Chris Groendyke* and Adam Combs

Modifying the network-based stochastic SEIR model to account for quarantine: an application to COVID-19

<https://doi.org/10.1515/em-2020-0030>

Received August 12, 2020; accepted February 10, 2021; published online March 16, 2021

Abstract

Objectives: Diseases such as SARS-CoV-2 have novel features that require modifications to the standard network-based stochastic SEIR model. In particular, we introduce modifications to this model to account for the potential changes in behavior patterns of individuals upon becoming symptomatic, as well as the tendency of a substantial proportion of those infected to remain asymptomatic.

Methods: Using a generic network model where every potential contact exists with the same common probability, we conduct a simulation study in which we vary four key model parameters (transmission rate, probability of remaining asymptomatic, and the mean lengths of time spent in the exposed and infectious disease states) and examine the resulting impacts on various metrics of epidemic severity, including the effective reproduction number. We then consider the effects of a more complex network model.

Results: We find that the mean length of time spent in the infectious state and the transmission rate are the most important model parameters, while the mean length of time spent in the exposed state and the probability of remaining asymptomatic are less important. We also find that the network structure has a significant impact on the dynamics of the disease spread.

Conclusions: In this article, we present a modification to the network-based stochastic SEIR epidemic model which allows for modifications to the underlying contact network to account for the effects of quarantine. We also discuss the changes needed to the model to incorporate situations where some proportion of the individuals who are infected remain asymptomatic throughout the course of the disease.

Keywords: contact network; COVID-19; SARS-CoV-2; stochastic SEIR epidemic.

Introduction

In late 2019, the SARS-CoV-2, which is sometimes known colloquially as the novel coronavirus, and which is the virus which causes the COVID-19 illness, began to spread from Wuhan, China, to other cities, countries, and continents. By mid-December 2020, it had reached almost every corner of the globe, infecting over 76 million people, and causing over 1.6 million deaths (The New York Times 2020), also resulting in wide-scale sociological and economic impacts. Because most people have not yet been vaccinated, it is likely to continue to spread – at least in some locations – for at least several more months, and perhaps longer. Thus, it is important to develop high-quality models to study the dynamics of the spread of this disease.

Efforts to control and combat the spread of this epidemic have included strategies such as social distancing, self-isolation, and quarantine (among others), all of which are designed to alter the contact patterns

*Corresponding author: Chris Groendyke, Robert Morris University, Department of Mathematics, 6001 University Blvd., Moon Township, PA, 15108-2574, USA, Phone: +4123976054, E-mail: groendyke@rmu.edu

Adam Combs, Robert Morris University, Department of Mathematics, Moon Township, PA, USA

of individuals in a population, and particularly those who are infected with the disease (even if they are unaware they are infected). Due to these changes in contact patterns, particularly those only initiated by individuals upon developing symptoms of the disease, most of the models traditionally used to model the spread of epidemics, such as the standard SEIR (susceptible-exposed-infectious-removed) model are likely to be inadequate to model the progression of this epidemic. It seems clear that modifications to the standard models are needed in order to sufficiently account for these new dynamics.

Because this novel coronavirus emerged only recently, the literature on modeling the spread of this virus is not yet as robust or thorough as it is for other viruses. However, several researchers have attempted to model the spread of this epidemic. Many of these efforts involve using compartmental disease models such as SEIR or modified versions thereof, along with a “mean field” model for the contacts between members of the population. These mean field (sometimes called “random mixing” or “well-mixed”) models, which have a long history in the study of disease dynamics (Bailey 1950; Kermack and McKendrick 1927), allow for any infectious member to infect any member of the susceptible class. Under assumptions about the various disease parameters, it is often possible to express differential equations governing the sizes of the various compartments.

Many researchers have taken just this approach to study the novel coronavirus. Iwata and Miyakoshi (2020) and Wan et al. (2020) conducted studies of this epidemic using standard random mixing SEIR models. Hou et al. (2020) used a mean field SEIR model to describe this disease, and varied the rate of contacts to model the differences in transmission events caused by the effects of self-isolation. Peng et al. (2020) modified the mean field SEIR model by adding **Q** (quarantined) and **P** (insusceptible) classes. López and Rodo (2020) took a similar approach by adding a **Q** class to represent the quarantining of infectious individuals, as well as a protected population compartment (**C**) to account for efforts to protect susceptible population members by confining them. Shi, Cao, and Feng (2020) added three new classes to the standard mean field SEIR model to account for the quarantining of individuals who are susceptible, exposed, and infectious. Similarly, Wang et al. (2020) added new classes to the mean field SEIR model to account for asymptomatic cases and the effects of hospitalization.

There is, however, an alternative to the mean field approach of modeling interactions between population members which posits an underlying network describing the contact relationships between individuals. In this framework, disease transmission can only occur between individuals who share an edge in this network. This network approach can sometimes yield very different – and in some cases more realistic – dynamics for the spread of an epidemic (Ferrari et al. 2006; Keeling and Eames 2005; Meyers et al. 2005). Simulation studies using network models have been used in the past as an effective tool to investigate the properties of epidemics; this approach was taken in such works as Volz (2008) and Barthelemy et al. (2005).

While there have been comparatively fewer published works along these lines in the efforts to study the spread of the SARS-CoV-2 virus, several authors have in fact implemented network or other agent-based models to simulate or study the spread of this disease through various populations. Aleta et al. (2020) used an agent-based model with three layers (school, workplace/community, household); the intra-network links had various weights based on the likelihood of transmission in the corresponding settings. Their work utilized a compartmental model with nine classes, and fixed latent and pre-symptomatic periods. They used their model to test various containment strategies, including school closures and “stay-at-home” orders and found that contact tracing and increased testing can diminish the need for other strategies.

Sewell and Miller (2020) focused on the computational aspects of running epidemic simulations over contact networks, noting that “equation-based models may be utilized more quickly during an outbreak setting of an emerging infectious disease”. These authors used an SEIR disease model and a contact network constructed based on a study of contacts among people in Hong Kong. They studied the effects of both mask mandates and quarantining (as well as their combined effects) on the spread of this disease and noted the importance in implementing these measures in a timely fashion.

Braun et al. (2020) used a simulation involving Watts–Strogatz small world networks to assess the effectiveness of three different control measures (social distancing, personal protective equipment such as masks, and quarantining/behavior modifications), in terms of changes in the total number of infections and the length of the epidemic. They used an SIR disease model, but partitioned the infectious class into three different subclasses to account for differences between individuals who are presymptomatic, symptomatic, or asymptomatic.

Hoertel et al. (2020) used a network-based model to study the efficacy of various proposed interventions in France, including extending their national lockdown, distancing measures, mask wearing, and shielding the most vulnerable members of the population (post-lockdown). They considered the impacts of these measures on ICU bed occupancy, numbers of total cases, and numbers of total deaths. Their study used Weibull random variables to describe the lengths of time individuals spent in the various disease stages.

Kerr et al. (2020) developed software in Python to aid in the simulation of this disease, supporting three different types of network models. They used a modified SEIR model, where the infectious class is partitioned into symptomatic and asymptomatic cases, with the symptomatic further broken down into presymptomatic, mild, severe, and critical subclasses. This work used lognormal random variables to govern the length of time between transition between classes. They used probabilities of susceptibility, disease progression, and death that varied by the age of the individual. Probability of transmission between connected individuals varied based on the reason the individuals were connected, with household contacts having the highest probability of transmission, followed by workplace, school, and community contacts.

We also note that Prasse et al. (2020) used a network model to study the spread of this disease; however, this contact network was between cities rather than individuals. Further, both Di Domenico et al. (2020) and Zhang et al. (2020) used contact matrices to model the interactions of individuals in the population; these matrices segmented the population into groups based on similar characteristics with respect to their making contacts.

Here we present a modification to the network-based SEIR model that has two key changes that make it more appropriate for the study of this disease. First, there is an introduction of a separate class to model those individuals in self-isolation; this is similar to some of the approaches above, but treats this new class as a subset of the infectious class. (We will hereafter refer to this modified SEIR model as the SEI(Q)R model. We refer to this state as the quarantined (**Q**) state, though “isolation” or “self-isolation” are more accurate descriptors of the behavior being modeled by this state.) Second, our network model allows for specific changes in network contacts upon the presence of symptoms (and also allows for a percentage of the infectious population to remain asymptomatic), which could be used to simulate the likely change in contact patterns of an individual in self-isolation or quarantine. Thus, our network model is a dynamic, rather than static, network model, though the changes we allow in the network for this study are rather straightforward; see Bansal et al. (2010) for a discussion of the role of dynamic contact networks in the study of disease dynamics. In addition, we make use of stochastic models (as opposed to the deterministic ones sometimes utilized in mean field models) to describe the lengths of time spent by individuals in the various disease stages, allowing for considerable flexibility in modeling, as well as a straightforward way to assess uncertainty. We believe that this approach to studying SARS-CoV-2 provides a good framework for the investigation of this epidemic, particularly in smaller and/or closed populations.

