

Each individual in the model has a demographically-appropriate activity sequence as well as activity locations assigned to it. This results in a person-location bipartite graph as illustrated in Figure 1b, where each edge is annotated with the start time and duration of the activity. When two people are at the same location for an overlapping duration, the duration and the disease transmissibility will give rise to a propensity as in the direct stochastic algorithm (Gillespie 1976). This induces a person-person co-location graph, which we use for the simulations with EpiHiper. A detailed analysis of the network structures generated by this process can be found in (Chen et al. 2021). Each interaction takes place at a particular location, so each edge e, in the network has a corresponding location denoted by $Location(e)$. The locations are grouped into types: $\{home, work, school, college, shopping, religion, other\}$. We now describe how contact tracing is implemented in the simulation using this modeled social contact network.

3.2 Formal Description of Contact Tracing in EpiHiper

Newly Symptomatic Individuals. Given the set of all vertices V, and given that a newly symptomatic individual v is in a symptomatic infectious state I_l, of which the COVID-19 disease model has several (see Section 4), we can define the set of newly symptomatic individuals V_{new} by:

$$V_{new} = \{v \in V \,|\, \text{healthstate}(v,t) \in \{I_1, I_2, ...\} \cap \text{healthstate}(v,t-1) \notin \{I_1, I_2, ...\}\} \tag{1}$$

Home Isolation or Quarantine. During home isolation or quarantine (HI), an individual is only in contact with household members, i.e., all edges with nodes not within the household will be eliminated for the duration of HI, which, for COVID-19, is set to 14 days; the number of edges between the individual and household members may also be reduced to deter spread within the household. Individuals who participate in HI may be selected in different ways, e.g., a percentage of symptomatic individuals (voluntary home isolation) or nodes based on contact tracing information (voluntary home quarantine). Independent from the process of the selection of individuals, we denote the set V_{HI} of individuals choosing to starting to participate in HI as, $V_{HI} = \{v \in V \,|\, \text{HI}(v) = \texttt{starting}\}$ Let E be the set of all edges. We define the set of non-household contacts, $E_{nonhome}$. Since the location of a contact does not change during simulation, i.e., the set is time independent, it is only evaluated once. $E_{nonhome} = \{e \in E \,|\, \text{Location}(e) \neq \texttt{home}\}$. And

thus the set of edges E_{HI} affected by HI is given by,

$$E_{HI} = E_{nonhome} \cap \{e \in E \,|\, \text{Vertex}(e) \in V_{HI}\}, \tag{2}$$

where $Vertex(e)$ refers to either of the nodes connected through edge e. All edges $e \in E_{HI}$ will be removed for 14 days by deactivating them immediately and scheduling an activation after 14 days.

Voluntary Home Isolation. Symptomatic individuals will voluntarily participate in home isolation. They make this decision on the first day of symptoms, i.e., when they are in the set V_{new}. We represent this decision by sampling a customizable percentage P of V_{new}. These sampled individuals will be added to the set V_{HI}. Their contacts with their household members will be reduced duration the isolation.

Contact Tracing with Distance 1. Contact tracing will encourage all individuals in contact with a newly symptomatic individual to participate in self-quarantine. Not all individuals will participate in this process. We select candidates of V_{new} based on the government's capacity C to perform contact tracing and the willingness W of the individuals to participate. For this selection, we perform 2 sampling steps: i) select at most C individuals from V_{new}; ii) sample W of the result. This resulting set of participants is V_{CT}.

Contact tracing is done along active edges E_{active}, i.e., edges through which a transmission may occur. Therefore the set of all edges with at least one node in V_{CT} is given by, $E_{CT} = E_{active} \cap \{e \in E \,|\, \text{Vertex}(e) \in V_{CT}\}$. All nodes touched by edges in the above set E_{CT} which are not participants are contacted:

$$V_{contacted} = \{v \in V \,|\, \text{Edge}(v) \in E_{CT}\} \cap \overline{V_{CT}}, \tag{3}$$

where $\overline{V_{CT}}$ is the set complement of V_{CT}. Of the contacted nodes, a percentage $\hat{P}$ will agree to participate in home quarantine and are put in the set V_{HI}. Unlike the people participating in voluntary home isolation, their contacts with their household members are not reduced.

