

Stochastic vs. Deterministic Modeling for the Spread of COVID-19 in Small Networks

Mohammad Mubarak

James Berneburg

Cameron Nowzari

Abstract—This paper proposes and analyzes a stochastic Susceptible-Exposed-Infected-Removed (SEIR) spreading model on networks. Imagine a nursing home housing 28 seniors and 7 staff workers, in which one of the staff has tested positive for COVID-19. Unfortunately, the results of this test are 3 days late and the infected person had not been quarantining while waiting for their test results. What is now the individual risk to the different people living in this nursing home? If the home has access to two rapid COVID-19 viral tests, who should they be given to and why? In order to answer questions like this, we need to study stochastic models rather than deterministic ones. Unlike the vast majority of works that analyze various deterministic models, stochastic models are required when analyzing the risk of COVID-19 to individual people rather than tracking aggregate numbers in a given region. More specifically, this paper compares the results provided by analyzing stochastic and deterministic models and investigating when it is suitable to use the different models. In particular, we show why it is not suitable to use deterministic models when analyzing the spread in small communities and how these questions can be better addressed using stochastic ones. Finally, we show the added complications that arise due to the relatively long incubation period of COVID-19, and how it can be addressed. A simulated case study of the spread of COVID-19 in a 35-person nursing home is used to help illustrate our results.

I. INTRODUCTION

The study of various epidemic processes has been a longstanding research area with the earliest models proposed by Bernoulli in 1760 [1], [2], [3], [4]. With the rapid onset of the COVID-19 pandemic, it is no surprise that researchers worldwide are collectively trying to apply existing results on epidemic processes to the novel coronavirus to help figure out how to best combat it. Given that this is a brand new virus that has never been seen before, there is not yet a single established compartmental model to best describe how it spreads. Most current works, in the literature, focus on deterministic models that are only good for tracking aggregate numbers. Unlike exact stochastic models that are much more suitable for understanding the spread of COVID-19 at person to person level. Better understanding of how the virus spreads at the person-to-person level will also be helpful in making contact tracing efforts more efficient. Various smartphone apps already exist to help with these efforts by notifying people when they might have been in contact with someone who has tested positive for COVID-19. Through the help of the technology developed by Google and Apple in [5] that anonymously reports when different devices are within 6 feet of one another, we imagine our results in this paper could be used to help improve these apps by better quantifying the risk of an individual person based on their local interactions. For instance, rather than simply

providing an end user with the potentially scary information that they may have been in contact with someone positive for COVID-19 three days ago, a more sophisticated app could quantify this risk and even make a recommendation to the user based on their individual data. This work identifies the challenges that must be overcome to develop an app like the ones proposed in [6], [7].

To this end, there are currently a myriad of different models being proposed, all with different numbers and types of compartments, all aimed at trying to better capture certain aspects of COVID-19 as we learn about them. For example some works add an additional compartment to try to capture the asymptomatic transmitters [8], [9], [10], [11]. Another common variation is adding a compartment for deceased individuals [12], [13]. Other works may try to incorporate human behavior (e.g., wearing masks or not) by adding additional compartments [14]. Regardless of how the final compartments are chosen, all the works mentioned above and the vast majority of similar works only study deterministic models [1], [15], [16], [17], [18]. Unfortunately, it is known the mean-field approximations are only suitable for large numbers of people N . While these models are useful for tracking the aggregate number of infections in a large population (or network of sub-populations), they are not suitable for tracking the spread of the virus in a 35-person nursing home.

Since we are considering only small networks, mean-field approximation and other mass-action models cannot accurately capture what is happening at the individual person level. Most closely related to our research are the few works in the literature that instead study exact stochastic compartmental models rather than deterministic approximations. The work [19] investigates the connections between the exact stochastic models (2^N dimensional Markov Chain) and their mean-field approximations for the SIS compartmental model. In [20], [21], [22] the authors extend this type of analysis to the slightly more complicated SIRS model (3^N dimensional Markov chain). In [23], [24] the exact SIR model is analyzed for various specific small graph structures or graphs with some special properties (e.g., no loops). Many of these established results are discussed in the book [4, Chapter 2]. In this work we extend these ideas to the SEIR model (4^N dimensional Markov Chain) as the abnormally long incubation period seems to be an especially important artifact of how COVID-19 spreads, as this paper will show. A similar type of motivation and problem is discussed in [25], [26] for the continuous time SIS model.