The remainder of this paper is organized as follows: in “Epidemic and network models” we describe the epidemic and network models we use for this study, including the changes we make to the standard models to account for the effects of self-isolation and quarantine; in “An application to COVID-19”, we discuss how this model can be applied to study the spread of the SARS-CoV-2 virus, conduct a simulation study to assess the sensitivity of the model to the various parameters, offer an extension to the network model, and discuss parameter estimation; “Conclusions” concludes and offers ideas for future work and extensions of this model.

Epidemic and network models

SEI(Q)R epidemic model

Compartmental models, i.e., models that assign each member of the population to one of a finite number of classes based on the state of the disease progression within that individual, have long been used to model the progress of various infectious diseases in individuals. Kermack and McKendrick (1927) initially introduced an SIR compartmental model. This model consists of three classes: the susceptible (**S**) class consists of those individuals who do not currently have the disease, but could contract it at some future point in time; the infectious (**I**) class contains those individuals who can infect others in the population; the removed (**R**) class consists of those members of the population who can no longer spread the disease to others. Depending on the disease in question, it is sometimes assumed – as will be the case here – that once an individual enters the **R** class, they can never be reinfected, and hence play no further role in the progression of the epidemic through the population.

The SEIR model adds an exposed (**E**) class, corresponding to individuals in the population who have contracted the disease, but cannot yet infect others. Some early examples of this model include Schwartz and Smith (1983), Aron and Schwartz (1984), and Hethcote and Tudor (1980). This added class is a potentially important generalization of the SIR model, as it has been shown (Wearing, Rohani, and Keeling 2005) that failing to account for latent or incubation periods can lead to inaccurate or biased results. The SEIR model has been used to model many types of infectious disease, including HIV/AIDS (Li, Smith, and Wang 2001), and measles (Grenfell 1992; Momoh et al. 2013), as well as various types of influenza (Dukic, Lopes, and Polson 2012; González-Parra, Arenas, and Chen-Charpentier 2014; Grais, Ellis, and Glass 2003), and the novel coronavirus (Hou et al. 2020; Kuniya 2020). For a thorough review of compartmental epidemic models, as well as other approaches to modeling epidemics in populations, see Keeling and Rohani (2011).

We propose a modification to the SEIR model in which a quarantined (**Q**) class is added, as a subset of the **I** class. This new class is designed to account for the fact that an individual becoming symptomatic with disease indicators – which would happen at some point *after* their becoming infectious – may change their behavioral patterns in a way that impacts their interactions with others in the population. That is to say, a person who becomes aware of their having disease symptoms may not behave similarly (e.g., they may self-isolate) to others in the **I** class who do not show symptoms and thus have no reason to believe they are infectious. Thus, it seems reasonable to account for this sub-group of the **I** class separately; since the **Q** class is a sub-class of the **I** class, members of the **Q** class are still assumed to be able to infect others. (Though they may have fewer opportunities to do so, since their number of contacts will typically be assumed to decrease; details of this are given in “Modeling the network changes over the course of the epidemic”). We assume that a person entering the **Q** class will remain there for the duration of the time they are infectious, and will then move to the **R** class.

We also assume, however, that there is some proportion of the population who enter the **I** class who never enter the **Q** class; this feature is used to model the set of people who catch the disease, but never show any symptoms, and hence never have any reason to change their contact patterns. (But it could also be used to model individuals who are symptomatic but choose for whatever reason not to alter their contact patterns.) These individuals would then follow the standard progression through the SEIR model; thus the modified model can be represented by:

Finally, we note that in our model, the lengths of time spent in the **E**, **I**, and **Q** states are all stochastic, as is the time to infection along any given edge in the contact network. The specifications of these random

variables and accompanying parameters for this application are given in “Applying the SEI(Q)R model to the SARS-CoV-2 virus”.

Network structure

The model we use to describe the structure of the contact network for the (N individuals in the) population is a type of model known as a Bernoulli graph model whereby every pair of individuals in the population has a common probability p of having a contact relationship. This is a dyadic-independent model, meaning that the probability of an edge for a given dyad is independent of the presence or absence of edges among other dyads in the network. This model was first described in Gilbert (1959) and Erdős and Rényi (1959), and will be henceforth referred to as the Gilbert–Erdős–Rényi (or GER) model. This choice is intentionally simple and generic; it does not attempt to model any specific population. It also minimizes network effects so that the effects of the disease model can more easily be examined. Further, it can provide a baseline for comparison for future work that incorporates more complicated network structures; an example of such a comparison to a more complex network is provided in “A generalization of the network model”.

An important aspect of our development is the recognition that the presence of the disease can change the nature of the contact network structure. This is reflected in our model by specifying two different network parameters. In particular, we will have one network parameter corresponding to the situation where neither member of the dyad is in the **Q** state (represented by p), and a second network parameter for dyads where one member of the dyad is in the **Q** state, and the other is not (represented by p^*).

Note that, for the purpose of simulating the spread of the epidemic through the population, it is not necessary to consider the case where both members of the dyad are in the **Q** state. Recall that a transmission event can only occur from a member of the **I** class (which includes the **Q** class) to a member of the **S** class. Thus, since any future disease transmission is impossible for two individuals who have entered the **Q** state, the presence or absence of an edge between these individuals cannot impact the progress of the epidemic. We do note, however, that our method of estimating the reproduction number for an epidemic relies on the degree distribution of the network (see “Estimation of reproduction numbers” for details). For this reason, we do indeed consider a third network parameter, corresponding to **Q–Q** dyads, even though it has no impact on the actual spread of the disease in our model.

Modeling the network changes over the course of the epidemic

At the point in the epidemic when an individual i enters the **Q** state – should such a transition occur – we reconsider the network connections involving this individual. Every dyad involving individual i , whether or not it currently has an edge in the contact network, is individually considered. Dyads not including individual i are ignored at this stage.

At the time an individual i enters quarantine, we will let E denote the event that a given dyad (involving individual i) is connected by an edge, and E' be the complementary probability of this edge not existing. We will also define E^* as the event that such an edge exists after the network adjustments are made as a result of quarantine. Let p represent the *a priori* probability of an edge between any individuals i and j when neither individual is in the **Q** state, and let p^* represent the *a priori* probability of such an edge when either i or j (but not both) is in the **Q** state.

In order to get the desired expected post-quarantine edge density for individual i , we will condition on the pre-quarantine existence of an edge, giving

$$\begin{aligned} p^* &= P[E^*] \\ &= P[E^* | E] \cdot P[E] + P[E^* | E'] \cdot P[E'] \\ &= P[E^* | E] \cdot p + P[E^* | E'] \cdot (1 - p). \end{aligned} \tag{1}$$

Because there is not a unique solution for the probabilities $P[E^*|E]$ and $P[E^*|E']$, we must impose a further condition. There are many possible options, and the choice will depend on the nature of the phenomenon we are trying to model; for this application, we choose to set $P[E^*|E'] = 0$. This choice implies that no new edges are formed as a result of moving to the **Q** state, and the only edges remaining for individual i post-quarantine will be a subset of that individual's pre-quarantine edges. Other choices are certainly possible; it might be reasonable in some cases to assume that there is some (presumably small) chance of forming new edges as a result of entering the **Q** state, though we disallow such possibilities here.

Then, under these assumptions, we can solve for the probability of a pre-quarantine edge involving individual i remaining in the network as

$$P[E^*|E] = \frac{p^*}{p}. \quad (2)$$

Note that in a typical application, we will have $p^* < p$, so that the ratio in Eq. (2) will be less than 1. However, this need not necessarily be the case, i.e., this model allows for situations in which $p^* > p$, and in this event, we could either bound this probability by 1 (and adjust the probability $P[E^*|E']$ accordingly to yield the correct edge density) or consider a different type of constraint on the conditional probabilities in Eq. (1) when making adjustments to the edges in the network. The process for adjusting the network in the case where both individuals in a dyad are in the **Q** state is analogous.

An application to COVID-19

In this section, we describe how we apply the model discussed above to study the spread of the COVID-19 disease. We first discuss the distributions and parameter values used to describe the lengths of time spent by individuals in the various states; we will refer to these as the **baseline** parameter values. We then describe how to estimate reproduction numbers (R_0 and R) using our network and epidemic models. Finally, through a simulation study, we analyze the impact of varying the different model parameters and network type on statistics of interest with respect to the spread of the epidemic.

Applying the SEI(Q)R model to the SARS-CoV-2 virus

To determine our baseline model distributions and parameter values, we adapt the results of several previous studies; this is complicated somewhat by the fact that different researchers have modeled somewhat different quantities than the ones in our study. He et al. (2020) assumed the distribution for the *incubation period* (length of time from exposure to becoming symptomatic) to be lognormal with mean of 5.2 days (see also Li et al. 2020). This is similar to results in Linton et al. (2020) which found the best fitting model for the incubation period to be lognormal with mean of 5.0 days. He et al. (2020) also estimated the number of days between the start of being infectious and the start of symptoms, that is the pre-symptomatic period, to be about 2.3 days. Combining these results, we use the lognormal distribution with mean of 2.9 days (and standard deviation of 2.51 days) to describe the length of time from being exposed to becoming infectious, i.e., the length of time spent in the **E** state.