4 EXPERIMENT SETTINGS AND DESIGN

The simulation's input parameters specify the population demographics and contact network, COVID-19 disease model, initial configuration S_0, contact tracing and quarantine, and other non-pharmaceutical interventions (NPIs). The simulation output is a dendogram: a directed graph that tells us who infects whom and on what day. From the output data, we can compute many epidemiological measures, such as daily new infections, cumulative infections, prevalence in each age group, total hospitalizations and deaths, as well as many others.

Disease model. The disease model is the *best guess* version of "COVID-19 Pandemic Planning Scenarios" prepared by the US Centers for Disease Control and Prevention (CDC) SARS-CoV-2 Modeling Team (Centers for Disease Control and Prevention 2020) and has been used by multiple researchers in their papers. It is an SEIR model where state transitions follow the parameters as defined in the document. The disease states and transition paths are shown in Figure 1a. Individuals of different age groups have different infectivity and susceptibility; dwell time distributions and state transition probability distributions are stratified by the following age groups: preschool (0-4 years), students (5-17), adults (18-49), older adults (50-64) and seniors (65+).

Initializations. The simulations are initialized at the county level by age group using the detailed data of confirmed cases from The New York Times (2020). The initialization specifies the health state of each individual. In this study we are interested in how contact tracing enhances effectiveness of vaccination in terms of reducing infections/hospitalizations/deaths. To this end, we initialize the simulations using the confirmed cases data up to Mar. 1, 2021, denoted by t, as follows. From state-level age distributions in cases up to t, we derive an age group distribution $\mathscr{D}$ that is applied at the county level to stratify county-level data. We assume a case ascertainment ratio of 3, i.e., for each confirmed case, there are 3 actual cases, including symptomatic as well as asymptomatic ones. Based on county-level cumulative confirmed cases up to $t-17$, we derive the number of prior infections in each county by scaling the cumulative number by

the case ascertainment ratio, then compute the number of prior infections in each age group of this county using the distribution $\mathscr{D}$. We randomly choose individuals in each age group in each county and set their health states to `recovered` to reflect that they have already been infected. Based on county-level daily confirmed cases from $t-16$ to t, we derive the number of individuals that are infected each day by scaling these cases by the case ascertainment ratio; compute the number of infections in each age group using the distribution $\mathscr{D}$; and seed the simulation by setting randomly chosen individuals to `exposed` by day in each age group of each county.

Non-pharmaceutical interventions. We consider four NPIs: (*i*) *Infectivity reduction (IR)*. Infectivity is universally reduced by 60% to account for preventive behavior, e.g., mask wearing and hand washing. (*ii*) *Generic social distancing (GSD)*. A fraction (25%) of the population chooses to reduce non-essential (shopping, religion, and other) activities. (*iii*) *Virtual learning (VL)*. A fraction (50%) of K-12 students choose virtual learning. (*iv*) *Voluntary home isolation of symptomatic cases (VHI)*. With probability 75%, a symptomatic person chooses to stay home for 14 days, reducing the weights on household contacts by 50%. For this person, all outside contacts are disabled and at-home contacts are reduced by 50% temporarily during these 14 days. Note that the choice of these parameters is motivated by the current situation but given that the ground reality changes, these are representative numbers.

Model calibration. We calibrate our disease model so that $R_{\text{effective}}$ at the beginning of our simulations is 0.9, which is the estimate based on the confirmed case data around Mar 1, 2021 for Virginia (Abbott et al. 2020). We increase infectivity of the disease by 2% for every week in our simulation, to model the gradual spread of more infectious new variants in the population and the resulting increase of overall transmission risks.

Vaccinations. We assume that the Pfizer-BioNTech and Moderna vaccines are administered to the Virginia population such that 140K people are fully vaccinated every week. We also consider prior vaccinations before the starting date of the simulations: by Feb. 2021, 150K people have been fully vaccinated and they are seniors (65+). For vaccine hesitancy, we assume that 70% of people are willing to get vaccinated. For vaccination prioritization, we assume that vaccines are given first to seniors (65+), then to older adults (50-64), then to adults (18-49). Vaccines protect anyone who is fully vaccinated (two doses) by reducing the probability of infection by 95% and reducing the probability of hospitalization and death by 50%. This is lower than the reductions reported in e.g. (Moghadas et al. 2021) so the parameterization can be taken as a lower bound.