Rather than a more sophisticated model with additional compartments, this work considers an SEIR model with only 4 compartments that captures when an individual is Susceptible, Exposed, Infected, or Removed. Instead of trying to

The authors are with the Department of Electrical and Computer Engineering, George Mason University, Fairfax, VA 22030, USA, {mmubara, jbernebu, cnowzari}@gmu.edu

figure out what type of compartmental model is best suited for COVID-19, we are interested in the difference between the deterministic models and their stochastic counterparts. We opt for the 4-state SEIR model rather than the simpler 3-state SIR model to show the additional challenges that arise when trying to properly model the abnormally long incubation period of COVID-19 at the individual level. Although various forms of the SEIR model have been used to capture the intermediate ‘Exposed’ state between becoming fully ‘Infected’, all the compartmental models mentioned above assume that these transitions are Poisson processes. Unfortunately, by now it is quite well known that the incubation period of COVID-19 does not follow an exponential distribution, but rather a Weibull/Log-normal one [27]. In this work, we assume that the incubation period follows the Log-normal distribution.

Statement of Contributions: This work helps better understand how COVID-19 spreads in small communities or groups of people (e.g., $N < 40$ people). The contributions of this paper are threefold. First, we show how to construct the 4^N -state Markov Chain describing the evolution of the 4-state SEIR model and how it can be used to track the expected states of individual people that may be given virus/antigen tests at different times. Second, we compare the results of the stochastic model to its deterministic mean-field approximation and discuss when it is suitable to use one over the other. Finally, we show the complications that arise when accounting for the unusually long incubation period of COVID-19. To the best of our knowledge, there are no other works suitable for even formalizing how to quantify the risk of COVID-19 to individual people. This paper provides a first step towards answering these types of questions by identifying the different challenges that must be addressed.

II. THE STOCHASTIC MODEL

Here we formalize the discrete-time network SEIR model of interest. Consider a group of N people (nodes), interacting according to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$, where the set of nodes $\mathcal{V} = \{1, \dots, N\}$ captures all the people in our community. The edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ represents the interactions between the different people, and the adjacency matrix $A = [a_{ij}]$, where $a_{ij} = 1$ represents the fact that person i and person j regularly interact with one another and the virus can pass between them.

Note that in general A should then be a time-varying graph. In particular, if real-time mobility data is available such as those produced from Apple and Google’s apps that can identify which devices are within 6 feet away, this data can be used to construct time-varying contact graphs $A(t)$. For simplicity, we are only considering static graphs here but note that all our results are easily generalizable to time-varying graphs if the information is available at all times.

We keep track of the compartmental state of node i at discrete timestep k by $X_i(k)$. Depending on the granularity of data available, one discrete timestep will generally represent at most one day. At any given time k , each person’s state $X_i(k)$ should belong to exactly one of the four compartments $X_i(k) \in \mathcal{C} = \{S, E, I, R\}$.

Figure 1 shows the compartmental model for a single person. A person in the Exposed state will naturally move

to the Infected state, then to the Removed state over time. However, a person that is in the Susceptible compartment can only move to the Exposed compartment through interactions with infected individuals. It is worth noting here that unlike some similar COVID-19 spreading models, we assume that a person in the S compartment can only become exposed to the virus through interactions with Infected people, not Exposed people. The probability that an individual might transition from one compartment to the next in one timestep are defined by $\beta^{\text{eff}}, \gamma, \delta > 0$, where β^{eff} depends on the states of the individual’s neighbors. The term β^{eff} will be explained soon, whereas the incubation rate γ and recovery rate δ are fixed constants that don’t depend on interactions with other people. Note that our model does not distinguish between people who have recovered or have died and we lump these individuals in the ‘Removed’ state. Denote the



Fig. 1. SEIR compartmental model for a single person. The chance that a person in the Susceptible state will move to the Exposed state is given by $\beta^{\text{eff}} = 1 - (1 - \beta)^{|\mathcal{N}_i^I|}$. The chance of moving from Exposed to Infected and Infected to Removed are given by γ and δ , respectively.

Markov Chain process for the entire system as $\{X(k)\}_{k \geq 0}$, where $X(k) \in \mathcal{C}^N$ is the Markov state and each element $X_i(k)$ of $X(k)$ is the state of node i . At each time k , each of the N nodes in the network will be in one of four possible states, the total number of possible states of the network is then $n = 4^N$. We refer to the collection of all distinct states as $\mathcal{S} = \{S_1, \dots, S_n\}$, where each S_c corresponds to a unique element in $\mathcal{C}^N$, for $c = 1, \dots, n$. Let $Y(k) \in [0, 1]^n$ be the probability simplex of the Markov Chain, such that $\mathbf{1}_n^T Y(k) = 1$ and the probability of the Markov Chain being in state S_c at time k is $Y_c(k)$. Stated mathematically, that is $\Pr[X(k) = S_c] = Y_c(k)$.

