

MODELING AND ANALYSIS OF THE SPREAD OF COVID-19 UNDER A MULTIPLE-STRAIN MODEL WITH MUTATIONS

Osman Yağan^{†,1}, Anirudh Sridhar^{‡,2}, Rashad Eletreby^{◇,3}, Simon A. Levin^{‡,4}
Joshua B. Plotkin^{§,5}, H. Vincent Poor^{‡,6}

[†] Carnegie Mellon University

[‡] Princeton University

[◇] Walmart Labs

[§] University of Pennsylvania

ABSTRACT. Since December 2019, COVID-19 has caused worldwide devastation. To devise effective countermeasures, it is important to develop mathematical models that help us to understand and predict the spreading of COVID-19, as well as to provide guidelines on what can be done to limit its spread. To this end, we leverage recent work of Eletreby et al. (2020) which studies a model where multiple strains of a virus propagate through a network while also undergoing mutations. Highlighting the recent reports on a mutation of SARS-CoV-2 that is thought to be more transmissible than the original strain, we discuss the importance of incorporating mutations and evolutionary adaptations in epidemic models. We also demonstrate how the results of Eletreby et al. (2020) can be used to assess the effectiveness of mask-wearing in limiting the spread of COVID-19. These are supported by simulation results showing the impact of various mutation and mask-wearing possibilities.

Keywords: COVID-19, mutations, mask-wearing

MEDIA SUMMARY

The spread of COVID-19 has led to a massive loss of life around the world since December 2019. Various countermeasures against the virus have been implemented, resulting in the closure of schools and businesses worldwide. It is of paramount importance to develop mathematical models that can accurately predict the spread of COVID-19 so that the virus can be contained and restrictions can be safely lifted.

An important fact missing from most epidemiology models is that viruses may mutate over time, leading to several strains with potentially different transmission properties, including highly infectious strains; recent reports show that this has already happened in SARS-CoV-2. Our recent work (Eletreby et al., 2020) aimed to bridge this gap by studying a model where a virus may mutate into a different strain when it infects a host. There, we established analytical results that

¹oyagan@andrew.cmu.edu, ²anirudhs@princeton.edu, ³reletreby@cmu.edu, ⁴slevin@princeton.edu,

⁵jplotkin@sas.upenn.edu, ⁶poor@princeton.edu

accurately predict the number of individuals infected by each strain, whereas models that do not consider the possibility of mutations are shown to give incorrect predictions.

In this article, we demonstrate how our recent work on predicting the spread of a virus with mutations can be utilized to address some important questions concerning COVID-19. We examine several scenarios where a virus may mutate into more infectious strains, and illustrate through simulations how the epidemic spreads in each of these cases. We also show how the multiple-strain model can be used to assess the effectiveness of mask-wearing in limiting the spread of COVID-19.

1. INTRODUCTION

From its start in December 2019 in Wuhan, China, the novel coronavirus (known to cause a respiratory disease known as COVID-19) has spread rapidly and broadly and has since devastated a significant fraction of the world population. The World Health Organization (WHO) has identified the spread of COVID-19 as a *pandemic*, and as of February 2021, over 100 million individuals were infected, with about 2.5 million dying of the disease. In addition, the spread of the virus and the countermeasures taken against it have severely impacted the economy, with industries such as tourism, travel and entertainment suffering the most. Schools at all levels are closed in several countries around the world; and highly anticipated events, including the Tokyo 2020 Summer Olympics and Euro 2020 Championship, were cancelled. In summary, the spread of COVID-19 is among the most catastrophic events affecting the health and well-being of humans worldwide since World War II.

A key scientific goal concerning COVID-19 is to develop mathematical models that help us to understand and predict its spreading behavior, as well as to provide guidelines on what can be done to limit its spread. With tight restrictions on traveling, large gatherings, and commercial entertainment in place across many jurisdictions, another important question is to understand the order and time in which these restrictions can be *safely* eliminated.

In the recent work (Eletreby et al., 2020), a class of spreading processes (including information propagation in online social networks and virus propagation) was studied under a *multiple-strain* model with mutations; i.e., a mutation may take place at each host leading to a different strain of the virus/information with different transmissibility from the originally acquired strain. Their work aimed to bridge the disconnect between how spreading processes propagate and evolve in real life versus mathematical and simulation models that ignore evolutionary adaptations. The results in Eletreby et al. (2020) are shown to predict accurately the epidemic threshold, expected epidemic size, and the expected fraction of individuals infected by each strain in this model. The key finding is that classical epidemic models that do not consider the evolution of the strain lead to incorrect predictions on the spreading dynamics when mutations that affect transmission are present.

The purpose of the current paper is to discuss how the findings in Eletreby et al. (2020) can help address some key questions concerning the spread of COVID-19. We highlight the recent reports on a mutation of SARS-CoV-2 that is thought to be more transmissible than the original strain, and discuss the importance of incorporating mutation and evolutionary adaptations (together with the network structure) in epidemic models. We also demonstrate how the multiple-strain transmission model studied in Eletreby et al. (2020) can be used to assess the effectiveness of mask-wearing in limiting the spread of COVID-19. Finally, we present simulation results on a few sample cases

to demonstrate our ideas and the utility of the findings of Eletreby et al. (2020) in the context of COVID-19.