4.1 Contact tracing: modeling and parameterization

The contact tracing process begins with the report of someone who has tested positive for SARS-CoV-2 or is diagnosed with COVID-19. Once an individual, called *index case*, is identified as meeting the probable or confirmed case definition (https://wwwn.cdc.gov/nndss/conditions/coronavirus-disease-2019-covid-19/case-definition/2020/08/05/), they are added to the index case queue. A case investigator chooses (or is assigned) index cases from the queue and interviews them to obtain contact information for their close contacts. If the index case provides information about whom they might have exposed while infectious, those are added to the contact queue. Another contact tracer is assigned contact cases from the contact queue, notifies them about their contacts with the index cases, and asks them to quarantine themselves at home for a certain number of days.

Due to limited human resources, there are a fixed number of employees; they are divided into a team of *case investigators* and another team of *contact tracers*, according to their different skill sets. Each case investigator has a capacity of a number of index cases processed per day, and each contact tracer can process a certain number of contacts per day. Depending on the size of the two teams, we can have different index case processing rates (IPR) and contact processing rates (CPR).

In the baseline scenario, we assume that among all confirmed cases, up to 1500 can be investigated every day and provide information about their close contacts, including household members and the people

they meet regularly in schools, colleges, and workplaces. The contact processing rate is 2500 cases per day; 50% of notified contacts comply with the voluntary quarantine and will start quarantining about 7 days after index case exposure. The self-quarantine lasts for 14 days, during which they do not have interactions with anyone outside of their households.

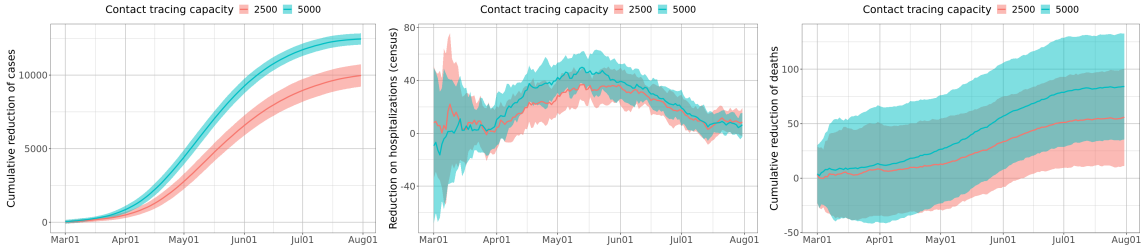
4.2 Experiment Design

In the experiment, we consider **three scenarios**: (*i*) No contact tracing; (*ii*) Baseline: contact processing rate is 2500 per day. This is the average rate of contact tracing in Virginia since August 2020, and (*iii*) More contact tracing: contact processing rate is 5000 per day. The simulations run from Mar. 1, 2021 to Jul. 31, 2021. Since the simulations are stochastic, each simulation is repeated for 20 replicates, and distributions of the measures are computed. The figures presented in Section 5 are all based on data from 20 replicates. The curves show an uncertainty of one standard deviation above and below the mean.

5 ANALYSIS AND RESULTS

5.1 Key Findings

(1) Contact tracing helps further reduce cases/hospitalizations/deaths even with ongoing vaccinations. As shown in Figure 2, even with ongoing vaccinations, contact tracing further reduces the number of confirmed cases, hospitalizations, and deaths. In Virginia alone over five months time, contact tracing can reduce 9K–10.7K cases with a daily tracing capacity of 2500 contacts. If the tracing capacity can be increased to 5000 per day, then the reduction can reach 12K–13K. Reduction of hospitalizations due to contact tracing can be about 40/day (daily capacity 2500) or 50/day (daily capacity 5000) at the pandemic peak. For deaths, contact tracing can reduce mortality in Virginia in five months by 50 (capacity 2500) or 80 (capacity 5000).



(a) Reduction of cases (b) Reduction of hospitalizations (c) Reduction of deaths

Figure 2: Contact tracing reduces: (a) the number of confirmed cases; (b) the number of hospitalizations; and (c) the number of deaths. In (a), cumulatively over five months, contact tracing can reduce confirmed cases by an estimated 10K when contract tracers have the capacity for processing 2500 contacts per day, and by 12K with a daily capacity of 5000. In (b), contact tracing with a daily capacity 2500 can reduce hospitalizations by almost 40 cases, and, with a daily capacity of 5000, can reduce hospitalizations by 50 at peak. In (c), cumulatively over five months, contact tracing can reduce deaths by around 50 or 80 with a daily capacity of 2500 or 5000, respectively.

