

Network Modeling and Analysis of COVID-19 Testing Strategies

Siqi Zhang, Marta J. Ventura and Hui Yang*

Abstract—The COVID-19 preparedness plans by the Centers for Disease Control and Prevention strongly underscores the need for efficient and effective testing strategies. This, in turn, calls upon the design and development of statistical sampling and testing of COVID-19 strategies. However, the evaluation of operational details requires a detailed representation of human behaviors in epidemic simulation models. Traditional epidemic simulations are mainly based upon system dynamic models, which use differential equations to study macro-level and aggregated behaviors of population subgroups. As such, individual behaviors (e.g., personal protection, commute conditions, social patterns) can't be adequately modeled and tracked for the evaluation of health policies and action strategies. Therefore, this paper presents a network-based simulation model to optimize COVID-19 testing strategies for effective identifications of virus carriers in a spatial area. Specifically, we design a data-driven risk scoring system for statistical sampling and testing of COVID-19. This system collects real-time data from simulated networked behaviors of individuals in the spatial network to support decision-making during the virus spread process. Experimental results showed that this framework has superior performance in optimizing COVID-19 testing decisions and effectively identifying virus carriers from the population.

Index Terms—Epidemic modeling, network models, testing strategies, data-driven risk scoring systems, COVID-19

I. INTRODUCTION

COVID-19 not only impacts everyone's daily life in a negative way but also poses significant challenges to the health and economy of our society. From the Centers for Disease Control and Prevention, effective testing strategies are critical to identify and isolate virus carriers in a timely manner, thereby mitigating the virus spread process. Such testing strategies usually involve a variety of decisions (e.g., sampling, testing frequency and contact tracing)

Simulation models are often utilized to capture the spatiotemporal dynamics of infectious diseases and aid the design of health policies [1]. From a broader view, simulation models can be briefly categorized into two classes [2]. The first class is called system dynamic models, which consider homogeneous human behaviors at a macro level for the characterization of epidemic dynamics. Because the heterogeneity of human behaviors can't be adequately modeled within population subgroups, these models tend to be limited in the provision of operational details for the design and evaluation of health policies. The second class is called discrete-event simulations (DES), which consider the virus

spread process as systems of individuals, as well as their activities and interactions in the spatial environment. With a detailed representation of human behaviors, the heterogeneity of human dynamics can be captured in the virus spread process. Hence, DES models offer great advantages for the design of COVID-19 testing strategies.

However, most of existing works related to testing strategies are based on system dynamics models. For example, Cashore et al. [3] developed a compartment model to simulate the spread of infectious diseases on campus during the fall semester and discussed the impact of different testing strategies for epidemic control. Paltiel et al. [4] analyzed the symptom-based screening and tests with varying testing frequency, sensitivity, specificity and cost to discuss the safe reopening of US campuses. Johansson et al. [5] analyzed the expected effectiveness of different testing strategies on the control of travel-related virus transmission. Note that these studies adopt the extended SEIR model and assume that screening and testing are performed on the subgroups of the population at each cycle, which tends to overlook the heterogeneity of individuals' behaviors and activities. Dickens et al. [6] considered the heterogeneity of travelers from different countries in the analysis of testing strategies at the entry point. However, very little work has been done to estimate individuals' infection risks in real-time for statistical sampling and testing of COVID-19.

Risk scoring systems are widely used in the ICUs to predict a patient's risk of mortality, organ dysfunction, disease severity. Examples include the Acute Physiology and Chronic Health Evaluation (APACHE), Simplified Acute Physiology Score (SAPS), Mortality Probability Model (MPM) [7]. These systems build predictive models to estimate the risk probability from a multivariate set of risk variables. For COVID-19, there are a variety of risk factors (e.g., age, social patterns, commute conditions) that impact the probability of infection [8]. However, little has been done to design a data-driven risk scoring system for the evaluation of individuals' exposure risk and then optimize COVID-19 testing strategies for the identification of virus carriers in a timely manner.

Hence, this paper presents a novel network-based simulation model for COVID-19 testing strategies. The main contributions are highlighted as follows:

- We design a data-driven risk scoring system for stratified sampling and testing of COVID-19.
- We leverage the network structure to simulate detailed human activities and characterize the spread dynamics of the infectious diseases in the spatial network.