The rest of the paper is organized as follows. In Section 2, we present the model studied by Eletreby et al. (2020) and summarize its main results. There, we also summarize some recent reports on the mutations that SARS-CoV-2 has exhibited so far. In Section 3, we present our ideas on how the model and findings of Eletreby et al. (2020) can be utilized to better understand the spread of COVID-19. There, we show, through an analogy between the multiple-strain virus propagation and a single-strain propagation where some individuals wear a mask, that the results in Eletreby et al. (2020) can help assess the effectiveness of mask-wearing in the spread of COVID-19. Finally, in Section 4 we present numerical results to demonstrate our main ideas in a few sample cases. We remark that our forthcoming conference publication (Sridhar et al., 2021) will support our work in this paper by providing an abridged discussion and additional simulations to validate our claims.

2. MODELING THE SPREAD OF A VIRUS

Modeling the spread of epidemics. What causes an outbreak of a disease? How can we predict its emergence and control its progression? Over the past several decades, multidisciplinary research efforts have been converging to tackle the above questions, aiming for providing a better understanding of the intricate dynamics of disease propagation and accurate predictions on its course (Anderson et al., 1992; Barabási, 2016; Daszak et al., 1999; Fraser et al., 2004; Granell et al., 2014; Lloyd-Smith et al., 2005; Moreno et al., 2002; Morens et al., 2004; Newman, 2002; Pastor-Satorras & Vespignani, 2001; Wei et al., 1995; Wolfe et al., 2007). At the heart of these research efforts is the development of mathematical models that provide insights on predicting, assessing, and controlling potential outbreaks (Brauer et al., 2012; Diekmann & Heesterbeek, 2000; Keeling & Rohani, 2011; Siettos & Russo, 2013). The early mathematical models relied on the *homogeneous mixing* assumption, meaning that an infected individual is equally likely to infect any other individual in the population, without regard to her location, age, or the people with whom she interacts. A finer approach is taken by *metapopulation models*, where the population is divided into several sub-populations in which the epidemic may have different propagation characteristics or may interact in different ways (Hanski & Hanski, 1999; Keeling & Rohani, 2002; Watts et al., 2005). More recently, *network epidemics* has emerged as a mathematical modeling approach that takes the underlying contact network between individuals into consideration (Allard et al., 2009; Barabási, 2016; Keeling & Eames, 2005; Miller & Kiss, 2014; Newman, 2002; Pastor-Satorras et al., 2015). The main goal of these mathematical models is to characterize the speed and scale of propagation and to provide insights into how the parameters of a disease, e.g., its *basic reproductive number* denoted by R_0 , can be used to predict the ultimate reach of the virus. In a nutshell, R_0 is defined as the mean number of secondary infections in a *naïve* population, i.e., the expected number of infections directly generated by one individual in a population where all individuals are susceptible to infection.

Evolution of infectious diseases. A common theme among the proposed models for network epidemics is the assumption that the virus is transferred across individuals without going through any modification or *evolution* (Anderson et al., 1992; Balhrop et al., 2004; Dodds & Watts, 2004; Newman et al., 2002; Qian et al., 2012; Sahneh et al., 2013; Yağan & Gligor, 2012; Yağan et al., 2013; Zhuang & Yağan, 2016). However, in real-life spreading processes, pathogens *evolve* in response to changing environments and medical interventions (Alexander & Day, 2010; Antia et al., 2003;

Leventhal et al., 2015; Morens et al., 2004; Pfennig, 2001). Although the vast majority of molecular changes are neutral or deleterious, there can be strong selection to promote the spread of rare adaptive changes in pathogens. In fact, 60% of the (approximately) 400 emerging infectious diseases that have been identified since 1940 are zoonotic¹ (Jones et al., 2008; Morse et al., 2012a). A zoonotic pathogen is usually poorly adapted, poorly replicated, and inefficiently transmitted when first introduced into the human population (Parrish et al., 2008), but it may eventually acquire and fix mutations that allow it to spread more efficiently from human to human (Jones et al., 2008; Morens et al., 2004; Morse et al., 2012b; Pfennig, 2001; Plowright et al., 2017; Woolhouse et al., 2005). Evolutionary adaptations are often the result of amino-acid substitutions in pathogen proteins that facilitate host cell binding, entry, and release. Some pathogens, such as seasonal influenza viruses, undergo ongoing amino-acid substitutions in immunogenic proteins that facilitate escape from immunity in the host population. But the molecular basis of adaptation can also include recombination and reassortment (e.g., *H5N1 influenza*) as well as hybridization (e.g., *Phytophthora alni*) (Woolhouse et al., 2005). In fact, a key event that caused the emergence of the 1918 H1N1 influenza pandemic was a re-assortment of viral RNA segments during co-infection, resulting in a novel virus with increased infectivity and virulence (Klempner & Shapiro, 2004).