In Figure 2a, we find that while vaccines are given to people of 18 years or older, contact tracing is able to reduce cases in school-aged children the most. This is because vaccines do not directly protect school-aged children who have the highest degrees in the network. So the marginal protection they receive from contact tracing is larger than the other age groups. The reduction of cases in adults (18–49) is also significant.

Figure 3 shows the percentage reduction of cases at the county and independent city levels in Virginia. It shows that the relative impact of contact tracing is higher in north and east counties of Virginia, which have higher population densities.

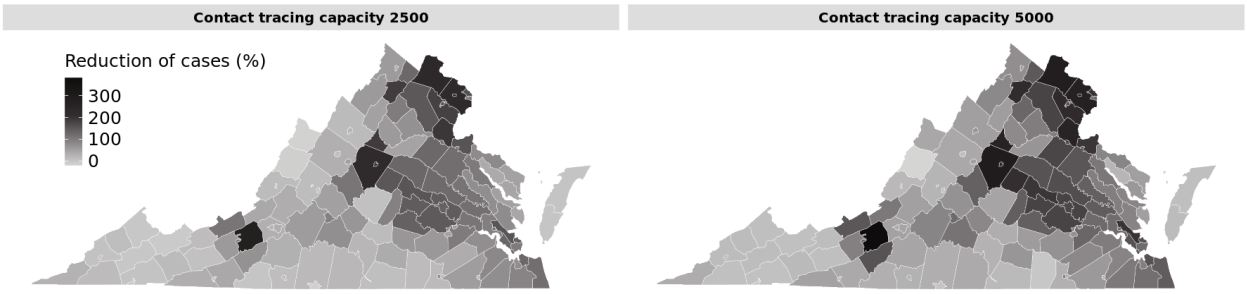
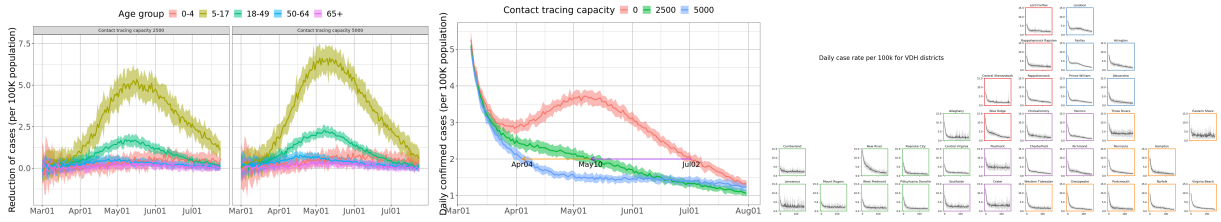


Figure 3: The reduction of cases,relative to the cases in the *no contact tracing* scenario, varies across the different counties of Virginia.



(a) Reduction in Age Groups (b) Daily confirmed Cases (c) Case Rate per District

Figure 4: (a) Contact tracing reduces cases in school-aged children the most. It also reduces cases in the adult age group (18–49) by as many as 2.5 per 100K population. (b) Contact tracing can help us to reach a target rate of 2 daily confirmed cases per 100K population two months or even three months earlier, depending on tracing capacity. (c) Figure showing daily case rate per 100K broken down by districts.

(2) Contact tracing helps us reach the targeted threshold much earlier. From Figure 4b, we find that, without contact tracing, confirmed cases will go up from April to May 2021, even with ongoing vaccinations, due to more infectious new variants spreading through the population. Cases will start decreasing in May and will eventually reach 2/100K in early July 2021. Suppose 2/100K is a targeted threshold for decision-making regarding public health, social, and economic policies; this target can be reached almost two months earlier if contact tracing is in place with a daily capacity of 2500. If the daily capacity can be increased to 5000, then the target can be reached in early April, three months earlier compared to the scenario where no contact tracing takes place.