*Corresponding author: Hui Yang (email:huy25@psu.edu)

The authors are with the Complex Systems Monitoring, Modeling and Control Lab, The Pennsylvania State University, University Park, PA 16802 USA

II. RESEARCH METHODOLOGY

In this section, we first discuss the proposed data-driven risk scoring system for real-time assessment of individuals' susceptibility and vulnerability. Then, we provide details about testing plans to identify virus carriers and control the virus spread process. Finally, we discuss the proposed framework to simulate detailed behaviors of human movements during the virus spread process in the spatial network.

A. Data-driven Risk Scoring System

Clinical testing is a critical step to identify virus carriers and control the virus spread in a spatial area. Because of practical issues like an inadequate supply of testing kits, prohibitive costs, it is often impossible to perform 100% testing of the population. In real-world practices, random sampling is often used. However, individuals' susceptibility and vulnerability to the virus are not random but depend on several personalized attributes (e.g., age groups, social patterns, protective measures). Hence, this paper presents a data-driven risk scoring system to first characterize individuals' infection risks, then stratify them into different risk groups, and finally optimize testing decisions.

The data-driven scoring system builds a predictive model to characterize the relationship between a multivariate set of risk factors and an individual's infection risk. Here, we consider 10 risk factors as predictors, as shown in Fig. 1. There are three predictors relating to personalized attributes, which include age, medical conditions, and symptomatic vs. asymptomatic. There are five predictors relating to interactions with others in the spatial environment, which include residential settings/commute, work/school mode, public gatherings, travel history, and contact tracing. Because human movements and interactions are dynamic, these risk factors are also time-varying in the virus spread process. Lastly, there are two predictors relating to protective measures and test history. When individuals wear masks and maintain 6-feet social distances, their infection risks are reduced. Furthermore, the risk factor of the test history is considered to measure the degree of the risk for an undetected virus carrier transmitting the virus. When a large number of these virus carriers exist in a spatial region, the remaining susceptible individuals are more likely to get infected.

In the system, each risk factor is worth a value from 0 to 4, where 0 refers to the least risky situation and 4 refers to the riskiest situation. For example, individuals with positive test results are given a value of 0 for the test history, while individuals who haven't been tested within two test cycles are given a value of 4. After collecting these risk factors, the cumulative infection risk y can be derived from a multivariate logistic regression model as $\log \frac{y}{1-y} = a + \sum_i b_i x_i$ where y is the infection risk, a is the intercept, b_i is the coefficient term for the risk factor x_i . Here, both training data and the domain knowledge can be used to adjust coefficients and benchmark the predictive model for real-time estimation of the infection risk for individuals in spatial regions.

B. Testing Plans

Figure 1 provides the diagram of detailed testing plans for statistical sampling and testing of virus carriers in a spatial region. Note that testing decisions are guided by the trained data-driven risk scoring system discussed above. Based on collected data from the questionnaire, the risk scoring system will provide estimated risks for individuals in a spatial area and then categorize them into different risk groups as follows: high (0.75~1), medium (0.5~0.75), and low (0~0.5). For the group of high-risk individuals, 100% testing is used to identify virus carriers for isolation and quarantine. For the group of medium-risk individuals, the Lot Quality Assurance Sampling (LQAS) strategy is utilized to sample test participants in order to better allocate testing resources based on budget constraints. For the group of low-risk individuals, 0% testing can be considered because they have a much lower probability of getting infected. For individuals who don't fill the questionnaire, LQAS can also be applied to make testing decisions. Then, it takes a certain amount of time (i.e., two days) for test participants to get their test results. Four different test results are possible, namely true positive, true negative, false positive, and false negative. When an individual gets a positive result, both contact tracing and isolation are triggered. The contact tracing strategy depends on whether a confirmed virus carrier is symptomatic or asymptomatic at the moment of getting their testing results. For symptomatic cases, the system tracks all close contacts from two days before the symptom onset until the case is isolated from the remaining population. For asymptomatic cases, contact tracing focuses on the identification of close contacts occurring 2 days before the test was taken. Once found, these close contacts go to self-quarantine for seven days. In addition, they will get tests both before they enter the status of quarantine and leave the quarantine in order to avoid releasing individuals with long latency time. Finally, isolated individuals will stay at home until they are recovered.