The emergence of COVID-19 provides an ongoing example of pathogen evolution, and it highlights the role of molecular evolution in facilitating pathogen establishment in a new host species. Originally a non-human disease, the ability of COVID-19 to undergo animal-to-human transmission and eventually human-to-human transmission was paved by mutation and selection that produced a strain that is well-adapted to the human host. In addition to the original mutation that led COVID-19 to spread among humans, there is evidence (Long et al., 2020; Tang et al., 2020) that the novel coronavirus has already diverged into distinct lineages with functional differences. Detailed molecular studies (Zhang et al., 2020) have identified a specific point mutation in the spike protein that allows the virus to infect host cells more readily; and this mutation also appears to dominate other variants epidemiologically. In the coming weeks and months, COVID-19 may evolve a greater variety of functional variants, including functions related to transmissibility, virulence, and even, eventually, vaccine escape. Functionally distinct strains of COVID-19 may co-evolve or compete with one another.

Multiple-strain model with mutations. In Eletreby et al. (2020), we studied the (inhomogeneous) multiple-strain² model and characterized the spread and evolution of a pathogen on a contact network. In particular, we i) developed a mathematical theory that characterizes the epidemic threshold, expected epidemic size and the expected fraction of individuals infected by each strain; ii) validated our results on a real-world contact network (a contact network among students, teachers, and staff at a US high school (Salathé et al., 2010) and a contact network among professional staff and patients in a hospital in Lyon, France (Vanhems et al., 2013)); and iii) provided a detailed

¹A zoonosis is any disease or infection that is naturally transmissible from vertebrate animals to humans.

²At a high level, strains represent homogeneous groups within species (Balmer & Tanner, 2011) and they generally possess unique features such as virulence, infectivity, growth rate, etc.

analysis of the case in which co-infection is possible. The model considered in our work³ is particularly appropriate for pathogens with short infectious periods and high mutation and generation rates, e.g., RNA viruses (Grenfell et al., 2004) such as COVID-19.

The multiple-strain model (Alexander & Day, 2010) can be briefly outlined for the two-strain case as follows; cases with arbitrary number m of strains can be modeled similarly. Consider a spreading process that starts with an individual, i.e., the seed, receiving infection (from an external reservoir) with strain-1 of a particular pathogen. The seed infects each of her contacts independently with probability T_1 , called the *transmissibility* of strain-1. Once a susceptible individual receives the infection from the seed, the pathogen may evolve within that new host prior to any subsequent infections. In particular, the pathogen may remain as strain-1 with probability μ_{11} or mutate to strain-2 (that has transmissibility T_2) with probability $\mu_{12} = 1 - \mu_{11}$. If the pathogen remains as strain-1 (respectively, mutates to strain-2) within a newly infected host, then that host infects each of her susceptible neighbors in the subsequent stages independently with probability T_1 (respectively, T_2). As the process continues to grow, if any susceptible individual receives strain-1, the pathogen may remain as strain-1 with probability μ_{11} or mutate to strain-2 with probability $\mu_{12} = 1 - \mu_{11}$ prior to subsequent infections. Similarly, if any susceptible individual receives strain-2, the pathogen may remain as strain-2 with probability μ_{22} or mutate to strain-1 with probability $\mu_{21} = 1 - \mu_{22}$ prior to subsequent infections. The process continues to grow until no additional infections are possible; see Figure 1. The key contribution of Eleteby et al. (2020) is the mathematical theory that enables calculating the number of individuals who will be infected by strain-1 and strain-2, respectively, as a function of the transmissibility parameters (i.e., T_1 and T_2), mutation probabilities (i.e., μ_{11} and μ_{22}), and the structure of the underlying contact network (e.g., its degree distribution).

It is important to note that the transmissibilities and mutation probabilities represent *average* values. In reality, the probability that individual A transmits the virus to individual B depends on numerous factors, such as the number of times A and B interact within close physical proximity and the amount of time A remains infectious. While the relatively simpler multiple-strain model described above abstracts out these complexities through the use of the transmissibilities and mutation probabilities, it can accurately compute quantities such as the probability that an epidemic emerges, the epidemic threshold, and the final size of the epidemic (see, e.g., (Newman, 2002)) in more complex models. On the other hand, the model involving transmissibilities does not capture the *time-varying behavior* of the epidemic, such as the location and size of the epidemic peak. We elaborate on models that can capture time-varying behavior in Section 4.