Now, we can more formally define the probability of a single node transitioning from each state into any other state. For convenience, we define $\mathcal{N}_i^I(k) \triangleq \{j \in \mathcal{N}_i | X_j(k) = I\}$ as the set of infected neighbors of person i , and $\beta^{\text{eff}} = 1 - (1 - \beta)^{|\mathcal{N}_i^I|}$. We mathematically represent the potentially nonzero individual transition probabilities over a graph as follows

$$\begin{aligned} \Pr[X_i(k+1) = S | X_i(k) = S] &= (1 - \beta^{\text{eff}}), \\ \Pr[X_i(k+1) = E | X_i(k) = S] &= \beta^{\text{eff}}, \\ \Pr[X_i(k+1) = E | X_i(k) = E] &= (1 - \gamma), \\ \Pr[X_i(k+1) = I | X_i(k) = E] &= \gamma, \\ \Pr[X_i(k+1) = I | X_i(k) = I] &= (1 - \delta), \\ \Pr[X_i(k+1) = R | X_i(k) = I] &= \delta, \\ \Pr[X_i(k+1) = R | X_i(k) = R] &= 1. \end{aligned} \quad (1)$$

All state transitions that are not listed here have probability zero. Now, given some initial distribution $Y(0)$, we wish to propagate the distribution forward in time, so that we know what the probability of the Markov Chain being in each state is. This can be accomplished using the transition matrix $P \in \mathbb{R}^{4^N \times 4^N}$ of the Markov Chain. The transition matrix is a

stochastic matrix where each entry P_{rc} , for $c, r = 1, \dots, n$, is the conditional probability

$$P_{rc} = \Pr[X(k+1) = \mathcal{S}_c | X(k) = \mathcal{S}_r].$$

By applying the Law of Total Probability, we can define the forward propagation of the distribution $Y(k)$ as

$$Y^T(k+1) = Y^T(k)P. \quad (2)$$

Now, we must calculate the transition probabilities for the entire Markov state so that we can write the entries of P . The transition probability from Markov state $\mathcal{S}_r$ to Markov state $\mathcal{S}_c$ can be calculated as the product of each node's state transition probability from Equation (1).

This works because transitions of individual nodes, given the previous Markov state, occur independently. Such a calculation can be performed as follows

$$P_{rc} = \prod_{i=1}^N \Pr[X_i(k+1) = \mathcal{S}_c(i) | X_i(k) = \mathcal{S}_r(i)],$$

where $\mathcal{S}_c(i)$ represents the state $X_i(k) \in \mathcal{C}$ of node i for the associated co-domain state $\mathcal{S}_c \in \mathcal{S}$.

Because the distribution of the Markov Chain can be thought of as a joint probability distribution for the individual nodes, we can calculate the individual distributions as marginal distributions. This can be done as follows

$$\Pr[X_i(k) = s] = \sum_{\ell \in \mathcal{L}_i^s} Y_\ell(k), \quad (3)$$

where $\mathcal{L}_i^s \triangleq \{\ell : \mathcal{S}_\ell(i) = s\} \subset \mathcal{S}$ is the set of co-domain states for which $X_i = s$, for $s \in \mathcal{C}$. This allows us to calculate the probability of each node i being in each compartment $\{S, E, I, R\}$ at each time step k .

A. Single Sample with Known Prior Distribution

Now that we know how to propagate the Markov distribution forward in time from an initial distribution $Y(0)$, we wish to determine how sampling a node affects that distribution, e.g., how does the result of a viral/antigen test given to a particular person update the posterior probabilities of other people being infected?

Suppose that we sample node i at time $k = K$, so that we know that $X_i(K) = s$, where $s \in \mathcal{C}$. In order to incorporate this information from the sample into our analysis, we must calculate a distribution conditioned on it. That is, we wish to find $Z(K) \in [0, 1]^n$, where $Z_c(K) = \Pr[X(K) = \mathcal{S}_c | X_i(K) = s]$. Additionally, we wish to propagate this distribution forward in time, so that we can calculate $Z(k)$ for $k \geq K$.

In order to simplify notation, we define $\mathcal{B}$ as the event $X_i(K) = s$. We begin by calculating each element of $Z(K)$:

$$Z_c(K) = \frac{\Pr[\mathcal{B} | X(K) = \mathcal{S}_c] \Pr[X(K) = \mathcal{S}_c]}{\Pr[\mathcal{B}]}, \quad (4)$$

where we've applied Bayes' Theorem. All these probabilities are known, because $\Pr[X(K) = \mathcal{S}_c] = Y_c(K)$, $\Pr[\mathcal{B}]$ is given by (3), and

$$\Pr[\mathcal{B} | X(K) = \mathcal{S}_c] = \begin{cases} 1, & \mathcal{S}_c(i) = s \\ 0, & \text{otherwise} \end{cases}.$$