C. Network Modeling of COVID-19 Testing

As shown in Figure 2, the proposed network model consists of five components, namely spatial network, human traffic, spread modeling, testing, and interventions in a closed loop to investigate detailed behaviors of human movements and interactions during the virus spread process in a spatial region. See more details in [8].

This framework is supported by the data-driven decision support system. Through real-time data collection and predictions of individuals' infection risks, testing decisions can be optimized. During the simulation, risk factors such as age, medical condition are not changing because they are invariant to time-varying human behaviors. For the risk factor of asymptomatic vs symptomatic, risk points are updated when the time for symptom onset is up for a virus carrier. We also assume that work/school mode and residential setting/commute remain the same until individuals get infected and also become aware of their infectious status. For symptomatic cases, this condition is met when individuals have

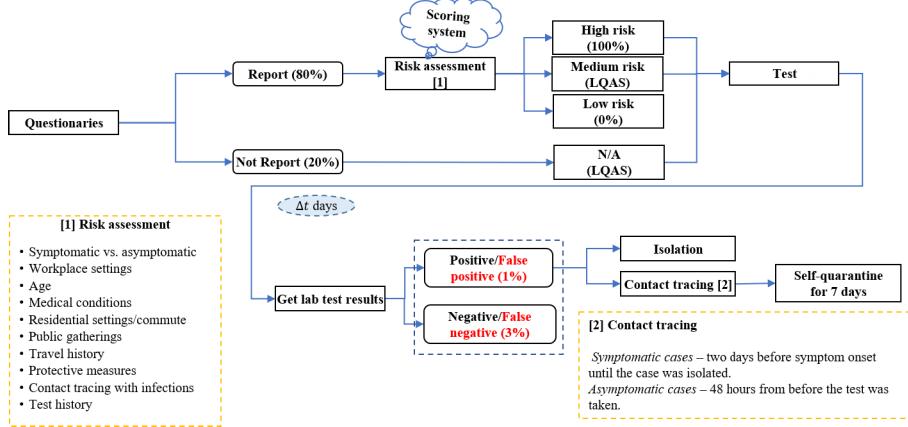


Fig. 1. Flow diagram of the risk scoring system and testing plans for individuals from different risk groups

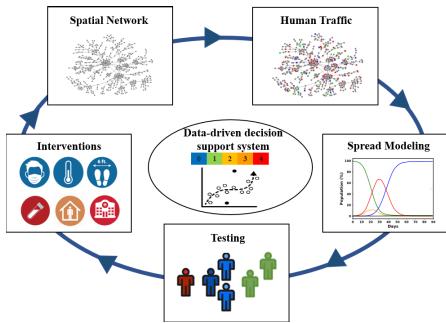


Fig. 2. Flow diagram of the proposed multi-layer network model for testing (See a demo of network models in this link: https://youtu.be/mKC_TXry3Lc)

symptom onset. For asymptomatic cases, this condition is only met when they are identified by the testing. Nodes of the spatial network are categorized into different gathering levels to model the heterogeneity of social patterns and estimate individuals' risks during the virus spread process. Because the node with a higher degree has more connections, it tends to be visited by many more individuals daily. Besides, a few nodes are randomly selected to represent travel locations of different risk levels. When individuals move and arrive at these nodes that are linked to risk factors, their corresponding risk points are immediately updated.

III. EXPERIMENTS AND RESULTS

A. Experimental Design

In order to evaluate and validate the proposed framework, a three-way layout experiment was designed, which includes three-factor groups, i.e., testing cycle τ , the ratio of symptomatic vs. asymptomatic cases γ , and the proportion of individuals filling the questionnaires β . In our simulation, three testing cycles (e.g., 2, 3, 7 days) are considered to evaluate how it will impact the virus control. Asymptomatic cases are often difficult to identify, diagnose and isolate. As such, we vary γ from 25/75 to 50/50 and finally to 75/25. In addition, the proportion of participants in questionnaires β is experimentally varied between 0.25 to 1.0 with a step size of

0.25. Two performance measures are considered, namely the percentage of identified virus carriers and the proportion of total infection cases. The first measure is defined to be the proportion of identified positive cases to the total number of positive cases in the spatial environment. When more virus carriers can be identified and isolated, the virus spread is slowing down. The second measure is defined to be the proportion of total infection cases to the entire population.