It has been widely observed that some proportion of the individuals contracting this disease never develop symptoms. For our baseline value of this asymptomatic percentage (which we label α), we use the “best estimate” scenario given by the CDC (Center for Disease Control and Prevention 2020) of 35%, which is similar to other estimates, such as Nishiura et al. (2020). For those individuals who do become symptomatic, we assume that the length of their pre-symptomatic period, i.e., the length of time spent in the **I** state before entering the **Q** state, is uniformly distributed between 2 and 3 days, which is consistent with the estimates given by He et al. (2020). The individuals who remain asymptomatic will skip the **Q** state and remain in the **I** state for the entirety of the time they are infectious.

For this study, we choose to use a gamma random variable with a mean of 12.5 days and a standard deviation of 5.0 days to describe the total length of time spent by an individual in the **I** state, to include

any time spent in the **Q** state. Estimates of the length of the infectious period for this disease have varied considerably, with most studies to date using the positive detection of a virus in an individual (for example, in the throat or stool) as a proxy for the individual being infectious. For example, Ling et al. (2020) finds that the virus is detectable in throat samples for a median time of 9.5 days after symptom onset (so perhaps roughly 11–14 days after the individual has become infectious).

We use an exponential random variable to model the length of time taken for a transmission event across a given edge in the contact network, from an infectious individual to a susceptible one. The reciprocal of the mean of this random variable is sometimes referred to as the “transmission rate” and is represented by β in many models. Because we model transmissions as occurring across a contact network (which is a substantially different model than has been commonly used to date for this virus) comparison of this model parameter is difficult. However, two studies (Fang, Nie, and Penny 2020; Radulescu and Cavanagh 2020) which have comparable interpretations of this parameter to ours both use a value of $\beta = 0.1$, which we also take as our baseline value for this parameter.

Finally, we set the values of the network parameters to $p = 0.047$ and $p^* = 0.011$ to account for the presumed tendency of individuals to reduce the number of people they are in contact with upon entering the quarantined state. We set the probability of two individuals who are both in the **Q** state sharing an edge in the contact network at 0.0067; though as discussed above, this value does not impact the actual spread of the disease and is only needed for the calculation of the reproduction number at the end of the outbreak. These values are somewhat arbitrary, as we are not trying to model the contact patterns of any specific population.

Estimation of reproduction numbers

When assessing disease dynamics, one of the metrics often used to describe the capacity of an epidemic to spread through a population is the **basic reproduction number**, R_0 . This is commonly defined as the number of disease transmissions that can be expected to be caused by an initial infected person in a population of otherwise susceptible individuals (Anderson and May 1992). As the disease spreads through the population, the number of transmissions caused by each infectious individual will tend to decrease, due to both the depletion of susceptible individuals (who have subsequently been infected), as well as any changes in the behaviors of individuals in response to the epidemic. To account for this change through time, we also consider the **effective reproduction number at time t** , R_t , which is the expected number of secondary cases for an infectious individual at time t .

For the network and epidemic models used here, Groendyke, Welch, and Hunter (2012) developed the following formula that can be used to estimate R , based on previous work by Andersson (1998), Meyers (2007), and Kenah (2011):

$$R = \left(\frac{E[D^2]}{E[D]} - 1 \right) \cdot \left(1 - \left[\frac{1}{1 + \beta\theta_1} \right]^{k_1} \right), \quad (3)$$

where D is the random variable governing the degree distribution of the contact network, and k_1 and θ_1 are the parameters of the gamma random variable describing the length of time an individual remains in the **I** state. Here, the change in the network structure due to quarantine will impact the degree distribution of the network, allowing us to assess the change in effective reproduction number from the start of the epidemic (R_0) to the end (R_ω).

Simulation study

This subsection gives the results of our simulation study. We first examine the simulated epidemics produced at the baseline parameter values. Then we analyze the impact of varying the transmission rate, probability of remaining asymptomatic, and mean lengths of time spent in the exposed and infectious states. We vary each of these variables from 50 to 150% of its baseline value to encompass a range of reasonable values for each parameter, holding the values of all other variables constant. We also assess the impact of the network structure by comparing results under the GER model to those under a network model allowing covariates.

Simulations at baseline parameter values

To establish a basis for comparison, we first run 1,000 simulated epidemics through a population of 100 individuals at the baseline parameter values. To get a sense of the trajectory of the epidemic, consider Figure 1A, which shows the number of individuals in the infectious (I) class over time, for each of the simulated epidemics. Most of the simulated epidemics follow a consistent pattern, with a relatively symmetric pattern of increase and decrease in infectious class size, peaking around day 40. (Time 0 is defined as the exposure time for the initially infected individual.) The transmission tree for one sample simulated epidemic is given in Figure 13 in Appendix A. The total lengths of the simulated epidemics are displayed in Figure 1B. We note that about 15% of the epidemics die out after the initially infected individual failed to infect any others, while another 6% only infect one other individual; these cases largely account for the leftmost mode. Of the other epidemics that spread to multiple individuals, the majority last from 70–90 days.

In many places, one of the biggest concerns of the COVID-19 pandemic has been the stress that it has placed on the local health care systems, with ICU beds, hospital staff, and ventilators being in short supply at various times. In this study, we consider two different metrics as proxies for the level of stress induced on the health care system by the epidemic: the maximum size of the infectious class ($I_{\max}$), and the number of days the size of the infectious class exceeds 15% of the population (this threshold is arbitrary, but serves as a useful benchmark); we refer to the latter metric as “stress days” (D_s).

Figure 2 gives histograms of both of these metrics for the simulated epidemics at the baseline parameter values. The size of the infectious class tends to peak at around 25–45% of the population in most cases; we see that the infectious class exceeds 15% of the population for 25–40 days in the bulk of the simulated epidemics.

We also calculate estimates of the effective reproduction numbers at the beginning and end of each of the simulated epidemics. Figure 3 gives a histogram of the calculated reproduction numbers pre- and post-epidemic for each of the simulated epidemics. We can see that the reproduction number tends to decrease significantly from the start to the end of the epidemic, due to the effects of the quarantine. Recall from Eq. (3) that in our model, R is a function of the degree distribution of the underlying contact network; as individuals lose contacts due to entering quarantine, their expected number of contact decreases significantly, hence bringing down the estimated value of the reproduction number. We again see the effect of the portion of epidemics where only a single individual is infected; these are the cases where R does not decrease significantly over the course of the epidemic. We note that our estimated values of R_0 , while broadly reasonable, are slightly higher than those found by some other researchers: Li et al. (2020) estimated a value of 2.2 based on the first

Figure 1: Trajectories of simulated epidemics under baseline parameter values. Each light grey line in (A) represents the size of the I class over time for a single simulated epidemic. A histogram of the total lengths of the epidemics is given in (B).

Figure 2: Impact of epidemic on health care system under baseline parameter values. Histograms of the maximum size of the infectious class (A) and the number of stress days induced by the epidemic (B).

Figure 3: Histogram of estimated reproduction numbers for epidemics simulated under baseline parameter values. Red represents the reproduction number at the start of an epidemic (R_0), while blue represents the reproduction number at the end of the epidemic (R_w).

425 cases in Wuhan, China; Nkwayep et al. (2020) calculated a value of approximately 2.95 using data from Cameroon; Liu et al. (2020) computed a median value of 2.79 from their meta-analysis of the reproduction number for this disease. The bulk of our R_0 values can be seen to fall in the range of 2.8–3.7. However, as mentioned above, in our model this calculation is impacted by the underlying contact network in the population, which we have chosen arbitrarily. Thus, we do not expect the corresponding R_0 values to be realistic estimates for any particular population. Rather, our main interests in calculating these values are to assess their change over the course of the epidemic, as well as to gauge how they vary with the model parameters.

Varying the transmission rate

We first vary the transmission rate β , which describes how quickly the disease can be expected to spread across a given edge in the contact network, from an infectious individual to a susceptible one, from 0.05 to 0.15. The lengths of the epidemics form a bimodal distribution for all values of β , similar to the baseline case (see Figure 1B). For the smaller values of the transmission rate, the leftmost mode is more pronounced, while the opposite is the case for the larger values of β . Table 1 gives the mean and median epidemic lengths for the epidemics simulated under the various values of the transmission rate.

We can see that for the lowest values of the transmission rate, the epidemics are short, because that fewer people are infected. The epidemic lengths rise rapidly with the lower values β , but taper off thereafter (for the larger values of β), as the faster transmission rates infect the population more rapidly. The maximum size of the infectious group rises monotonically (nearly linearly) with the transmission rate (see Figure 4A).

Table 1: Mean and median lengths (in days) of simulated epidemics for various transmission rates. Values for the baseline parameter case are bolded.

Transmission rate	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15
Length (mean)	53.0	63.7	69.2	71.7	71.9	70.0	69.2	67.7	67.4	65.8	64.5
Length (median)	34.0	54.5	80.5	82.0	81.0	78.0	75.0	73.0	71.0	69.0	67.0

Figure 4: Impact of epidemic on health care system for various values of the transmission rate. Boxplots of the maximum size of the infectious class (A) and the number of stress days induced by the epidemic (B). Values for the baseline parameter case are in blue.