(3) Queue size and its impact. In Figure 5, we show how the number of people in the contact queue and the number of people in quarantine change over time under the two scenarios with different contact tracing capacities. As contacts are collected from index cases, they are added to the contact queue. They are removed from the queue as they are processed and notified about their exposure, or if they are still not processed after three days. Since the number of confirmed cases is lower than the case investigation capacity and is decreasing over time, the contact queue size will continue to decrease as well. With a larger tracing capacity, the queue size will decrease faster. When we look at the number of people in quarantine, however, we find that it is larger in the *more contact tracing* scenario (5000 capacity) than in the *baseline* scenario (capacity 2500), at the beginning when the capacity is saturated. In early April, in the *more contact tracing* scenario, cases become so low that the contacts processed daily are lower than the capacity. More contacts ending quarantine than those starting quarantine leads to a reduction in the number of people in quarantine. This starts to happen to the *baseline* scenario much later, in mid-June, at which time the number of people in quarantine are actually higher than those in quarantine in the *more contact tracing* scenario. This implies that **tracing more contacts earlier may potentially lead to fewer people in quarantine later on**.

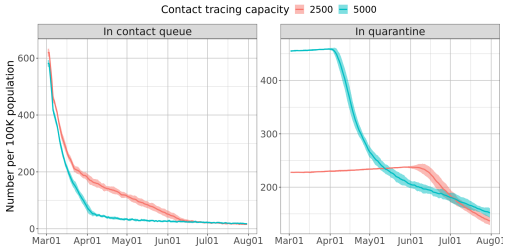


Figure 5: The number of people in the contact queue and the number of those in quarantine, with different tracing capacities. The numbers are normalized as per 100K population.

5.2 Network Analytics to Explain Contact Tracing

The contact network undergoes structural changes at every time step in response to the disease dynamics. Some of these changes can be attributed to the progression of the disease itself, and others to the intervention responses in place. In order to further our understanding on the effect of contact tracing and subsequent quarantining on the disease progression, we measured various graph properties on a sequence of temporal snapshots of the network to study how the network changes structurally. The properties that we considered include average degree, spectral radius, clustering coefficient, k-core, and diameter. These measures have been typically used for structural and dynamical characterization of real-world networks, and are frequently reported in the literature, e.g. (Cattuto et al. 2010; Génois and Barrat 2018)). Suppose $G(V,E)$ is the population network at time 0 of the simulation instance. The graph $G_t(V_t,E_t)$ corresponds to a subgraph of G defined based on the state of the simulation instance at time t. The node set V_t corresponds to individuals whose health state is either susceptible, exposed or infectious, the last of which has several variants. The graph G_t is the subgraph of G induced by V_t. We considered a set of eight snapshots spaced 15 simulation steps apart over the entire simulation time period to study the long term changes in the network. These sets were created for each scenario (CT 0, 2500, and 5000).

The results for one representative simulation instance are in Figure 6. We note that the network is shrinking gradually with removal of nodes due to individuals moving out of the infectious states. However, we do not see significant change in these properties with increase in contact tracing. This can be attributed to the nature and extent of this intervention. Firstly, the removal of edges corresponding to the contact traced individuals is temporary. These edges are reintroduced after a number of time steps. Also, the number of individuals quarantining as a result of contact tracing is small relative to the size of the network. Therefore, at any time step, the structural changes that can be attributed to contact tracing is minimal.

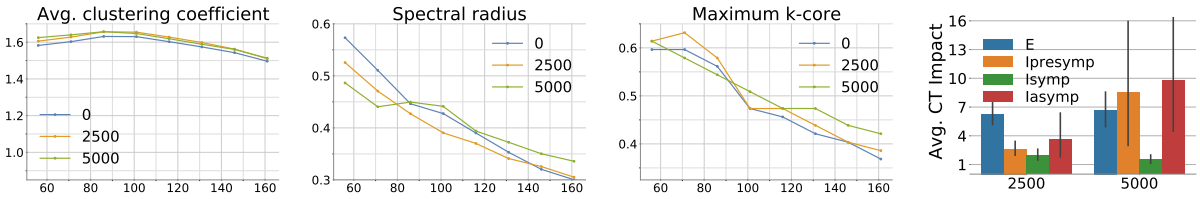


Figure 6: The first three plots correspond to structural properties of the network snapshots for one representative simulation instance. The x-axis corresponds to time. The y-axis corresponds to the property value relative to the network at time 0. The last plot corresponds to impact of contact tracing by comparing the transmission trees of contact traced individuals v with non-contact traced individuals v'. The y-axis is the average of $|T_{v'}|/|T_v|$. The x-axis corresponds to the different contact tracing scenarios.