Modeling the contact network. In modeling the underlying contact network, we utilize *random graphs with arbitrary degree distribution* generated by the configuration model (Molloy & Reed, 1995; Newman et al., 2001). The configuration model generates random graphs with specified degree

³An important focus of Eleteby et al. (2020) was to understand the role of mutation and evolution in the spread of *information* in social networks. The similarity between evolution of viruses and information was inspired by the mechanism of *natural selection*. For example, a virus may go through a number of mutations, which is mostly a random process. However, among them, the mutation that leads to highest infectiousness may be “naturally selected” and become prevalent. A similar process can be envisioned for the spread of information. Several people can attempt to modify a piece of information to make it more attractive, e.g., to get more “likes” or “retweets”. Then, it will be the *audience* who may “select” the most appealing version of the information, and make it much more prevalent than others.

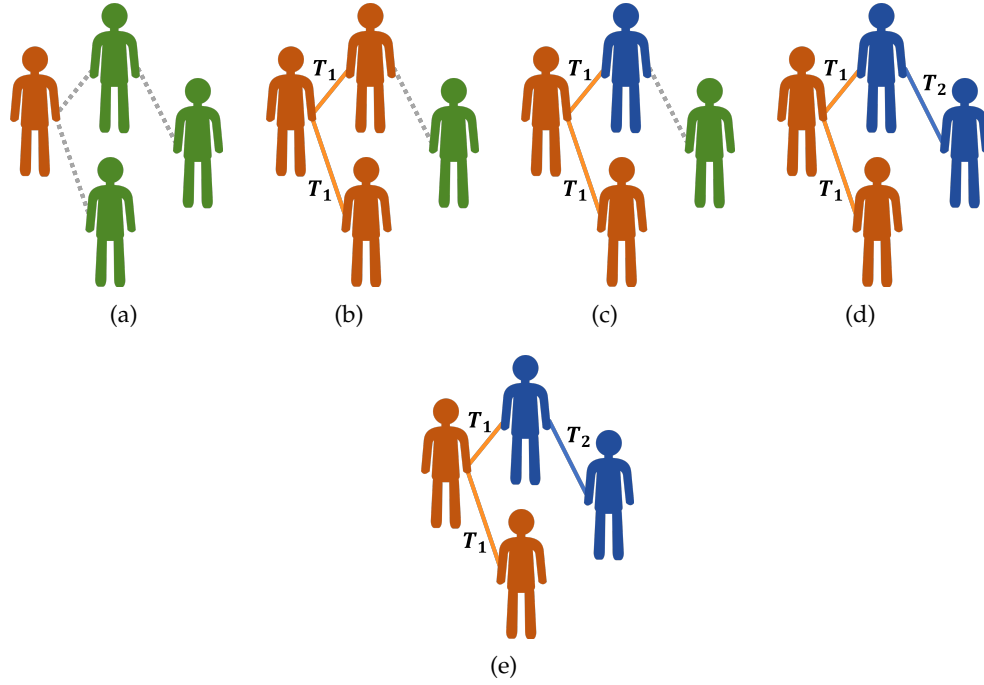


Figure 1. The multiple-strain model for evolution. (a) The process starts with a single individual, i.e., the seed, receiving infection with strain-1 (highlighted in orange) from an external reservoir. (b) The seed infects each of her susceptible neighbors (highlighted in green) independently with probability T_1 . (c) The pathogen mutates independently within hosts. The pathogen remains as strain-1 with probability μ_{11} or mutates to strain-2 (highlighted in blue) with probability μ_{12} . (d) Individuals whose pathogen has mutated to strain- i infect their neighbors independently with probability T_i . (e) The pathogen mutates independently within hosts. The pathogen remains as strain-2 with probability μ_{22} or mutates to strain-1 with probability μ_{21} . Figure (e) demonstrates the case where no second mutation took place.

sequence (sampled from an arbitrary degree distribution), but are otherwise random, by taking a uniformly random matching on the half-edges of the specified degree sequence. The model provides a tractable mathematical framework that allows the investigation of several key properties related to the spreading process and how it interacts with the structure of the underlying graph, as specified by its degree distribution. In addition, since the model could match the degree sequence of real-world social networks, it would essentially generate graphs that resemble such real-world networks to some extent.

3. UTILIZING THE MULTIPLE-STRAIN MODEL WITH MUTATIONS TO BETTER UNDERSTAND THE SPREAD OF COVID-19.

As mentioned earlier, the main goal of this paper is to discuss potential ways that the model and results in Eletreby et al. (2020) can help shed light on the spread of COVID-19. We believe that the multiple-strain model of spreading processes can be utilized towards i) understanding

the impact of potential mutations in COVID-19, and ii) assessing the impact of some mitigation strategies (e.g., mask-wearing) on the spread of COVID-19. Below, we explain in more detail how to achieve these goals, and present sample simulation results in Section 4.

3.1. Understanding the potential impact of mutations. In light of the discussion above, we believe it is of utmost importance to incorporate potential mutations and evolution into the mathematical models used for predicting the future spread of COVID-19. This may help us better prepare for different mutation scenarios, including worst-cases for the current or future pandemics. For example, it may help us understand the effects of a new strain emerging with significantly different spreading characteristics than existing ones. In Section 4, we provide several examples from simulations under different mutation scenarios to demonstrate these ideas.