However, our other indicator of the impact of the epidemic on the health care system, stress days, levels off at a median of about 30 days for $\beta \geq 0.1$ (see Figure 4B). While higher transmission rates result in more individuals becoming infected, they also cause the epidemic to end (slightly) more rapidly; these two factors roughly offset to keep the number of stress days level for the larger values of β .

We also consider the values of the reproduction number for this disease, at the start and end of each simulated epidemic. Table 2 gives the median values of R_0 and R_ω for the simulated epidemics under the various values of the transmission rate. The shapes of these distributions are similar to those seen in the baseline case; R_0 yields a symmetric, roughly bell-shaped distribution, whereas the distribution of R_ω is bimodal for the same reason as in the baseline case.

We can see that R_0 increases monotonically with β . On the other hand, R_ω drops from values near 2 for the lowest transmission rates, to values near 1.4 for all $\beta \geq 0.07$. In our formulation, the reproduction number is a function of both the degree distribution as well as the transmission rate (see Eq. (3)). As the transmission rate increases, a greater proportion of the population ultimately becomes infected, and hence enters quarantine, reducing their contacts. This has the effect of lowering their expected number of contacts, which largely offsets the direct effect that the increase in β has on the reproduction number.

Table 2: Reproduction numbers at the start and end of simulated epidemics for various transmission rates. Values for the baseline parameter case are bolded.

Transmission rate	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.13	0.14	0.15
R_0 (median)	2.06	2.33	2.57	2.78	2.96	3.12	3.27	3.40	3.51	3.61	3.70
R_ω (median)	1.90	1.99	1.44	1.34	1.34	1.33	1.34	1.35	1.37	1.41	1.40

Varying the probability of remaining asymptomatic

We next vary the probability that a given individual who becomes infected with the disease remains asymptomatic from values of $\alpha = 0.175$ to $\alpha = 0.525$. As α increases, the average length of the epidemics changes little, and retains the same basic bimodal distribution shape as previously seen. However, the dispersion of the length of the epidemics decreases as α increases; see Table 3 for summary statistics of the distributions of epidemic lengths under the various values of α .

Next we consider the impact of varying α on the two metrics we use to assess the strain on the health care system imposed by the epidemic. Table 4 gives the median values of these metrics across the various values of the asymptomatic probability. We can see that in both cases, there is a roughly linear increase with α , though we do note that the number of stress days levels off toward the higher end of the table. Also, while the differences in these metrics across the values of α are significant, they are not particularly large in magnitude, compared to the changes in the parameter α .

We again consider the values of the reproduction number for this disease, at the start and end of each simulated epidemic. Table 5 gives the median values of R_0 and R_ω for the simulated epidemics under the various values of α . The shapes of these distributions are similar to those seen previously; R_0 yields a symmetric, roughly bell-shaped distribution, whereas the distribution of R_ω is again bimodal.

We note that the distribution of R_0 does not vary as the value of α changes. In Eq. (3), we see that R is a function of the degree distribution; this degree distribution will change as individuals enter quarantine, but this does not occur until the epidemic is under way. Hence, we should expect that R_ω will vary with α , whereas R_0 will not, and this is indeed the case. The median value of R_0 , as noted earlier, is slightly higher than most other researchers have calculated or estimated, but is broadly reasonable and is suitable for our purposes. With respect to R_ω , it increases monotonically with α , as we might expect. Specifically, as the proportion of infected individuals who remain asymptomatic increases, fewer people enter quarantine. Each infectious person who fails to enter quarantine continues to have unfettered opportunities to infect other

Table 3: Mean and standard deviation of lengths (in days) of simulated epidemics for various probabilities of remaining asymptomatic. Values for the baseline parameter case are bolded.

α	0.175	0.210	0.245	0.280	0.315	0.350	0.385	0.420	0.455	0.490	0.525
Length (mean)	69.8	70.0	69.7	69.5	69.7	70.0	70.5	71.7	71.3	70.9	70.9
Length (st.dev.)	37.1	36.9	35.8	34.0	33.3	32.1	31.0	31.1	30.1	28.8	28.0

Table 4: Impact of α on health care system strain. Median values of the maximum size of the infectious class and the number of stress days in simulated epidemics for various probabilities of remaining asymptomatic. Values for the baseline parameter case are bolded.

α	0.175	0.210	0.245	0.280	0.315	0.350	0.385	0.420	0.455	0.490	0.525
I_{max} (median)	24	25	27	28	30	31	33	34	35	37	38
D_s (median)	24	25	26	27	28	29	29	30	30	30	30

Table 5: Reproduction numbers at the start and end of simulated epidemics for various probabilities of remaining asymptomatic. Values for the baseline parameter case are bolded.

α	0.175	0.210	0.245	0.280	0.315	0.350	0.385	0.420	0.455	0.490	0.525
R_0 (median)	3.12	3.12	3.12	3.12	3.12	3.12	3.12	3.12	3.12	3.12	3.12
R_ω (median)	1.05	1.07	1.13	1.19	1.24	1.33	1.41	1.49	1.57	1.69	1.75

individuals – that is, the number of edges associated with the individual in the contact network does not decrease – which prevents the reproduction number from dropping.

Varying the length of time spent in exposed state

To assess the effect of the length of time spent by individuals in the **E** state on the dynamics of this disease, we vary the mean of the (lognormal) distribution we use to model this time period from 1.45 to 4.35 days; we adjust the standard deviation in each case in order to maintain a constant coefficient of variation. As the mean length of time spent in the **E** state increases, the duration of the epidemics increases accordingly; this is quite intuitive, as individuals remaining latent for longer would be expected to lengthen the total duration of the epidemic. The changes in epidemic duration are noticeable, but not dramatic. This is also intuitive, considering that the baseline mean exposed period is relatively short (2.9 days). Thus, increasing and decreasing this mean by 50% should not be expected to make a large impact on the dynamics of the disease. Table 6 gives the mean and median epidemic durations for the various mean exposure times.

We find that varying the length of time spent in the **E** does not tend to have a particularly great impact on the strain put on the health care system by the epidemic. In particular, the maximum size of the infectious class shrinks monotonically with the mean length of the exposed period; this is again due to the effect of spreading out the epidemic over a longer period of time (sometimes known as “flattening the curve”). The impact of mean **E** time on the number of stress days per epidemic is more subtle; while the median number of stress days stays roughly constant as the mean exposed time increases, there are an increasing number of epidemics with very few or no stress days, again a sign that the epidemic is being flattened. Figure 5 shows boxplots for these two metrics as a function of the mean **E** time.

Table 6: Mean and median lengths (in days) of simulated epidemics for various mean times in the exposed state. Values for the baseline parameter case are bolded.

Mean E time	1.45	1.74	2.03	2.32	2.61	2.90	3.19	3.48	3.77	4.06	4.35
Length (mean)	59	61	64	66	69	70	74	76	79	80	84
Length (median)	66	67	70	73	76	78	82	84	87	89	93

Figure 5: Impact of epidemic on health care system for various values of the mean time spent in the **E** state. Boxplots of the maximum size of the infectious class (A) and the number of stress days induced by the epidemic (B). Values for the baseline parameter case are in blue.

Based on our model and our methodology for calculating R , we would expect little or no impact on either R_0 or R_ω as a result of varying the mean time spent in the exposed state. Our results indicate that this is indeed the case, with the simulated distributions of both reproduction numbers staying very similar to those produced in the baseline case, regardless of the mean time spent in the **E** state; we see slightly more variability in R_ω than R_0 , which conforms with our intuition, as the former metric reflects some additional variability in the empirical degree distribution due to the effects of quarantine.

Varying the length of time spent in infectious state

To assess the effect of the length of time spent by individuals in the **I** state on the dynamics of this disease, we vary the mean of the (gamma) distribution we use to model this time period from 6.25 to 18.75 days; we adjust the standard deviation in each case in order to maintain a constant coefficient of variation. As the mean length of time spent in the **I** state increases, the duration of the epidemics increases monotonically and roughly linearly. The impact of the time spent in the infectious state on the total length of the epidemic is somewhat greater than that for the exposed state. In addition to the direct impact on epidemic length of the changes in the **I** state, there is a secondary, indirect effect caused by an increase in the number of infectious individuals. This latter effect occurs because when an individual spends a longer amount of time in the **I** state, they have more chances to infect others, thereby contributing to a lengthened epidemic duration. The shape of the distribution of epidemic lengths remains bimodal, as in the other simulations. Table 7 gives the mean and median epidemic lengths for the various mean infectious times.

Unlike the previous section where we vary the length of time spent in the **E** state, we find that varying the length of time spent in the **I** state has a very significant impact on the strain put on the health care system by the epidemic. The maximum size of the infectious class increases substantially with the mean length of the **I** period; this is due both the direct effect of each individual remaining infectious for a longer period of time as well as an indirect effect as the result of a larger number of people becoming infected per epidemic. These effects combine to produce resulting epidemics whose mean $I_{\max}$ varies from 9.2 for the shortest mean infectious time up to 41.0 for the longest scenario. Examining the number of stress days reveals a similar dynamic; this metric is impacted greatly by the changes in mean infectious time, with the mean number of stress days ranging from 2.1 days for the shortest mean infectious time up to 35.4 days for the longest mean infectious time. Figure 6 shows boxplots for these two metrics as a function of the mean **I** time.