The latter probability simply checks if the Markov state $\mathcal{S}_c$ has node i in a particularly state, and so is not a function of time. Recalling that $\mathcal{S}_c(i)$ denotes the individual state of node i in Markov state $\mathcal{S}_c$, we define a diagonal matrix $C_{\mathcal{B}} = \text{diag}([\dots, \Pr[\mathcal{B} | X(K) = \mathcal{S}_c], \dots])$, and we can write the matrix version of (4):

$$Z(k) = \frac{C_{\mathcal{B}} Y(k)}{\mathbf{1}_n^T C_{\mathcal{B}} Y(k)}.$$

Now, we must determine how to propagate this conditional distribution $Z(K)$ forward in time. Using the Law of Total Probability conditioned on an event, we can write the c th element of $Z(k+1)$, for $c = 1, \dots, n$, as

$$Z_c(k+1) = \sum_{r=1}^n \Pr[X(k+1) = \mathcal{S}_c | X(k) = \mathcal{S}_r \cap \mathcal{B}] Z_r(k). \quad (5)$$

Now, due to the Markov property, we know that, as long as $k \geq K$, $\Pr[X(k+1) = \mathcal{S}_c | X(k) = \mathcal{S}_r \cap \mathcal{B}] = \Pr[X(k+1) = \mathcal{S}_c | X(k) = \mathcal{S}_r] = P_{rc}$. As a result, we can rewrite Equation (5) in matrix form as

$$Z^T(k+1) = Z^T(k)P,$$

for $k \geq K$. Therefore, the same transition matrix P can be used to propagate the conditional distribution $Z(k)$ forward from time K .

B. Multiple Samples with Known Prior Distribution

In the previous section, we considered only one sample, so that the state of node i is known at time K . However, it is more practical to consider multiple samples, potentially occurring at different times.

Assume that, rather than a single sample, we have a set of M samples. Let $\mathcal{B}_m$ denote the event $X_{i_m}(K_m) = s_m$, where $s_m \in \mathcal{C}$ is the measured state, $K_m \geq K_{m-1}$ is the time at which the sampling occurred, and i_m is the index of the sampled node, for $m = 1, \dots, M$. For the sake of mathematical rigor, we also define $K_0 = 0$. Now, with a slight abuse of notation, define

$$\mathcal{B} \triangleq \cap_{m=1}^M \mathcal{B}_m.$$

As before, we ultimately want to find the distribution $Z(k) = Y^{\mathcal{B}}(k)$, where $Y^{\mathcal{B}}(t) = \Pr[X(t) = \mathcal{S}_c | \mathcal{B}]$, for $k \geq K_M$. We do this by sequentially applying the analysis in the previous section, using $Y(k)$ to find $Y^{\mathcal{B}_1}(k)$ for $k \geq K_1$, then using $Y^{\mathcal{B}_1}(k)$ to find $Y^{\mathcal{B}_1 \cap \mathcal{B}_2}(k)$ for $k \geq K_2$, and so on. Writing this out in terms of the initial conditions, we have

$$Z^T(K_M) = \frac{Y^T(0) \left(\prod_{m=1}^M P^{K_m - K_{m-1}} C_{\mathcal{B}_m} \right)}{Y^T(0) \left(\prod_{m=1}^M P^{K_m - K_{m-1}} C_{\mathcal{B}_m} \right) \mathbf{1}_n}, \quad (6)$$

where $C_{\mathcal{B}_m} = \text{diag}([\dots, \Pr[\mathcal{B}_m | X(K_m) = \mathcal{S}_c], \dots])$. The derivation is omitted because it is almost identical to the work in the previous section. Note that the transition matrix is present to propagate the conditional distributions forward in time. As before, we can use the transition matrix P to continue to propagate $Z(k)$ forward for $k \geq K_M$.

III. THE DETERMINISTIC MODEL

In this section, we proposed a discrete-time model derived from a deterministic mean field approximation of the 4^N state Markov Chain model presented in Section II, for the sake of comparison. Our discrete time mean-field approximation SEIR model follows a similar idea proposed in [18] for SIS.

Denoting $p_i^S(k), p_i^E(k), p_i^I(k), p_i^R(k)$ the approximated probability that node i is in S state at time k , E state at time k , I state at time k , R state at time k , respectively. Such that $p_i^S(k), p_i^E(k), p_i^I(k), p_i^R(k) \in [0, 1]$, where

$$p_i^S(k) + p_i^E(k) + p_i^I(k) + p_i^R(k) = 1.$$

The discrete time deterministic mean-field approximation of the SEIR model is given by

$$p_i^S(k+1) = \prod_{j \in \mathcal{N}_i} (1 - \beta p_j^I(k)) p_i^S(k),$$

$$p_i^E(k+1) = (1 - \gamma) p_i^E(k) + \left(1 - \prod_{j \in \mathcal{N}_i} (1 - \beta p_j^I(k)) \right) p_i^S(k),$$

$$p_i^I(k+1) = (1 - \delta) p_i^I(k) + \gamma p_i^E(k),$$

$$p_i^R(k+1) = p_i^R(k) + \delta p_i^I(k).$$

As in [28], and based upon our simulations, we conjecture that the exposed and infected states of our approximation upper-bound the expectations of these states in the exact 4^N state Markov Chain stochastic model. This is known to be true for ‘‘I-state’’ for the SIS model [29], [30], [31], and was shown to be true for SIRS and SIR models [20], [21], [32].