3.2. Understanding the impact of mitigation strategies. In addition to understanding the potential impact of different mutation scenarios, we believe that the multiple-strain model (and the accompanying results) can also help in evaluating the impact of *some* mitigation strategies implemented to slow down the spread. At a high level, this is motivated by the intuitive similarity between “a virus mutating and becoming less easy to infect individuals” and “people following mitigation measures (e.g., social distancing, wearing masks, etc.) and thus becoming less easy to contract the disease”. The analogy here is strengthened by the fact that those who do not obey mitigation strategies are more likely to infect each other (in a similar manner to carrying a different strain of the virus) than those who are obeying the guidelines.

Below, we formally show how the impact of people wearing masks can be captured by the multiple-strain model in Eletreby et al. (2020). In Section 4, we provide simulation results for a few cases to demonstrate this analogy and show that the analytical results in Eletreby et al. (2020) can be directly used to determine the *probability of emergence* under the *mask model* introduced below.

Mapping the impact of wearing masks to the mutation model. We introduce a potential approach to understand the impact of mask wearing through the mutation model introduced in Eletreby et al. (2020). First, some notation is defined. Let $T_{NN}, T_{NM}, T_{MN}, T_{MM}$ denote the probabilities that an infectious individual will transfer the virus to each of their contacts in the network (independently from each other) according to the four possibilities given below.

- T_{NN} : probability that a non-mask-wearing individual infects a non-mask-wearing individual.
- T_{MN} : probability that a mask-wearing individual infects a non-mask-wearing individual.
- T_{NM} : probability that a non-mask-wearing individual infects a mask wearing individual.
- T_{MM} : probability that a mask-wearing individual infects a mask-wearing individual.

We remark that $T_{NN}, T_{MN}, T_{NM}, T_{MM}$ represent *average* values of transmission probabilities across the entire population based on mask-wearing behavior⁴. Intuitively, we have (e.g., see Eikenberry et al. (2020), Ngonghala et al. (2020), Stutt et al. (2020))

$$(3.1) \quad T_{MM} < T_{MN} < T_{NM} < T_{NN}$$

⁴To obtain a more accurate picture of the spreading dynamics, it might be important to take into account the *heterogeneity* in the population that may lead to *large* variations in transmission probabilities; e.g., it was shown that some people act as super-spreaders (Liu et al., 2020; Lloyd-Smith et al., 2005; Riou & Althaus, 2020).

The assumption $T_{MN} < T_{NM}$ can be explained as follows. The mask is likely to be useful in limiting the droplets emanating from an infectious person, but not as effective in preventing a healthy person catching the virus from a non-mask-wearing, infectious individual (Eikenberry et al., 2020). Finally, we assume that each individual has a probability p of wearing a mask independently from others. As a starting point, it is convenient to assume that mask-wearing behavior is independent from the network structure. Future studies may focus on more complex cases to capture the notion that if most of one's friends are not wearing a mask, then he/she is also likely to not wear a mask (and, vice versa). Henceforth, we refer to the spreading model with parameters p , T_{MM} , T_{MN} , T_{NM} , T_{NN} described above as *the mask model*.

In the SIR/bond percolation model (Allard et al., 2009; Newman, 2002; Yağan et al., 2013), the main parameter of interest is the *transmissibility* of the virus, which represents the *mean* probability that an infectious person transfers the virus to a susceptible person. Here, the mean is taken due to the fact that not every individual has the same contact frequency/behavior with each of their neighbors in the network. In the formulation described above, we can define and calculate two transmissibility parameters, one for mask-wearing individuals (say, T_1) and one for non-mask-wearing individuals (say, T_2). Initially, the probability that a susceptible vertex wears a mask is p , so we may compute the initial transmissibilities T_1 and T_2 by first conditioning on the mask-wearing status of a susceptible neighbor and then taking an expectation:

$$(3.2) \quad \begin{aligned} T_1 &= T_{MM} \cdot p + T_{MN} \cdot (1 - p), \\ T_2 &= T_{NM} \cdot p + T_{NN} \cdot (1 - p). \end{aligned}$$

With the ordering given at (3.1), we obtain $T_1 < T_2$. The parameters T_1, T_2 in (3.2) can be viewed as the transmission probabilities of two different strains of the virus. Those wearing a mask are assumed to be carrying strain-1 that has a smaller transmissibility than strain-2 carried by non-mask-wearers. Next, we calculate the mutation probabilities; e.g., probability that an individual who is infected by a strain-1-carrying individual (i.e., a mask-wearer) transmits to its contacts strain-2 of the virus (i.e., transmits with the probability T_2 associated with non-mask-wearers). For a newly infected individual whose mask-wearing behavior is unknown, assume that we know the mask-wearing behavior of the person who infected them. We can calculate the posterior probability of this newly infected person wearing a mask as

$$\begin{aligned} &\mathbb{P}[\text{newly infected is wearing a mask at time } t \mid \text{they were infected by a mask-wearer}] \\ &= \frac{T_{MM} * p}{T_{MM} * p + T_{MN} * (1 - p)} \\ &=: \mu_{11}. \end{aligned}$$

In the *mutation model* of Eletreby et al. (2020), μ_{11} represents the probability that a person infected by strain-1 of the virus remains in strain-1. In the mask model, the quantity μ_{11} has a potentially useful meaning as well. In particular, each person infected by a mask-wearer will also be wearing a mask with probability μ_{11} (and not wearing a mask with probability $\mu_{12} = 1 - \mu_{11}$). Put differently, for a mask-wearing infectious person, μ_{11} (respectively, μ_{12}) represents the *expected* fraction of mask-wearers (respectively, *non-mask-wearers*) among those that they infect.