When we examine the reproduction number at the start of the epidemic, we see that it increases with the mean **I** length; this is expected, as under our model, the reproduction number is a function of the two parameters governing the length of time that each individual is infectious; see Eq. (3). Counterbalancing this effect in the calculation of R_ω (but not R_0) is the effect that longer periods of infectiousness lead to more individuals becoming infected. This in turn leads to an increase in quarantined individuals, which decreases the mean degree in the underlying contact network. The net result is that R_ω is relatively constant as a function of mean infectious period. Table 8 gives the median values of R_0 and R_ω for the various mean infectious lengths.

Table 7: Mean and median lengths (in days) of simulated epidemics for various mean times in the infectious state. Values for the baseline parameter case are bolded.

Mean I time	6.25	7.50	8.75	10.00	11.25	12.50	13.75	15.00	16.25	17.50	18.75
Length (mean)	47	54	58	62	66	70	74	78	81	84	86
Length (median)	52	62	67	70	74	78	81	85	87	90	92

Figure 6: Impact of epidemic on health care system for various values of the mean time spent in the I state. Boxplots of the maximum size of the infectious class (A) and the number of stress days induced by the epidemic (B). Values for the baseline parameter case are in blue.

Table 8: Reproduction numbers at the start and end of simulated epidemics for various mean times in the infectious state. Values for the baseline parameter case are bolded.

Mean I time	6.25	7.50	8.75	10.00	11.25	12.50	13.75	15.00	16.25	17.50	18.75
R_0 (median)	2.06	2.33	2.57	2.78	2.96	3.12	3.27	3.40	3.51	3.61	3.70
R_ω (median)	1.45	1.29	1.31	1.30	1.31	1.33	1.35	1.37	1.38	1.41	1.46

A generalization of the network model

Here we consider an extension of the GER network model used in the simulation study above. In particular, suppose that rather than assuming that any two individuals i and j have a common probability p of sharing an edge in the underlying population network, we instead assume that the probability of such an edge is given by p_{ij} , where

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \sum_{s=0}^{k-1} \eta_s \mathbf{X}_{\{i,j\},s}, \quad (4)$$

$\mathbf{X}$ is an $\left(\frac{N}{2}\right) \times k$ matrix of dyadic covariates, and $\{\eta_s\}$ is the corresponding vector of k parameters. We note that an assumption of independence is made between dyads; this assumption is not overly restrictive, but does preclude the inclusion of some potentially relevant effects, such as transitivity.

The network model given in Eq. (4) admits many types of covariates; any statistic that can be computed from the characteristics of one or both members of a dyad can be used. We might consider, in some circumstances, homophily (matching) statistics – also sometimes known as “assortativity” effects (Cauchemez et al. 2011) – which take values of 0 or 1, according to whether two individuals have matching values of some characteristic; see McPherson, Smith-Lovin, and Cook (2001) for a thorough discussion of homophily effects in networks. Another possibility is to use the (absolute value of the) difference or distance between the characteristics of the individuals; difference in age and Euclidean distance between residences are a couple of examples of this type of effect. For a more thorough explanation of these and other types of effects that may be included in network models, the reader is referred to Morris, Handcock, and Hunter (2008). We previously described this network model as an extension of the GER model; when the covariate matrix in Eq. (4) is taken to be a single column of ones, the model reduces to the simpler GER network model.

We demonstrate this type of network model via an extension of the simulation study described above. We do this by comparing the results of our computed epidemic statistics under the GER model to the corresponding results under a (non-trivial) network model as described in Eq. (4) (henceforth called the “alternative model”), using the baseline values for all other model parameters. We again start with a hypothetical population of 100 individuals, but also generate additional characteristics for them, including household locations, and group memberships.

The individuals are assigned to 34 distinct households, with the number of individuals in each household generated according to a zero-truncated Poisson distribution. The simulated household sizes are given in Table 9.

The spatial location of each household is simulated via a bivariate normal distribution. These locations are given in Figure 7.

We assign hypothetical group memberships (which we might think of as corresponding to churches, social clubs, etc.) to the households such that five households belong to group 1 and 10 households belong to group 2; two households belong to both groups.

Given these population characteristics, we then consider a network model given by Eq. (4) with the following covariates:

- **Household homophily.** Clearly individuals in the same household are more likely to share an edge in the underlying contact network. This is a “uniform homophily” effect, as we assume that the magnitude of this effect is the same across households.
- **Group homophily.** We also assume that group membership increases the likelihood of disease transmission; we treat this effect, however, as a “differential homophily” effect in that we assume a different magnitude of the effect for common membership in group 1 than common membership in group 2.
- **Spatial distance.** We posit that (Euclidean) physical distance between individuals’ households is inversely related to the probability of disease transmission.

Thus, there are five parameters in the alternative network model, corresponding to an intercept term to measure baseline edge propensity (η_0), household (η_1), membership in groups 1 and 2 (η_2 and η_3 , respectively),

Table 9: Distribution of household sizes.

Household size	1	2	3	4	5	6
Number of households	4	10	10	5	4	1

Figure 7: Plot of spatial locations of households. The size of each circle is proportional to the number of individuals in the corresponding household.

Table 10: Parameter values for alternative network model.

Effect	Baseline	Household	Group 1	Group 2	Distance
Parameter	η_0	η_1	η_2	η_3	η_4
Value	-3.5	3.0	2.5	1.5	-0.025

Figure 8: Degree distributions for 1,000 networks simulated under the GER model (A) and the alternative network model containing covariates (B). Each light grey line represents the observed empirical degree distribution in a single simulation.

and spatial distance (η_4). For our simulation, the parameter values assigned for the initial network are given in Table 10.

And by way of comparison, for the parameterization of the GER given by Eq. (4), we use a lone parameter value of $\eta = -3$. We simulate 1,000 networks from both network models; Figure 8 gives the degree distributions resulting from these simulations, while Figure 9 shows the mean degree frequencies averaged across all simulations.

We note that while the parameter values for these network models were chosen in order to generate network distributions having the same average degree (4.7), the shape of the degree distributions differs between the two models. As expected, the GER model produces a binomial degree distribution, whereas the alternative model has a heavier right tail in its degree distribution. This latter feature is often present in

Figure 9: Histogram of mean degree frequencies observed across all 1,000 networks simulated from each model. Red represents the GER model, and blue corresponds to the alternative network model with covariates.

observed contact networks (Wasserman and Faust 1994) and hence would seem to indicate a more realistic contact structure. An important implication of a heavy right tail in the degree distribution is that these individuals with large numbers of contacts are able to act as “super spreaders” of the virus, and are often significant drivers of the spread of the disease through the population (Ferrari et al. 2006).

We assume that upon an individual entering the $\mathbf{Q}$ state, some portion of the edges they share are removed from the network, as given in Eq. (2). The same phenomenon occurs in our alternative network model, where Eq. (2) is generalized to

$$P[E^* | E] = \frac{p_{ij}^*}{p_{ij}}, \quad (5)$$

where p_{ij} is given by Eq. (4) and

$$\log \left(\frac{p_{ij}^*}{1 - p_{ij}^*} \right) = \sum_{s=0}^{k-1} \eta_s^* \mathbf{X}_{\{i,j\},s}, \quad (6)$$

with $\{\eta_s^*\}$ being the vector of k modified network parameters. For these modified network parameters, we choose to keep the parameters corresponding to household and distance constant, and reduce the baseline, group 1, and group 2 parameters to account for the effects of self-isolation. The specific network parameter values used are given in Table 11. Again, for the sake of comparison, for the parameterization of the GER given by Eq. (4), we use a lone parameter value of $\eta^* = -4.5$.

We simulated epidemics over each of the 1,000 networks generated under each network model, with all of the non-network (η) parameters held at their baseline values. Computing the various metrics discussed above allows us to assess the impact of the network structure on the dynamics of the spread of this disease. We also introduce one additional metric: the day one which the infectious class reaches its maximum size ($D_{\max}$), which can be seen as a measure of how quickly the epidemic has spread. Means and medians of these metrics for the epidemics simulated under the two network models are given in Tables 12 and 13, respectively.

First we consider the size of the infectious class over time. Figure 10 gives the mean size of the $\mathbf{I}$ class for both network models; from this plot, we can see that the disease appears to spread more quickly under the alternative network model, leading to a peak infectious class that is larger in size (by roughly 20%) and comes earlier in the epidemic (by approximately 8 days) than under the GER model.

Table 11: Modified parameter values for alternative network model.

Effect	Baseline	Household	Group 1	Group 2	Distance
Parameter	η_0^*	η_1^*	η_2^*	η_3^*	η_4^*
Value	-3.0	3.0	0.5	0.5	-0.025

Table 12: Means of metrics of interest under network models.