The simulation model is tested on a population of 6,000 people living, moving, and interacting with each other on a scale-free network with 5,000 spatial entities. There are four age groups: (1) 18-35, (2) 36-45 (3) 46-60 (4) 60+ and each group is 70%, 15%, 12% and 3% of the entire population, respectively. The simulation time is 60 days with one hour for each time step. In the human movement module, there are five activity groups that are considered to capture the heterogeneity of human activities. In the virus spread module, we assume that the percentage of mortality is 0.65% and 3.4% of symptomatic individuals are hospitalized [9]. When individuals wear masks for personal protection, the infection probability is reduced by 86%.

B. Experimental Results

As is shown in Figure 3, the boxplot is used to visualize different percentages of individuals who fill the questionnaire β against distributions of identified cases and total infection cases. The red, blue, and green boxplots represent experiments when the ratio of asymptomatic vs. symptomatic cases γ is 25/75, 50/50, and 75/25 respectively. At a given γ , identified cases increase as more individuals report their risk factors to the data-driven support system. This indicates that an increased availability of data improves the performance of risk assessment and thereby optimizes testing decisions. When comparing among different ratios of asymptomatic vs. symptomatic cases, identified cases are almost the same for the second and third ratios. For total infection cases, as the percentage of individuals who fill the questionnaire β increases from 25% to 100%, the median of total infection cases is reduced from 69.98% to 65.72% in the red boxplot. This indicates that data availability is critical in

identifying virus carriers and flattening the infection curve. Although more infection cases are identified at the higher ratio of asymptomatic vs. symptomatic cases, total infections are smaller at the smaller γ , as shown in Figure 3(b). This indicates that the existence of few asymptomatic virus carriers that are not identified can still be a concern for epidemic control. Therefore, in order to control the virus spread process, individuals are highly suggested to report their data to the data-driven decision systems for the risk assessment.

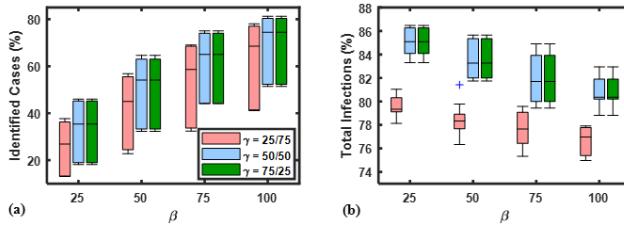


Fig. 3. Performance comparison of two performance measures: (a) identified cases (b) total cases on the variations of the percentage of individuals who fill the questionnaire

Figure 4(a) shows the performance comparison of two performance measures on the variations of test cycles in the boxplots. When the test cycle is prolonged, the median of identified cases quickly drops from 42.34% to 33.33% and finally to 18.41% in the red boxplot. That is because the testing cycle of seven days is much longer than the time interval between the infectious state and the symptom onset state (i.e., an average number of 3 days). The failure of identifying and isolating these virus carriers will cause more infection cases in the spatial environment. Figure 4(b) shows an increasing trend of infection cases when the test cycle is increasing. This implies that a larger testing cycle is not effective in identifying virus carriers and slowing down the virus spread. Hence, the choice of testing cycles can be critical to identify, isolate virus carriers and thereby mitigate the virus spread process.