Following the same approach, it is easy to compute all four mutation probabilities and thus the mutation probability matrix. We have

$$\begin{aligned}
 \mu_{11} &:= \frac{T_{MM} \cdot p}{T_{MM} \cdot p + T_{MN} \cdot (1-p)}, \\
 \mu_{12} &:= \frac{T_{MN} \cdot (1-p)}{T_{MM} \cdot p + T_{MN} \cdot (1-p)}, \\
 \mu_{21} &:= \frac{T_{NM} \cdot p}{T_{NM} \cdot p + T_{NN} \cdot (1-p)}, \\
 \mu_{22} &:= \frac{T_{NN} \cdot (1-p)}{T_{NM} \cdot p + T_{NN} \cdot (1-p)}.
 \end{aligned}
 \tag{3.3}$$

This completes the analogy between the mask model and the multiple-strain model (Eletreby et al., 2020), paving the way to understanding the transmission behavior of the virus while taking into account mask-wearing behavior. To provide a concrete numerical example, if the parameters of the mask model are

$$p = 0.8; \quad [T_{MM} \ T_{MN} \ T_{NM} \ T_{NN}] = [0.2 \ 0.3 \ 0.4 \ 0.8],
 \tag{3.4}$$

then the corresponding parameters for the multiple-strain model with mutations are given by

$$[T_1 \ T_2] = [0.22 \ 0.48]; \quad \mu = \begin{bmatrix} 8/11 & 3/11 \\ 2/3 & 1/3 \end{bmatrix}$$

Using the analogy between the mask model and the multiple-strain model with mutations, we can leverage the analytical results obtained in Eletreby et al. (2020) and Alexander and Day (2010) to study the mask model. It is important to note that the two models are equivalent (with the analogy introduced above) as long as the parameter p gives the probability that a *susceptible* individual wears a mask throughout the spreading process. In other words, the analogy would lead to exact results at the early stages of the spreading process, i.e., when the population is *naïve* until a *significant* fraction of the population becomes infected. After that point, we would expect the fraction of susceptible individuals who wear a mask to increase as more people are infected (given that non-mask-wearers are more likely to be infected than mask-wearers), making it necessary to render the parameter p , and thus the mutation probabilities in (3.3) and transmissibilities in (3.2), time-varying. Consequently, the probability that an epidemic emerges is the same in both models, meaning that previous results on the mutation model (Alexander & Day, 2010; Eletreby et al., 2020) can be directly used in calculating the probability of emergence in the mask model; see Section 3.3 for a formal discussion of this. Although the results of Eletreby et al. (2020) do not lead to exact predictions for the expected size of the epidemic under the mask model, they can still be used to obtain rough estimates for the epidemic size as we illustrate in Figure 6 in Section 4.

3.3. The probability of emergence. We now provide a formal argument to show that the analytical results from Alexander and Day (2010), Eletreby et al. (2020) on the mutation model yield precise results for the probability of epidemic emergence under the mask model. Recently, Tian et al. (2020) derived the probability that an epidemic emerges in the mask model by a direct analysis of the model. We briefly review their results. Let $\{p_k\}_{k=0}^{\infty}$ be the degree distribution of the contact network, where p_k is the probability that a vertex has degree k . Let g be the probability generating

function (PGF) for the degree distribution, defined by

$$g(z) := \sum_{k=0}^{\infty} p_k z^k.$$

We also define the G to be the PGF for the excess degree distribution, defined by

$$G(z) := \sum_{k=0}^{\infty} \frac{k p_k}{\langle k \rangle} z^k,$$

where $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$ is the mean degree. They showed that the PGF for the number of infected neighbors of patient zero of each type (mask-wearing and non-mask-wearing) is given by

$$(3.5) \quad \begin{aligned} \gamma_M(s, t) &= g(1 - T_1 + T_1(s\mu_{11} + t\mu_{12})), \\ \gamma_N(s, t) &= g(1 - T_2 + T_2(s\mu_{21} + t\mu_{22})) \end{aligned}$$

where γ_M is the PGF if patient zero wears a mask, and γ_N is the PGF if patient zero does not wear a mask. Similarly, the PGF for the number of infected neighbors of a *later-generation infective* of each type is given by