IV. MODELING INCUBATION PERIOD DISTRIBUTION

It is quite known that the spreading of the COVID-19 virus has a long incubation period with non-exponential distribution. As the abnormally long incubation period seems to be an especially important artifact of how COVID-19 spreads. Thus, in this section we extend the SEIR model (4^N -state Markov Chain) to (13^N -state Markov Chain). We assume that all transitions are Poisson processes except the transition from exposed to infected, which represents the incubation period, is not exponential. Empirical studies found that Log-normal distribution fits the incubation period of COVID-19 with a mean of 7.76 days [27].

In the construction of this approximation, an important role is played by the class of probability distributions called phase-type distributions [33]. Hence, modeling the Log-normal distribution can be done by expanding the Exposed state in the stochastic 4-state SEIR model, from a single state to p -internal state. We utilize the expectation-maximization algorithm proposed in [33]. Such that, we choose $p = 10$, which represents the number of the Exposed states in the expanded model.

As a result, we expand the stochastic 4-state SEIR model into stochastic 13-state SEIR model by adding 10 compartments into the Exposed state as illustrated in Figure 2. As can be noted, a new transition probability α appears in the model, where its value is highly dependent on the mean and standard deviation (STD) of the Log-normal probability distribution.

We leave it to the reader to check more details about this expansion technique in [34].



Fig. 2. 13-state SEIR compartmental model for a single person. The chance that a person in the Susceptible state will move to the Exposed state (E_0) is given by $\beta^{\text{eff}} = 1 - (1 - \beta)^{|\mathcal{N}_i^I|}$. The chance of moving from the internal Exposed compartments ($E_0, \dots, E_9$) to Infected and Infected to Removed are given by α and δ , respectively.

V. CASE STUDIES

Here we consider a case study in simulating nursing home with 28 seniors and 7 staff for a total of $N = 35$ people. Let us number the staff workers 1-7 and assume that person 4 has tested positive for COVID-19. How is this virus now expected to propagate throughout the nursing home? If additional tests are available, how should incorporating them affect the virus propagation?

The model parameters we used in these simulation were $\beta = 1/2.5$, $\gamma = 1/7.76$, $\delta = 1/15$, and $\alpha = 1.299$ with mean of 7.76 and STD of 1.6. First we will only consider the stochastic 4-state SEIR model. We initialize the entire network so that $X_4(0) = I$, and $X_i(0) = S$ for all $i \neq 4$. Figure 3 shows the initial condition along with 2 possible outcomes of what might happen after 12 days according to the dynamics (1). Due to the highly stochastic process, each snapshot looks significantly different after 12 days.

In order to compute the expected state of the individual nodes, we must analyze the exact stochastic model which is a 4^N -dimensional Markov Chain. Due to the intractability of this problem even for relatively small N , we will analyze the spread among the $N = 7$ staff workers to demonstrate our results. In particular we are interested in highlighting the need for analyzing the exact models even if they are computationally difficult, as the various deterministic approximations are not suitable here as we show.

This paper has discussed three different ways of computing the expected states of the different nodes over time. Figure 4 (a) shows the expected states of the nodes using the stochastic 4-state SEIR model, (b) shows the the expected states of the nodes using the deterministic 4-state SEIR mean-field approximation (MFA) model, and (c) shows the expected states of the nodes using the stochastic 13-state SEIR model. The colors correspond only to the maximally likely state of the different nodes at day 12, the exact probability distributions are shown for person 2 as an example. The main takeaway message here is the significant differences in the expected state of person 2 depending on the model being used. This suggests that none of the models provide similar outputs and so it does not make sense to use any of these as an approximation for any other one. In particular, the MFA cannot be used for such small numbers of nodes to accurately predict the individual persons’ probability distributions. Similarly, when considering small numbers of people the proper modeling of the log-normal distribution of the incubation period has a significant effect on the outputs, meaning that using Poisson processes as an approximation for the actual incubation period is also not very useful.