Since contact tracing targets nodes in the infectious states, it is effective in mitigating the disease spread with few structural changes. To measure its impact on the infected population, we traced the transmission trees of sampled contact traced individuals and compared them with the transmission trees of similar

individuals who were never contact traced in the simulation instance. For a network snapshot G_t, we sampled 100 individuals from the set of individuals who were either exposed or infectious (symptomatic, asymptomatic, presymptomatic) and contact traced in the time window $[t+1, t+10]$. For each sampled individual v, we chose 10 individuals who were in the same state as v at time t and have the same degree as v in G_t, but were never contact traced in the simulation instance. For each considered individual u, we constructed the transmission tree with T_u as the root node. For each pair (v, v') of contact traced individual and corresponding non-contact traced individual, we computed the ratio $|T_{v'}|/|T_v|$. The average value of this impact is plotted in Figure 6 (last plot) for different states and different contact tracing scenarios ($t = 56$). We observe that there is high value in contact tracing and quarantining exposed, asymptomatic and presymptomatic individuals. Since symptomatic individuals are very likely isolated (to some extent) from the rest of the network regardless of whether they are contact traced or not, the effect is minimal.

6 CONCLUSIONS AND LIMITATIONS

We presented a scalable agent-based modeling environment to study realistic contact tracing protocols. Our work represents a six month long collaborative effort between academic, policy and public health groups. The resulting work was thus able to capture some of the realistic constraints faced by decision makers.

Our analysis suggests that: (*i*) increasing the contact early during the pandemic might be more impactful, (*ii*) although contact tracing does not result in significant changes to macroscopic structural properties of the network, it does have significant impact on disease dynamics due to targeted intervention; (*iii*) spatial heterogeneity of cases should be taken into account while allocating limited resources.

Acknowledgments. This work was partially supported by NIH Grant 1R01GM109718, NSF BIG DATA Grant IIS-1633028, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, University of Virginia Strategic Investment Fund award number SIF160, Virginia Dept. of Health Grant and DTRA under Contract No. HDTRA1-19-D-0007.

REFERENCES

Abbott, S., J. Hellewell, R. N. Thompson, K. Sherratt, H. P. Gibbs, N. I. Bosse et al. 2020. "Temporal variation in transmission during the COVID-19 outbreak". https://epiforecasts.io/covid/, accessed 1st March 2021.

Abueg, M., R. Hinch, N. Wu, L. Liu, W. J. Probert, A. Wu, P. Eastham, Y. Shafi, M. Rosencrantz, M. Dikovsky et al. 2020. "Modeling the combined effect of digital exposure notification and non-pharmaceutical interventions on the COVID-19 epidemic in Washington state". *medRxiv*.

Adiga, A., A. Agashe, S. Arifuzzaman et al. 2015. "Generating a Synthetic Population of the United States". Technical Report NDSSL 15-009, Network Dynamics and Simulation Science Laboratory.

Ahmed, N., R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha. 2020. "A Survey of COVID-19 Contact Tracing Apps". *IEEE Access* 8:134577–134601.

Aleta, A., D. Martin-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler et al. 2020. "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19". *Nature Human Behaviour* 4(9):964–971.

Anglemyer, A., T. H. Moore, L. Parker, T. Chambers, A. Grady, K. Chiu, M. Parry, M. Wilczynska, E. Flemyng, and L. Bero. 2020. "Digital contact tracing technologies in epidemics: a rapid review". *Cochrane Database of Systematic Reviews* 8.

Barrat, A., C. Cattuto, M. Kivelä, S. Lehmann, and J. Saramäki. 2020. "Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data". *medRxiv*.

Cattuto, C., W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. 2010. "Dynamics of person-to-person interactions from distributed RFID sensor networks". *PloS one* 5(7):e11596.

Centers for Disease Control and Prevention 2020. "COVID-19 Pandemic Planning Scenarios". https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-h.pdf. accessed 15th January 2021.

Chan, J., S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, S. Singanamalla, J. Sunshine et al. 2020. "PACT: Privacy Sensitive Protocols and Mechanisms for Mobile Contact Tracing". *arXiv preprint arXiv:2004.03544*.

Chen, J., S. Hoops, B. L. Lewis, H. S. Mortveit, S. Venkatramanan, and A. Wilson. 2019. "EpiHiper: Modeling and Implementation". Technical report, NSSAC Technical Report Series: No. 2019–003.

Chen, J., S. Hoops, A. Marathe, H. Mortveit, B. Lewis, S. Venkatramanan, A. Haddadan, P. Bhattacharya, A. Adiga, A. Vullikanti, A. Srinivasan, M. L. Wilson, G. Ehrlich, M. Fenster, S. Eubank, C. Barrett, and M. Marathe. 2021. "Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness". *medRxiv preprint:2021.02.04.21251012*.