Metric	$I_{\max}$	D_s	R_0	R_ω	Length	$D_{\max}$
GER mean	26.53	23.00	3.14	1.70	69.95	33.25
Alternative network mean	31.38	23.40	3.97	2.30	67.97	29.88

Table 13: Medians of metrics of interest under network models.

Metric	$I_{\max}$	D_s	R_0	R_ω	Length	$D_{\max}$
GER median	31	29	3.12	1.33	78	36
Alternative network median	37	28	3.97	1.97	71	31

Figure 10: Mean size of the infectious (I) class through time, averaged across all 1,000 simulations of the GER model (red) and alternative network model (blue).

Next we look at the length of the epidemics simulated under both network models. The alternative network model tends to produce epidemics that spread more quickly, and hence are somewhat shorter in duration (a median of 74 days) than we see for the GER network model (a median of 78 days). The shape of the distribution of epidemic lengths is not significantly different under the two network models; the alternative network model produces the same pattern of lengths as seen in Figure 1B.

We then consider the two metrics that we use as proxies for the impact of the epidemic on the health care system: the maximum size of the infectious class and the number of stress days induced by the pandemic. Figure 11 gives boxplots of these metrics for the two network models. Consistent with Figure 10, we see that the maximum size of the infectious class is somewhat larger for the alternative model than under the GER. While the mean and median number of stress days are roughly equal under the two network models, the GER model allows for more epidemics that produce small numbers of stress days; there is also a bit less variability in the number of stress days under the alternative network model.

Next we look at the reproduction number at both the beginning and the end of the simulated epidemics under both network models. We can see from Tables 12 and 13 that the reproduction numbers are substantially higher for the alternative network model than under the GER. This is to be expected, considering Eq. (3), since the only difference between the two simulations is the second moment of the network model; in particular, the greater level of dispersion of degrees in the alternative model leads to greater calculated values of both R_0

Figure 11: Boxplots for the maximum size of the infectious class (A) and number of stress days (B) for epidemics simulated under both network models.

Figure 12: Distributions of the day when the infectious class reaches its maximum size under the GER model (red) and the alternative network model (blue).

and R_0 . The shape of the distributions of these reproduction numbers under the alternative network model is not substantively different from that under the GER model; its distribution is slightly more dispersed and, of course, shifted to the right.

Finally, we consider $D_{\max}$, the day at which the infectious class reaches its maximum size over the course of the simulated epidemic. Figure 12 gives plots of the distributions of this metric under the simulations run across both network models. We can see that these two distributions have similar shapes; both are bimodal, with the leftmost mode corresponding to the small portions of the simulations in which the epidemics fail to spread widely through the population. The bulk of the distributions appear in the range of day 25 to day 50, with the maximum occurring about 5 days sooner under the alternative network model than for the GER model, again providing evidence that the epidemics are spreading more quickly under this more complex network model.

Discussion

Because the SARS-CoV-2 virus is relatively new, it has yet to be studied in as much depth as many of the other viruses that commonly spread through populations. However, as more researchers study the spread of this novel coronavirus, the understanding of its properties will continue to be refined, likely resulting in more definitive knowledge regarding model parameter values. Our simulation study gives some indications about where this knowledge will be the most useful in shaping our understanding of the dynamics of this disease.

The transmission rate β can be seen to have a significant impact on the dynamics of this disease, as seen in all of the metrics we calculate. Namely, quicker transmission rates (greater values of β) tend to lead to epidemics having longer durations and ultimately infecting greater numbers of people. We do note, however, that some of the metrics, in particular, the number of stress days and R_0 level off for the larger values of β . That is, the epidemics are sensitive to the value of β for values of this parameter less than about 0.1, but relatively insensitive (by most metrics) to changes in the transmission rate above this point. We also note that in our model infectiousness is a binary property; some other researchers allow for different levels of infectiousness as a function of time, e.g., Lin et al. (2020).

Increasing the probability that an individual remains asymptomatic throughout the course of the disease led to more severe epidemics by all of the metrics we examined. This is a result of fewer individuals quarantining, and hence reducing their contacts, which would lead to fewer infections. However, while the effect of the parameter α is monotonic, the magnitude of the effects – as measured by all of our metrics – is comparatively small. Thus, it appears that more precise knowledge of this parameter may not lead to great changes in the epidemic outcomes for this disease.

As the mean length of time spent in the exposed state increases, the trajectory of the disease tends to be elongated somewhat. That is, the total length of the disease is increased, while the maximum size of the infectious class decreases. We note that these effects are relatively small in magnitude, and that the median

number of stress days is mostly unchanged as the mean **E** time increases. We can conclude that the epidemics involving this disease are relatively insensitive to this particular parameter.

We also see that the length of time spent in the infectious class makes a very significant impact on the dynamics of this disease. With the exception of R_{ω} , all of the metrics we calculate indicate that the severity of an epidemic is quite sensitive to the length of time spent in the **I** state. Hence, it seems likely that better knowledge about this model parameter is very likely to lead to significantly more accurate modeling of the spread of this disease through populations.

Finally, when we look at the simulated epidemics under different network models, we see that using a more complex network model which includes covariates does have a substantial impact on the dynamics of the disease spread, especially in terms of the speed with which the epidemics move through the population. While the number of stress days is not greatly affected by the network model, under the alternative model, the infectious class tends to reach a larger peak size and to reach this size more quickly. This increase in speed is likely due to the presence under the alternative model of super-spreaders in the right tail of the network degree distribution. Also of great importance is the effect on the reproduction number that results from the heavier tailed degree distributions generated by the network model with covariates; a significantly larger reproduction number portends more serious epidemics.

Parameter estimation

The data that can be used to estimate the parameters in the epidemic and network models consist of (1) the set of times at which individual in the population (who was ultimately infected over the course of the epidemic) entered each of the **E**, **I**, **Q**, and **R** states, and (2) information for each individual sufficient to calculate the covariate values for any dyadic covariates included in the network model. While some previous works (Britton and O'Neill 2002; Ray and Marzouk 2008) have explored (often more realistic) situations where only some subsets of the disease progressions are available, we will assume here that we possess full and complete data.

Several previous works that have performed inference for parameters in network-based SEIR epidemic models have used a Bayesian approach. This approach allows for the incorporation of information about model parameters learned from other sources, in the form of the prior distributions assigned to the model parameters; a Bayesian approach in some cases will also offer computational advantages over frequentist methods.

Britton and O'Neill (2002) developed a Markov chain Monte Carlo (MCMC) algorithm to perform Bayesian inference on an SIR model and a GER network, using the sets of **I** and **R** times. They modeled the time spent in the **I** state using an exponential random variable. Ray and Marzouk (2008) generalized this methodology by using an SEIR disease model and using gamma random variables to describe the lengths of time spent in the exposed and infectious states; they also used a somewhat more general network model than the GER model posited by Britton and O'Neill (2002), allowing for two levels of mixing.

Groendyke, Welch, and Hunter (2011) extended this methodology by considering a more general network model of the type given in Eq. (4), again using a stochastic SEIR model. Groendyke and Welch (2018) presented an implementation of an MCMC algorithm that produces the requisite samples from the joint posterior distribution of the parameters. This algorithm includes inference for the underlying network as well as the path taken by the disease through the population (sometimes called the “transmission tree”). These items may or may not be of specific interest but are included in order to simplify the calculation of the likelihood function, which would otherwise be intractable.

While many aspects of this MCMC algorithm could be used in this setting, some aspects would need to be adjusted to accommodate the modifications made to the disease model. The parameters governing the lengths of the **E** and **I** times, the transmission tree, transmission rate, and network parameters, can all be updated via standard Metropolis–Hastings steps, though Gibbs sampling is also possible under specific choices of prior distributions; the specific form of the likelihood will change slightly due to the addition of the **Q** state.

Since the network model in question is a dyadic-independent model, the presence or absence of edges in the network can be updated for each dyad individually and independently of any of the other dyads; the inference would be done with respect to the initial state of the network. When the data are not complete, it

is necessary to infer the missing times, as well as the identity of the initially exposed individual; this would not be necessary under the assumption of full data. Finally, the asymptomatic probability can be estimated independently of the other model parameters.

Previous authors have reported difficulties arising from poor mixing of this type of algorithm; for this reason, it may be necessary to generate a large number of samples from the posterior distribution and apply thinning. It is also worth noting that the run time for this algorithm scales roughly with the square of the size of the population. See Neal and Roberts (2005) for a discussion of the computational impacts of alternative (“non-centered”) parameterizations of these types of models.

Conclusions

The SARS-CoV-2 virus and accompanying COVID-19 pandemic has had an enormous impact across the globe, disrupting many aspects of society, from health outcomes to individual behavior to financial markets. With no effective treatments having yet been discovered, vaccines not yet being widely available to much of the public, and new strains of this virus circulating, it is clearly important to work to improve our understanding of the dynamics of this disease. Our work furthers this effort by implementing a viable alternative to the mean field model, which has been the most utilized framework in the early studies of this pandemic.

In this paper, we have presented a modification to the network-based stochastic SEIR model to account for the effects of quarantine or self-isolation on the contact patterns of individuals in a population. Our model also allows us to incorporate the effects of a percentage of the population remaining asymptomatic during the entirety of their time in the infectious state. Both of these model features are important to the accurate modeling of the SARS-CoV-2 virus.