$$(3.6) \quad \begin{aligned} \Gamma_M(s, t) &= G(1 - T_1 + T_1(s\mu_{11} + t\mu_{12})), \\ \Gamma_N(s, t) &= G(1 - T_2 + T_2(s\mu_{21} + t\mu_{22})) \end{aligned}$$

where, as before, Γ_M is the PGF if the later-generation infective wears a mask, and Γ_N is the PGF if the later-generation infective does not wear a mask. Finally, using well-known results from branching process theory, if (Q_1, Q_2) is the smallest non-negative solution of the equation $(s, t) = (\Gamma_M(s, t), \Gamma_N(s, t))$, then we have $(P_1, P_2) = (\gamma_M(Q_1, Q_2), \gamma_N(Q_1, Q_2))$ where P_1 (resp., P_2) is the probability that the epidemic dies out in finite time if patient zero wears a mask (resp., does not wear a mask). The probability of emergence is the probability of the complement of the aforementioned event, given by $1 - P_1$ (resp., $1 - P_2$) if patient zero wears a mask (resp., does not wear a mask). A key insight from Tian et al. (2020) is that the PGFs in (3.5), (3.6) are the same PGFs used in the derivation of the probability of emergence in a two-strain model with mutations with transmissibilities T_1, T_2 and mutation probabilities μ . This shows that the probability of emergence is the same in the mask model and the mutation model.

In Tian et al. (2020), the authors also derived the *epidemic threshold*. Define the matrices

$$\mathbf{T} := \begin{pmatrix} T_{MM} & T_{MN} \\ T_{NM} & T_{NN} \end{pmatrix} \quad \text{and} \quad \mathbf{p} := \begin{pmatrix} p & 0 \\ 0 & 1 - p \end{pmatrix}.$$

Then the basic reproduction number is given by

$$(3.7) \quad R_0 = \left(\frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle} \right) \rho(\mathbf{T}\mathbf{p}),$$

where $\rho(\mathbf{T}\mathbf{p})$ is the spectral radius of $\mathbf{T}\mathbf{p}$. Then, the epidemic threshold is expressed in the usual manner through the basic reproduction number, with a positive probability of epidemic emergence only if $R_0 > 1$; if $R_0 < 1$, the spreading process almost surely dies out without reaching a positive fraction of the population.

It would be of interest to extend the results of Tian et al. (2020) and study a model that incorporates both mutations *and* the impact of mask-wearing. This would involve deriving a new set of formulas for the PGFs of number of infected nodes under different strains and mask-wearing behavior. It might be possible to accomplish this by extending (3.5)-(3.6) in combination with the

corresponding formulas from Eletreby et al. (2020) to a new, and a more complicated, set of formulas; e.g., we would need to have four formulas for the case with two viral strains. Another interesting direction for future work would be to use a *multi-layer* contact network with layers representing different types of relationships between nodes; e.g., co-worker, neighbor, school, etc. This might help in understanding the impact of reducing, or entirely eliminating, the contact rate in certain layers (e.g., by closing schools) in slowing down the spread of the virus.

4. SIMULATION RESULTS

In this section, we present some simulation results pertaining to the mutation model and mask model. The goal of our simulations is to qualitatively illustrate how, in the case of the mutation model, an unlikely but highly virulent strain of a virus can rapidly spread through a contact network, and, in the case of the mask model, how the efficacy of masks and the fraction of mask-users influences the spread of a virus. It is of special interest to understand how masks and viral mutations can affect the spread of SARS-CoV-2, so we choose our simulation parameters to match that of SARS-CoV-2. We emphasize that there is still much that is unknown about the spreading dynamics of COVID-19, particularly as to how asymptomatic individuals spread the virus. Thus our results should be taken as a qualitative assessment of how the COVID-19 pandemic may change and evolve.

We begin by reviewing our simulation model. As mentioned earlier, the mask and mutation models abstract the complex mechanics of viral spreading into the transmissibility parameters, which represent the probability that an individual eventually infects a given neighbor. While such a model can be readily analyzed from a theoretical lens, it does not describe the time-varying behavior of the epidemic (e.g., where the peak in cases occur, how long the epidemic persists before the population recovers, etc.). To properly account for time-varying behavior, we assume that the time it takes for an infected individual to transmit the virus to a neighbor is an $\text{Exp}(r)$ random variable, which is independent of all other randomness in the system. The parameter r is the *contact rate*, which represents the average number of potentially viral-spreading interactions in a given day. The contact rate may depend on the strain that is being transmitted, or could depend on whether the host or target is wearing a mask. If we denote the *infectious period*, which is the amount of time an infected person is contagious, by IP , the transmissibility of the virus is given by

$$T = 1 - e^{-IP \cdot r}.$$

Parameter selection. The reproductive number, R_0 , of SARS-CoV-2 is estimated to be between 1.4 and 3.9 based on data from the earliest 425 confirmed cases in Wuhan, China (Li et al., 2020); we will use the conservative estimate $R_0 = 4.0$. The same authors estimate the serial intervals, which is the average time between infections, to be 7.5 days. For our model, this implies that the contact rate is $r = 1/7.5 \text{ days}^{-1}$. We next estimate the infectious period. There have been multiple studies on the viral shedding period (Bullard et al., 2020; van Kampen et al., 2020; Wölfel et al., 2020), which indicate that the median viral shedding period after the onset of symptoms is at most 8 days. A recent survey of epidemiological studies of COVID-19 by McAloon et al. (2020) showed that the median time to symptom onset is 5 days. Since viral shedding can occur even before the onset of symptoms (He et al., 2020), a conservative estimate of the median infectious period (the amount of time for which an individual may infect others) is 13 days. Since the time between infections is an exponential random variable with a mean of 7.5 days, the transmissibility can be