Eames, K. T. 2007. "Contact tracing strategies in heterogeneous populations". *Epidemiology & Infection* 135(3):443–454.

Eames, K. T., and M. J. Keeling. 2003. "Contact tracing and disease control". *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270(1533):2565–2571.

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser. 2020. "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing". *Science*.

Génois, M., and A. Barrat. 2018. "Can co-location be used as a proxy for face-to-face contacts?". *EPJ Data Science* 7(1):11.

Giabbanelli, P. J., and J. Li. 2020. "Identifying Synergistic Interventions to Address COVID-19 Using a Large Scale Agent-Based Model". *medRxiv*.

Gillespie, D. T. 1976, Dec. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". *Journal of Computational Physics* 22(4):403–434.

Keeling, M. J., T. D. Hollingsworth, and J. M. Read. 2020. "Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19)". *J Epidemiol Community Health* 74(10):861–866.

Kerr, C. C., D. Mistry, R. M. Stuart, K. Rosenfeld, G. R. Hart, R. C. Nunez, P. Selvaraj, J. A. Cohen, R. G. Abeysuriya, L. George et al. 2020. "Controlling COVID-19 via test-trace-quarantine". *medRxiv*.

Kerr, C. C., R. M. Stuart, D. Mistry, R. G. Abeysuriya, G. Hart, K. Rosenfeld, P. Selvaraj, R. C. Nunez, B. Hagedorn, L. George et al. 2020. "Covasim: an agent-based model of COVID-19 dynamics and interventions". *medRxiv*.

Kiss, I. Z., D. M. Green, and R. R. Kao. 2005, Jul. "Disease contact tracing in random and clustered networks". *Proceedings. Biological sciences* 272(1570):1407–1414. 16006334[pmid].

Kiss, I. Z., D. M. Green, and R. R. Kao. 2008. "The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing". *Journal of the Royal Society Interface* 5(24):791–799.

Klinkenberg, D., C. Fraser, and H. Heesterbeek. 2006. "The effectiveness of contact tracing in emerging epidemics". *PloS one* 1(1):e12.

Kretzschmar, M. E., G. Rozhnova, M. C. J. Bootsma, M. van Boven, J. H. H. M. van de Wijgert, and M. J. M. Bonten. 2020. "Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study". *The Lancet Public Health* 5(8):e452 – e459.

Kwok, K. O., A. Tang, V. W. Wei, W. H. Park, E. K. Yeoh, and S. Riley. 2019. "Epidemic models of contact tracing: Systematic review of transmission studies of severe acute respiratory syndrome and middle east respiratory syndrome". *Computational and structural biotechnology journal* 17:186–194.

Liu, F., W. T. A. Enanoria, J. Zipprich, S. Blumberg, K. Harriman, S. F. Ackley, W. D. Wheaton, J. L. Allpress, and T. C. Porco. 2015, May. "The role of vaccination coverage, individual behaviors, and the public health response in the control of measles epidemics: an agent-based simulation for California". *BMC Public Health* 15(1):447. PMCID: PMC4438575.

Lorch, L., W. Trouleau, S. Tsirtsis, A. Szanto, B. Schölkopf, and M. Gomez-Rodriguez. 2020. "A Spatiotemporal Epidemic Model to Quantify the Effects of Contact Tracing, Testing, and Containment". *arXiv preprint arXiv:2004.07641*.

Moghadas, S. M., T. N. Vilches, K. Zhang, C. R. Wells, A. Shoukat, B. H. Singer, L. A. Meyers, K. M. Neuzil, J. M. Langley, M. C. Fitzpatrick, and A. P. Galvani. 2021, 01. "The Impact of Vaccination on Coronavirus Disease 2019 (COVID-19) Outbreaks in the United States". *Clin. Inf. Dis.*. ciab079.

Olu, O. O., M. Lamunu, M. Nanyunja, F. Dafae, T. Samba, N. Sempiira, F. Kuti-George, F. Z. Abebe, B. Sensasi, A. Chimbaru et al. 2016. "Contact tracing during an outbreak of Ebola virus disease in the Western area districts of Sierra Leone: lessons for future Ebola outbreak response". *Frontiers in Public Health* 4:130.