Using a GER model for the baseline underlying contact network, our simulation study examined the sensitivity of the severity of epidemics to four basic model parameters: transmission rate, probability of remaining asymptomatic, and the mean lengths of time spent in the **E** and **I** states. We found that the length of time spent in the infectious state was the most important driver of the epidemic severity and that the transmission rate was also important; most metrics were far less sensitive to changes in the time spent in the exposed state and the probability of remaining asymptomatic. The metrics examined in the study included the total duration of the epidemic, the maximum size of the infectious class and number of stress days (as proxies for strain on the health care system), and the estimated reproduction number at the start and end of the epidemic. We also examined the effects of modeling the contact structure in the population via a more general network model allowing for dyadic covariates. This more complex network model generated heavier-tailed degree distributions, leading to super-spreaders and hence more rapid spread of the disease through the population.

Our application of this model to the novel coronavirus assumes that individuals will self-isolate or quarantine at the onset of disease symptoms, which will tend to reduce their number of connections to other individuals. We note that our framework is more general, though, and could be used to model any situation where an individual might change their contact patterns, either by adding or removing contacts.

There are several potential directions for future research based on this model, especially in terms of extensions of the model and applications to various types of populations. While we have used relatively simple models to describe the underlying contact network for this study, it has been previously demonstrated (and again shown in our simulation study) that accounting for more accurate contact network patterns can produce more realistic disease dynamics. Hence, an obvious avenue for future work would be to consider the spread of this virus through specific types of populations, incorporating more complex network models. While the more complex network model we used in “A generalization of the network model” allowed for dyadic covariates, it was nonetheless a dyadic-independent model. This model was able to support a heavier tailed degree distribution, which is an important feature, but it did not allow for the incorporation of effects such as transitivity, colloquially, the notion that a friend of my friend is also likely to be my friend. One possibility would be to utilize a more general type of Exponential-family Random Graph Model (ERGM) or p^* model (Holland and Leinhardt 1981; Wasserman and Pattison 1996). This family of models is very flexible (allowing

for dyadic dependence or independence), well-studied (Hunter et al. 2008; Robins et al. 2007), and contains some important special cases, such as the models used in this study.

We also note that in our model formulation, infectiousness is a binary property; when an individual is in the **I** class, they are capable of infecting others, and when they are not in the infectious class, they cannot spread the disease. For many diseases (and it appears that COVID-19 likely falls into this category), it is more realistic to describe infectiousness as a function of time (since being exposed). He et al. (2020) develops an infectiousness profile for this disease, and provides evidence that the level of infectiousness does indeed vary over time. Incorporating this idea would likely result in a more realistic model. While it would be relatively straightforward to incorporate this into simulations, we note that parameter inference would become more difficult in the resulting model.

As the parameter inference described here requires very detailed and complete data for the entire population in question, finding appropriate data can be a challenge. As the COVID-19 pandemic continues to spread, though, it seems likely that such data may become more prevalent (as well as providing informative prior distributions), allowing for inference of the network and epidemic model parameters.

Due to the explicit modeling of each individual's contacts, the framework presented here is most useful for relatively small and closed populations. While the modeling of a country or state is likely not feasible, there are important sub-populations that would likely fit well into this framework; some possibilities include settings like nursing homes, small colleges, prisons, or cruise ships. Our model not only allows researchers to simulate the spread of epidemics through such populations, but also enables explicit testing of various containment strategies that might be implemented in such sub-populations. This type of study would seem to be both important and timely.

Acknowledgments: The authors would like to acknowledge David Welch, who wrote the original code in the *epinet* package (Groendyke and Welch 2018) for simulating an epidemic using a stochastic network-based SEIR model; the code we used in this study was based on this work. We would also like to thank two anonymous reviewers, whose comments and suggestions greatly improved the quality of this manuscript.

Research funding: None declared.

Author contribution: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

Appendix A: Transmission tree for sample simulated epidemic

See figure 13.

Figure 13: Transmission tree for one simulated epidemic in a population of 100 individuals at baseline parameter values. Each horizontal line segment represents an infected individual. The black portion represents the time in the exposed state, the red portion represents the time in the infectious (but not quarantined) state, and the blue portion represents the time in the quarantine state. Black vertical dotted lines represent transmission events.

References

- Aleta, A., D. Martin-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, Jr, S. Merler, A. Pentland, A. Vespignani, E. Moro, and Y. Moreno. 2020. "Modeling the Impact of Social Distancing, Testing, Contact Tracing and Household Quarantine on Second-Wave Scenarios of the COVID-19 Epidemic." *Nature Human Behavior* 4: 964–71.
- Anderson, R. M., and R. May. 1992. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press.
- Andersson, H. 1998. "Limit Theorems for a Random Graph Epidemic Model." *Annals of Applied Probability* 8: 1331–49.
- Aron, J. L., and I. B. Schwartz. 1984. "Seasonality and Period-Doubling Bifurcations in an Epidemic Model." *Journal of Theoretical Biology* 110: 665–79.
- Bailey, N. 1950. "A Simple Stochastic Epidemic." *Biometrika* 37: 193–202.
- Bansal, S., J. Read, B. Pourbohloul, and L. A. Meyers. 2010. "The Dynamic Nature of Contact Networks in Infectious Disease Epidemiology." *Journal of Biological Dynamics* 4: 478–89.
- Barthelemy, M., A. Barrat, R. Pastor-Satorras, and A. Vespignani. 2005. "Dynamical Patterns of Epidemic Outbreaks in Complex Heterogeneous Networks." *Journal of Theoretical Biology* 235: 275–88.
- Braun, B., B. Taraktas, B. Beckage, and J. Molofsky. 2020. "Simulating Phase Transitions and Control Measures for Network Epidemics Caused by Infections with Presymptomatic, Asymptomatic, and Symptomatic Stages." *PloS One* 15: e0238412.
- Britton, T., and P. O'Neill. 2002. "Bayesian Inference for Stochastic Epidemics in Populations with Random Social Structure." *Scandinavian Journal of Statistics* 29: 375–90.
- Cauchemez, S., A. Bhattarai, T. L. Marchbanks, R. P. Fagan, S. Ostroff, N. M. Ferguson, D. Swerdlow, and the Pennsylvania H1N1 Working Group. 2011. "Role of Social Networks in Shaping Disease Transmission during a Community Outbreak of 2009 H1N1 Pandemic Influenza." *Proceedings of the National Academy of Sciences* 108: 2825–30.
- Center for Disease Control and Prevention. 2020. "COVID-19 Pandemic Planning Scenarios." <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html#table-2> (accessed June 22, 2020).
- Di Domenico, L., G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza. 2020. "Impact of Lockdown on COVID-19 Epidemic in Île-De-France and Possible Exit Strategies." *BMC Medicine* 18: 240.
- Dukic, V., H. F. Lopes, and N. G. Polson. 2012. "Tracking Epidemics with Google Flu Trends Data and a State-Space SEIR Model." *Journal of the American Statistical Association* 107: 1410–26.
- Erdős, P., and A. Rényi. 1959. "On Random Graphs." *Publicationes Mathematicae Debrecen* 6.
- Fang, Y., Y. Nie, and M. Penny. 2020. "Transmission Dynamics of the COVID-19 Outbreak and Effectiveness of Government Interventions: A Data-Driven Analysis." *Journal of Medical Virology* 92: 645–59.
- Ferrari, M. J., S. Bansal, L. A. Meyers, and O. N. Bjørnstad. 2006. "Network Frailty and the Geometry of Herd Immunity." *Proceedings of the Royal Society B: Biological Sciences* 273: 2743–8.
- Gilbert, E. "Random Graphs." *The Annals of Mathematical Statistics* 1959, 1141–4. <https://doi.org/10.1214/aoms/1177706098>.
- González-Parra, G., A. J. Arenas, and B. M. Chen-Charpentier. 2014. "A Fractional Order Epidemic Model for the Simulation of Outbreaks of Influenza A (H1N1)." *Mathematical Methods in the Applied Sciences* 37: 2218–26.
- Grais, R. F., J. H. Ellis, and G. E. Glass. 2003. "Assessing the Impact of Airline Travel on the Geographic Spread of Pandemic Influenza." *European Journal of Epidemiology* 18: 1065–72.
- Grenfell, B. 1992. "Chance and Chaos in Measles Dynamics." *Journal of the Royal Statistical Society: Series B (Methodological)* 54: 383–98.
- Groendyke, C., and D. Welch. 2018. "Epinet: An R Package to Analyze Epidemics Spread across Contact Networks." *Journal of Statistical Software* 83: 1–22.
- Groendyke, C., D. Welch, and D. R. Hunter. 2011. "Bayesian Inference for Contact Networks Given Epidemic Data." *Scandinavian Journal of Statistics* 38: 600–16.
- Groendyke, C., D. Welch, and D. R. Hunter. 2012. "A Network-Based Analysis of the 1861 Hagelloch Measles Data." *Biometrics* 68: 755–65.
- He, X., E. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, and G. M. Leung. 2020. "Temporal Dynamics in Viral Shedding and Transmissibility of COVID-19." *Nature Medicine* 26: 672–5.
- Hethcote, H. W., and D. W. Tudor. 1980. "Integral Equation Models for Endemic Infectious Diseases." *Journal of Mathematical Biology* 9: 37–47.
- Hoertel, N., M. Blachier, C. Blanco, M. Olsson, M. Massetti, M. Sánchez Rico, F. Limosin, and H. Leleu. 2020. "A Stochastic Agent-Based Model of the SARS-CoV-2 Epidemic in France." *Nature Medicine* 26: 1417–21.
- Holland, P., and S. Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76: 33–50.
- Hou, C., J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, J. Zhang, G. Xu, and E. Jia. 2020. "The Effectiveness of Quarantine of Wuhan City against the Corona Virus Disease 2019 (COVID-19): A Well-Mixed SEIR Model Analysis." *Journal of Medical Virology* 92: 841–48.

- Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau, and M. Morris. 2008. “ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks.” *Journal of Statistical Software* 24: nihpa54860.
- Iwata, K., and C. Miyakoshi. 2020. “A Simulation on Potential Secondary Spread of Novel Coronavirus in an Exported Country Using a Stochastic Epidemic SEIR Model.” *Journal of Clinical Medicine* 9: 944.
- Keeling, M., and K. Eames. 2005. “Networks and Epidemic Models.” *Journal of the Royal Society Interface* 2: 295.
- Keeling, M. J., and P. Rohani. 2011. *Modeling Infectious Diseases in Humans and Animals*. Princeton, NJ: Princeton University Press.
- Kenah, E. 2011. “Contact Intervals, Survival Analysis of Epidemic Data, and Estimation of R_0 .” *Biostatistics* 12: 548–66.
- Kermack, W., and A. McKendrick. 1927. “A Contribution to the Mathematical Theory of Epidemics.” *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character (1905-1934)*, 115: 700–21.
- Kerr, C. C., R. M. Stuart, D. Mistry, R. G. Abeysuriya, G. Hart, K. Rosenfeld, P. Selvaraj, R. C. Nuñez, B. Hagedorn, L. George, A. Izzo, A. Palmer, D. Delport, C. Bennette, B. Wagner, S. Chang, J. A. Cohen, J. Panovska-Griffirhs, M. Jastrzebski, A. P. Oron, E. Wenger, M. Famulare, and D. J. Klein. 2020. “Covasim: An Agent-Based Model of COVID-19 Dynamics and Interventions.” medRxiv.
- Kuniya, T. 2020. “Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020.” *Journal of Clinical Medicine* 9: 789.
- Li, M. Y., H. L. Smith, and L. Wang. 2001. “Global Dynamics of an SEIR Epidemic Model with Vertical Transmission.” *SIAM Journal on Applied Mathematics* 62: 58–69.
- Li, Q., X. Guan, P. Wu, and X. Wang. 2020. “Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia.” *The New England Journal of Medicine* 382: 1199–207.
- Lin, Q., S. Zhao, D. Gao, Y. Lou, S. Yang, S. S. Musa, M. H. Wang, Y. Cai, W. Wang, L. Yang, and D. He. 2020. “A Conceptual Model for the Outbreak of Coronavirus Disease 2019 (COVID-19) in Wuhan, China with Individual Reaction and Governmental Action.” *International Journal of Infectious Diseases* 93: 211–16.
- Ling, Y., S.-B. Xu, Y.-X. Lin, D. Tian, Z.-Q. Zhu, F.-H. Dai, F. Wu, Z.-G. Song, W. Huang, J. Chen, B.-J. Hu, S. Wang, E.-Q. Mao, L. Zhu, W.-H. Zhang, and H.-Z. Lu. 2020. “Persistence and Clearance of Viral RNA in 2019 Novel Coronavirus Disease Rehabilitation Patients.” *Chinese Medical Journal* 133: 1039–43.
- Linton, N., T. Kobayashi, Y. Yang, and K. Hayashi. 2020. “Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data.” *Journal of Clinical Medicine* 9: 538.
- Liu, Y., A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. 2020. “The Reproductive Number of COVID-19 Is Higher Compared to SARS Coronavirus.” *Journal of Travel Medicine* 27: taaa021.
- López, L., and X. Rodo. 2020. “A Modified SEIR Model to Predict the COVID-19 Outbreak in Spain and Italy: Simulating Control Scenarios and Multi-Scale Epidemics.” Available at SSRN 3576802.
- McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27: 415–44.
- Meyers, L. 2007. “Contact Network Epidemiology: Bond Percolation Applied to Infectious Disease Prediction and Control.” *Bulletin of the American Mathematical Society* 44: 63–86.
- Meyers, L., B. Pourbohloul, M. Newman, D. Skowronski, and R. Brunham. 2005. “Network Theory and SARS: Predicting Outbreak Diversity.” *Journal of Theoretical Biology* 232: 71–81.
- Momoh, A., M. Ibrahim, I. Uwanta, and S. Manga. 2013. “Mathematical Model for Control of Measles Epidemiology.” *International Journal of Pure and Applied Mathematics* 87: 707–17.
- Morris, M., M. S. Handcock, and D. R. Hunter. 2008. “Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects.” *Journal of Statistical Software* 24: 1548.
- Neal, P., and G. Roberts. 2005. “A Case Study in Non-centering for Data Augmentation: Stochastic Epidemics.” *Statistics and Computing* 15: 315–27.
- Nishiura, H., T. Kobayashi, T. Miyama, A. Suzuki, S.-m. Jung, K. Hayashi, R. Kinoshita, Y. Yang, B. Yuan, A. R. Akhmetzhanov, and N. M. Linton. 2020. “Estimation of the Asymptomatic Ratio of Novel Coronavirus Infections (COVID-19).” *International Journal of Infectious Diseases* 94: 154.
- Nkwayep, C. H., S. Bowong, J. Tewa, and J. Kurths. 2020. “Short-term Forecasts of the COVID-19 Pandemic: Study Case of Cameroon.” *Chaos, Solitons & Fractals* 140: 110106.
- Peng, L., W. Yang, D. Zhang, C. Zhuge, and L. Hong. 2020. “Epidemic Analysis of COVID-19 in China by Dynamical Modeling.” arXiv preprint arXiv:2002.06563.
- Prasse, B., M. A. Achterberg, L. Ma, and P. Van Mieghem. 2020. “Network-based Prediction of the 2019-nCoV Epidemic Outbreak in the Chinese Province Hubei.” arXiv preprint arXiv:2002.04482.
- Radulescu, A., and K. Cavanagh. 2020. “Management Strategies in a SEIR Model of COVID-19 Community Spread.” arXiv preprint arXiv:2003.11150.

- Ray, J., and Y. Marzouk. 2008. "A Bayesian Method for Inferring Transmission Chains in a Partially Observed Epidemic." In *Proceedings of the Joint Statistical Meetings: Conference Held in Denver, Colorado, August 3-7, 2008*. American Statistical Association.
- Robins, G., P. Pattison, Y. Kalish, and D. Lusher. 2007. "An Introduction to Exponential Random Graph (p^*) Models for Social Networks." *Social Networks* 29: 173–91.
- Schwartz, I. B., and H. Smith. 1983. "Infinite Subharmonic Bifurcation in an SEIR Epidemic Model." *Journal of Mathematical Biology* 18: 233–53.
- Sewell, D. K., and A. Miller. 2020. "Simulation-free Estimation of an Individual-Based SEIR Model for Evaluating Nonpharmaceutical Interventions with an Application to COVID-19 in the District of Columbia." *PloS One* 15.
- Shi, P., S. Cao, and P. Feng. 2020. "SEIR Transmission Dynamics Model of 2019 nCoV Coronavirus with Considering the Weak Infectious Ability and Changes in Latency Duration." medRxiv.
- The New York Times. 2020. "Coronavirus Map: Tracking the Global Outbreak." <https://www.nytimes.com/interactive/2020/world/coronavirus-maps.html> (accessed December 21, 2020).
- Volz, E. 2008. "SIR Dynamics in Random Networks with Heterogeneous Connectivity." *Journal of Mathematical Biology* 56: 293–310.
- Wan, K., J. Chen, C. Lu, L. Dong, Z. Wu, and L. Zhang. 2020. "When will the Battle against Novel Coronavirus End in Wuhan: A SEIR Modeling Analysis." *Journal of Global Health* 10: 011002.
- Wang, C., L. Liu, X. Hao, H. Guo, Q. Wang, J. Huang, N. He, H. Yu, X. Liu, A. Pan, S. Wei, and T. Wu. 2020. "Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China." MedRxiv.
- Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, England: Cambridge University Press.
- Wasserman, S., and P. Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^* ." *Psychometrika* 61: 401–25.
- Wearing, H. J., P. Rohani, and M. J. Keeling. 2005. "Appropriate Models for the Management of Infectious Diseases." *PLoS Medicine* 2: e320.
- Zhang, J., M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, and H. Yu. 2020. "Changes in Contact Patterns Shape the Dynamics of the COVID-19 Outbreak in China." *Science* 368: 1481–6.