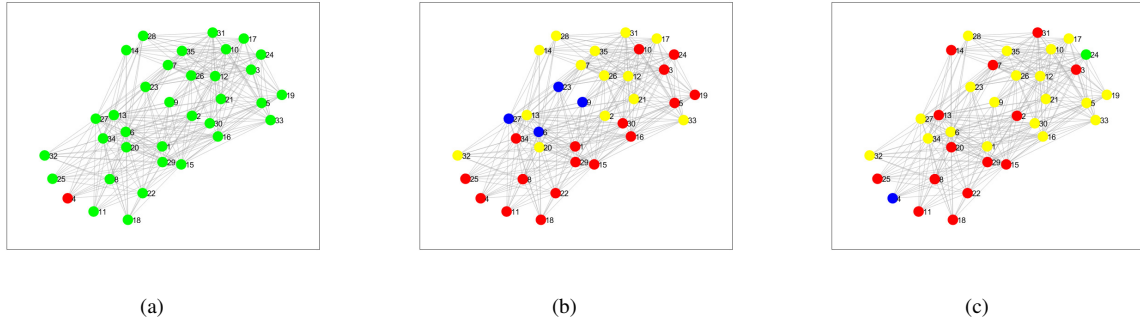


Fig. 3. Visualizing the entire network of $N = 35$ people in the nursing house using the stochastic 4-state SEIR model. (a) plot shows the initial condition at day 0. (b), (c) plots, show 2 possible outcomes at day 12. The states of some nodes are changing due to the highly stochastic process.

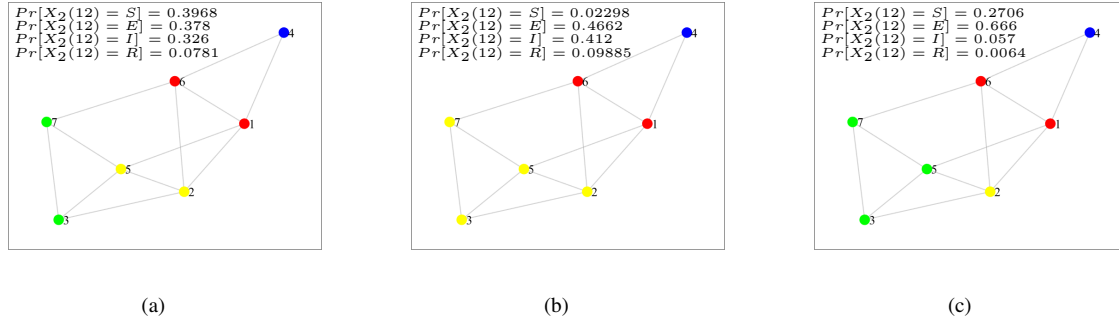


Fig. 4. Figure shows three possible outcome, of the $N = 7$ staff network at day 12, using the three proposed models. The colors correspond only to the maximally likely state of the different nodes. (a) shows the expected states of the nodes using the stochastic 4-state SEIR model. (b) shows the expected states of the nodes using the deterministic 4-state SEIR MFA model. (c) shows the expected states of the nodes using the stochastic 13-state SEIR model. The exact probability distribution are shown for node 2 as an example to demonstrate the discrepancy in the results based on the model's being used.

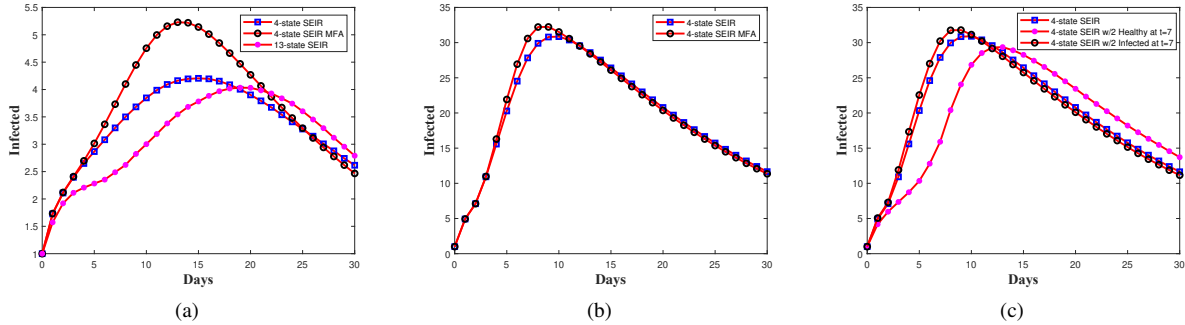


Fig. 5. The aggregate (average) number of expected infected (in I+E states) over time, of 1 day time step. (a) shows three curves for the 7 staff workers network in the nursing house; stochastic 4-state SEIR (blue marked), deterministic 4-state SEIR MFA (black marked), and stochastic 13-state SEIR (pink marked). The results demonstrate the difference between the models. (b) shows two curves for the 35-pepole network in the nursing house; 4-state SEIR (blue marked) and 4-state SEIR MFA (black marked). The plot shows the closeness in the results between the two models for large network. (c) shows three curves, adopting the stochastic 4-state SEIR model; the 4-state SEIR given the initial prior distribution of each node (blue marked), the 4-state SEIR model when incorporating one sample (test result), at day 7, of person 2 being healthy or infected, along with the known prior distribution.