Salathé, M., C. L. Althaus, N. Anderegg, D. Antonioli, T. Ballouz, E. Bugnion, S. Capkun, D. Jackson, S.-I. Kim, J. Larus et al. 2020. "Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland". *medRxiv*.

Salathé, M., C. L. Althaus, R. Neher, S. Stringhini, E. Hodcroft, J. Fellay, M. Zwahlen, G. Senti, M. Battegay, A. Wilder-Smith et al. 2020. "COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation". *Swiss medical weekly* 150(1112).

Swanson, K. C., C. Altare, C. S. Wesseh, T. Nyenswah, T. Ahmed, N. Eyal, E. L. Hamblion, J. Lessler, D. H. Peters, and M. Altmann. 2018. "Contact tracing performance during the Ebola epidemic in Liberia, 2014-2015". *PLoS neglected tropical diseases* 12(9):e0006762.

The New York Times 2020. "Coronavirus (Covid-19) Data in the United States". https://github.com/nytimes/covid-19-data, accessed 7th January 2021.

AUTHOR BIOGRAPHIES

STEFAN HOOPS is a Research Associate Professor in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative at the University of Virginia. His email is shoops@virginia.edu.

JIANGZHUO CHEN is a Research Associate Professor in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative at the University of Virginia. His email is chenj@virginia.edu.

ABHIJIN ADIGA is a Research Assistant Professor in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative, University of Virginia. His email is abhijin@virginia.edu.

BRYAN LEWIS is a Research Associate Professor in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative at the University of Virginia. His email is brylew@virginia.edu.

HENNING MORTVEIT is an Associate Professor in the Networks, Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative, and at the Department of Engineering Systems and Environment at the University of Virginia. His email is Henning.Mortveit@virginia.edu.

JUSTIN CROW is a Director, Division of Social Epidemiology at Virginia Department of Health. His email is justin.crow@vdh.virginia.gov.

ELENA DISKIN is a Epidemiologist Program Manager at Virginia Department of Health. Her email is elena.diskin@vdh.virginia.gov.

SETH LEVINE is a Senior Epidemiologist at Virginia Department of Health. His email is seth.levine@vdh.virginia.gov.

HELEN TAZELAAR is a Health Equity Specialist at Virginia Department of Health. Her email is helen.tazelaar@vdh.virginia.gov.

BROOKE ROSSHEIM M.D., M.P.H. is a Public Health Physician Specialist on the Virginia Department of Health (VDH) COVID-19 Health Information Team. His email is brooke.rossheim@vdh.virginia.gov.

CHRIS GHAEMMAGHAMI is a Professor of Emergency and Internal Medicine at the University of Virginia School of Medicine. Virginia Department of Health's COVID-19 response team. His email is chris.ghaemmaghami@vdh.virginia.gov.

CARTER PRICE is a Senior Mathematician at the RAND Corporation and Codirector of the Center for Scalable Computing and Analysis. His email is price@rand.org.

HANNAH BAEK is a Senior Scientist in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative at the University of Virginia. Her email is yc3rz@virginia.edu.

REBECCA EARLY is VEDSS Manager, Epidemiologist Lead at Virginia Department of Health. Her email is jrebecca.early@vdh.virginia.gov.

MANDY WILSON is a Research Scientist in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative at the University of Virginia. Her email is alw4ey@virginia.edu.

DAWEN XIE is a Research Scientist in the Network Systems Science in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute at the University of Virginia. His email is dawenx@virginia.edu.

SAMARTH SWARUP is a Research Associate Professor in the Network Systems Science and Advanced Computing division at the Biocomplexity Institute at the University of Virginia. His email is swarup@virginia.edu.

SRINIVASAN VENKATRAMANAN is a Research Scientist at the Network Systems Science and Advanced Computing division at the Biocomplexity Institute at the University of Virginia. His email is srini@virginia.edu.

CHRIS BARRETT Executive Director, Distinguished Professor at the Biocomplexity Institute and Initiative, and Professor of Computer Science, School of Engineering and Applied Science at the University of Virginia. His email is ChrisBarrett@virginia.edu

MADHAV V. MARATHE Distinguished Professor in Biocomplexity, Division Director Network Systems Science and Advanced Computing division at the Biocomplexity Institute and Initiative, and Professor of Computer Science, School of Engineering and Applied Science at the University of Virginia. His email is marathe@virginia.edu