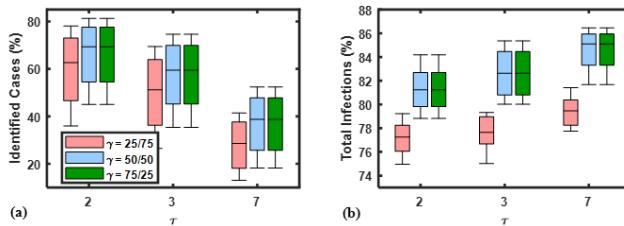


Fig. 4. Performance comparison of two performance measures (a) identified cases (b) total cases on the variations of test cycles

IV. CONCLUSIONS

The pandemic outbreak of COVID-19 has greatly changed our life. To control the spread of the virus, clinical tests of the population are required to identify both symptomatic and asymptomatic individuals before they infect others. Given a limited supply of testing kits and budget constraints, it is often impossible to test the entire population in a spatial region. However, prior efforts for the evaluation of COVID-19 testing strategy are more concerned about system dynamic models,

which are designed to study macro-level and aggregated behaviors of population subgroups in the virus spread process. Very little has been done to model detailed human behaviors for the evaluation of COVID testing strategies. Hence, this paper presents a network-based simulation model for the analysis of COVID-19 testing strategies. Specifically, we simulate the movements of human activities at a micro-level and characterize the spatiotemporal dynamics of COVID-19 in the spatial network. Experimental results show that the proposed framework not only provides detailed behaviors of individuals in the spatial network for the evaluation of COVID-19 testing strategies but also optimizes the testing decisions through data-driven support systems for effective identification of virus carriers.

ACKNOWLEDGEMENT

The authors would like to acknowledge the NSF I/UCRC Center for Healthcare Organization Transformation (CHOT), NSF I/UCRC award IIP-1624727, and NSF RAPID grant IIP-2026875 for funding this research.

DISCLAIMER

This paper doesn't involve experimental procedures on either human subjects or animal models, but rather use the data available in the public domain. Thus, approval is not needed from the Institutional Review Board.

REFERENCES

- [1] D. Du, H. Yang, A. R. Ednie, and E. S. Bennett, "Statistical metamodeling and sequential design of computer experiments to model glyco-altered gating of sodium channels in cardiac myocytes," *IEEE journal of biomedical and health informatics*, vol. 20, no. 5, pp. 1439–1452, 2015.
- [2] H. Yang, S. T. Bukkanpatnam, and L. G. Barajas, "Continuous flow modelling of multistage assembly line system dynamics," *International Journal of Computer Integrated Manufacturing*, vol. 26, no. 5, pp. 401–411, 2013.
- [3] J. M. Cashore, N. Duan, A. Janmohamed, J. Wan, Y. Zhang, S. Henderson, D. Shmoys, and P. Frazier, "Covid-19 mathematical modeling for cornell's fall semester (2020)," URL <https://people.orie.cornell.edu/frazier/COVID/>, vol. 19.
- [4] A. D. Paltiel, A. Zheng, and R. P. Walensky, "Assessment of sars-cov-2 screening strategies to permit the safe reopening of college campuses in the united states," *JAMA network open*, vol. 3, no. 7, pp. e2016818–e2016818, 2020.
- [5] M. A. Johansson, H. Wolford, P. Paul, P. S. Diaz, T.-H. Chen, C. M. Brown, M. S. Cetron, and F. Alvarado-Ramy, "Reducing travel-related sars-cov-2 transmission with layered mitigation measures: symptom monitoring, quarantine, and testing," *BMC medicine*, vol. 19, no. 1, pp. 1–13, 2021.
- [6] B. L. Dickens, J. R. Koo, J. T. Lim, H. Sun, H. E. Clapham, A. Wilder-Smith, and A. R. Cook, "Strategies at points of entry to reduce importation risk of covid-19 cases and reopen travel," *Journal of travel medicine*, vol. 27, no. 8, p. taaa141, 2020.
- [7] Y. Chen, F. Leonelli, and H. Yang, "Heterogeneous sensing and predictive modeling of postoperative outcomes," *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement*, pp. 463–501, 2016.
- [8] H. Yang, S. Zhang, R. Liu, A. Krall, Y. Wang, M. Ventura, and C. Delftch, "Epidemic informatics and control: A holistic approach from system informatics to epidemic response and risk management in public health," in *AI and Analytics for Public Health - Proceedings of the 2020 INFORMS International Conference on Service Science*. Springer, 2021, ch. 1, pp. 1–46.
- [9] D. Baud, X. Qi, K. Nielsen-Saines, D. Musso, L. Pomar, and G. Favre, "Real estimates of mortality following covid-19 infection," *The Lancet infectious diseases*, vol. 20, no. 7, p. 773, 2020.