Figure 5 (a), shows the aggregate (average) number of expected infected (in I+E states) over time of the $N=7$ staff workers network to further demonstrate the differences between the models, after averaging 30,000 simulations for the stochastic 4-state and the 13-state SEIR. Figure 5 (b), shows the average results of the total number of infected people over time of the $N=35$ people network, in the nursing house, after averaging 30,000 simulations for the stochastic 4-state SEIR, showing that the MFA is getting closer to the exact expected values when looking at aggregate numbers. However, the MFA model is not suitable in estimating the

state of any individual person.

In order to show the effect of nodes sampling on the average number of infected (I+E states), we suggest that person 2 was tested at day 7. The testing result was incorporated, in the $N = 35$ -people network, in the nursing house. Two cases were suggested, in the first case person 2 has tested positive for COVID-19 at day 7. In this case, we have a single sample for person 2 with a known prior distribution for all people in the network, so that $X_4(0) = I$, $X_2(7) = I$, and $X_i(0) = S$ for all $i \neq 4$. While in the second case, person 2

has tested negative for COVID-19 at day 7. In this case, we have $X_4(0) = I$, $X_2(7) = S$, and $X_i(0) = S$ for all $i \neq 4$. Figure 5 (c), shows the aggregate number of infected (in I+E states) over time to further demonstrate the differences between the models when incorporating single sample of person 2 in the network of 35 person in the nursing house. As a result, based on the available sampling information, the results demonstrates how such information can make the average number of infected people more accurate.

VI. CONCLUSIONS

We introduced the stochastic 4-state SEIR model by adopting the 4^N state Markov Chain, and we lately connected it to its deterministic mean-field approximation model. We also presented how to incorporate samples of nodes in a graph, into the model. Furthermore, we showed how to properly model the abnormal long incubation period with Log-normal distribution. At the end, we presented a case study in simulating a nursing house with a total number of 35 person, in which we compared the difference in the results between the proposed models. We confirmed that the stochastic models are much more suitable for the spread of COVID-19 on small networks. We also confirmed how the results of the stochastic model and it's deterministic mean-field approximation were close when dealing with larger number of population on a network, where the stochastic model becomes intractable. Moreover, the results showed how important including sampling information into the model. This motivates the need for further work on how to properly analyze networks that are too small for mean-field approximations, but too large to analyze the exact stochastic model.

ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 2028523.

REFERENCES

- [1] C. Nowzari, V. Preciado, and G. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, pp. 26–46, 2016.
- [2] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Royal Society*, vol. 115, p. 772, 1927.
- [3] N. T. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*. London: Griffin, 1975.
- [4] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of Epidemics on Networks From Exact to Approximate Models*. Springer, 2017.
- [5] Apple and Google, "Apple and google partner on covid-19 contact tracing technology," <https://www.blog.google/inside-google/company-announcements/apple-and-google-partner-covid-19-contact-tracing-technology/>, April 2020.
- [6] J. Abeler, M. Bäcker, U. Buermeyer, and H. Zillesen, "Covid-19 contact tracing and data protection can go together," *JMIR mHealth and uHealth*, vol. 8, p. e19359, 2020.
- [7] S. Altmann, L. Milsom, H. Zillesen, R. Blasone, F. Gerdon, R. Bach, F. Kreuter, D. Nosenzoa, S. Toussaert, and J. Abeler, "Acceptability of app-based contact tracing for covid-19: Cross-country survey study," *JMIR mHealth and uHealth*, vol. 8, p. e19857, 2020.
- [8] L. Basnarkov, "Seair epidemic spreading model of covid-19," *Chaos, Solitons and Fractals*, vol. 142, p. 110394, 2021.
- [9] J. B. Aguilar, J. S. Faust, L. M. Westafer, and J. B. Gutierrez, "Investigating the impact of asymptomatic carriers on covid-19 transmission," <https://www.medrxiv.org/content/10.1101/2020.03.18.20037994v2>, 2020.
- [10] S. Park, D. Cornforth, J. Dusho, and J. Weitz, "The time scale of asymptomatic transmission affects estimates of epidemic potential in the covid-19 outbreak," *Epidemics*, p. 100392, 2020.
- [11] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. D. Filippo, A. D. Matteo, and M. Colaneri, "Seir modeling of the covid-19 and its dynamics," *Nature medicine*, vol. 26, pp. 855–860, 2020.
- [12] G. C. Calafiore, C. Novara, and C. Possieri, "A modified sir model for the covid-19 contagion in italy," in *IEEE Conference on Decision and Control (CDC)*, (Jeju Island, Korea (South)), pp. 3889–3894, 2020.
- [13] D. Caccavo, "Chinese and italian covid-19 outbreaks can be correctly described by a modified sird model," *medRxiv*, 2020.
- [14] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. B. Gumel, "To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic," *Infectious Disease Modelling*, vol. 5, pp. 293–308, 2020.
- [15] F. D. Sahneh, C. Scoglio, and P. V. Mieghem, "Generalized epidemic mean-field model for spreading processes over multilayer complex networks," *IEEE/ACM Transactions on Networking*, vol. 21, pp. 1609–1620, Oct. 2013.
- [16] P. E. Paré, J. Liu, C. L. Beck, B. E. Kirwan, and T. Başar, "Analysis, estimation, and validation of discrete-time epidemic processes," *IEEE Transactions on Control of Network Systems*, 2018.
- [17] P. E. Paré, C. L. Beck, and A. Nedić, "Epidemic processes over time-varying networks," *IEEE Transactions on Control of Network Systems*, vol. 5, pp. 1322–1334, 2017.
- [18] H. J. Ahn and B. Hassibi, "Global dynamics of epidemic spread over complex networks," in *CDC IEEE on decision and control*, (Florence, Italy), pp. 4579–4585, 2013.
- [19] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1–14, 2008.
- [20] N. A. Ruhi and B. Hassibi, "Sirs epidemics on complex networks: Concurrence of exact markov chain and approximated models," in *IEEE Conference on Decision and Control (CDC)*, (Osaka, Japan), pp. 2919–2926, 2015.
- [21] S. Ottaviano and S. Bonaccorsi, "Some aspects of the markovian sirs epidemics on networks and its mean-field approximation," *arXiv preprint arXiv:1912.11287*, 2019.
- [22] P. V. Mieghem, "Exact Markovian SIR and SIS epidemics on networks and an upper bound for the epidemic threshold," *arXiv:1402.1731*, 2014.
- [23] K. MCCULLOCH, M. ROBERTS, and C. R. LAING, "Exact analytical expressions for the final epidemic size of a sir model on small networks," *The ANZIAM Journal*, vol. 57, no. 4, p. 429–444, 2016.
- [24] I. Kiss, C. Morris, F. Sélley, P. Simon, and R. Wilkinson, "Exact deterministic representation of markovian epidemics on networks with and without loops," *Journal of mathematical biology*, vol. 70, pp. 437–464, 2015.
- [25] M. Ogura and V. Preciado, "Stability of sis spreading processes in networks with non-markovian transmission and recovery," *IEEE Transactions on Control of Network Systems*, vol. 7, pp. 349–359, 2019.
- [26] P. V. Mieghem and Q. Liu, "Explicit non-markovian susceptible-infected-susceptible mean-field epidemic threshold for weibull and gamma infections but poisson curings," *Physical Review E*, vol. 100, p. 022317, 2019.
- [27] J. Qin, C. You, Q. Lin, T. Hu, S. Yu, and X. H. Zhou, "Estimation of incubation period distribution of covid-19 using disease onset forward time: A novel cross-sectional and forward follow-up study," *Science Advances*, vol. 6, 2020.
- [28] G. P. C. Nowzari, V. M. Preciado, "Optimal resource allocation for control of networked epidemic models," *IEEE Transactions on Control of Network Systems*, vol. 4, pp. 159–169, 2015.
- [29] P. V. Mieghem, "The n-intertwined SIS epidemic network model," *Computing*, vol. 93, no. 2-4, pp. 147–169, 2011.
- [30] C. Li, R. van de Bovenkamp, and P. V. Mieghem, "Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations," *Physical Review E*, vol. 86, p. 026116, 2012.
- [31] E. Cator and P. V. Mieghem, "Second-order mean-field susceptible-infected-susceptible epidemic threshold," *Physical Review E*, vol. 85, p. 056111, 2012.
- [32] M. Youssef and C. Scoglio, "An individual-based approach to sir epidemics in contact networks," *Journal of theoretical biology*, vol. 283, pp. 136–44, 06 2011.
- [33] S. Asmussen, O. Nerman, and M. Olsson, "Fitting phase-type distributions via the em algorithm," *Scandinavian Journal of Statistics*, vol. 23, no. 4, pp. 419–441, 1996.
- [34] C. Nowzari, M. Ogura, V. M. Preciado, and G. J. Pappas, "A general class of spreading processes with non-markovian dynamics," in *CDC IEEE Conference on Decision and Control*, (Osaka, Japan), pp. 5073–5078, 2015.