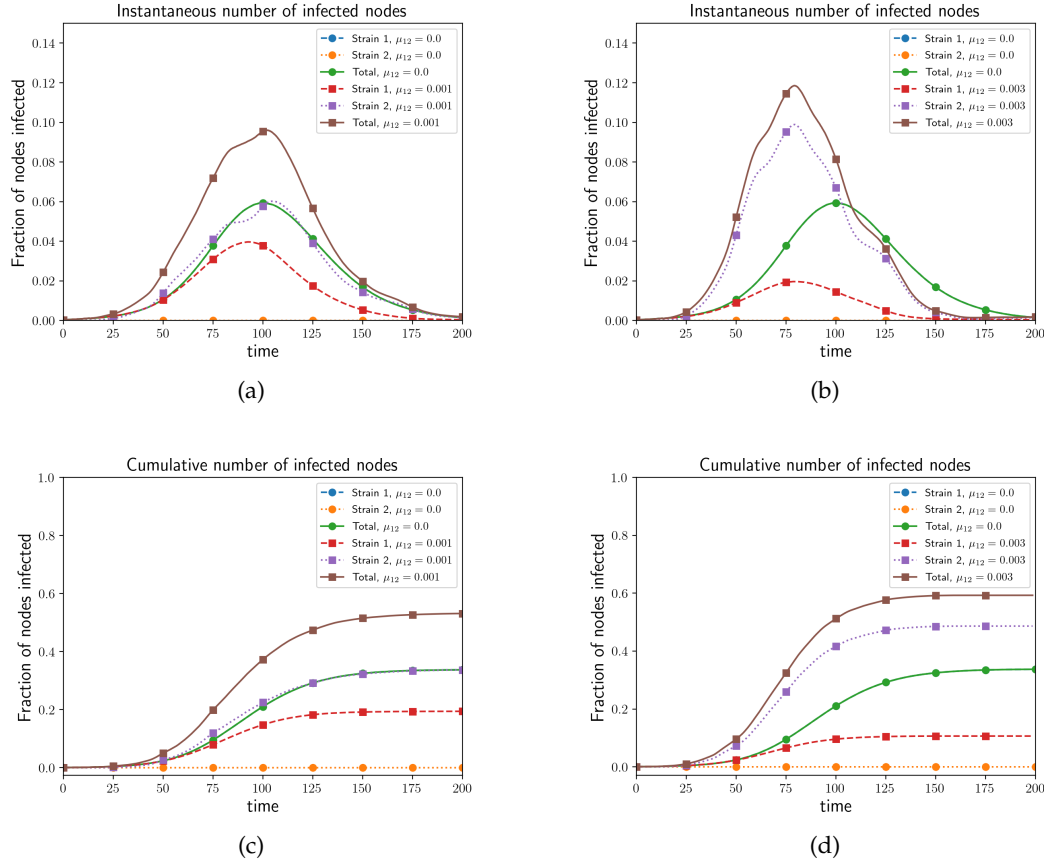


Figure 2. Plots (a) and (c) compare the instantaneous and cumulative number of infections when $\mu_{12} = 0$ and $\mu_{12} = 0.001$, while plots (b) and (d) compare the instantaneous and cumulative number of infections when $\mu_{12} = 0$ and $\mu_{12} = 0.003$. Though the difference in mutation probability is quite small, the infection curves have noticeably different characteristics, such as the height and location of the epidemic peak.

computed as

$$T = 1 - e^{-13/7.5} \approx 0.8.$$

If we assume that the contact network as a $\text{Poisson}(\lambda)$ degree distribution, (3.7) implies that $R_0 = \lambda T$. If we use the conservative estimate $R_0 = 4$, this shows that a reasonable approximation is $\lambda = 5$; i.e., each individual *interacts* with 5 others, on average.

Emergence of highly contagious strains. We consider a scenario where there is initially a single strain of the virus, but with some very small probability it may mutate into a highly contagious strain. In our simulations, we set $T_1 = 0.3$, $T_2 = 0.8$, $\mu_{12} = 0, 0.001, 0.003$ and $\mu_{22} = 1.0$. We simulated the epidemic on random networks with 5000 nodes and a $\text{Poisson}(5)$ degree distribution. We conducted 1000 independent simulations, and averaged the time-varying behavior across

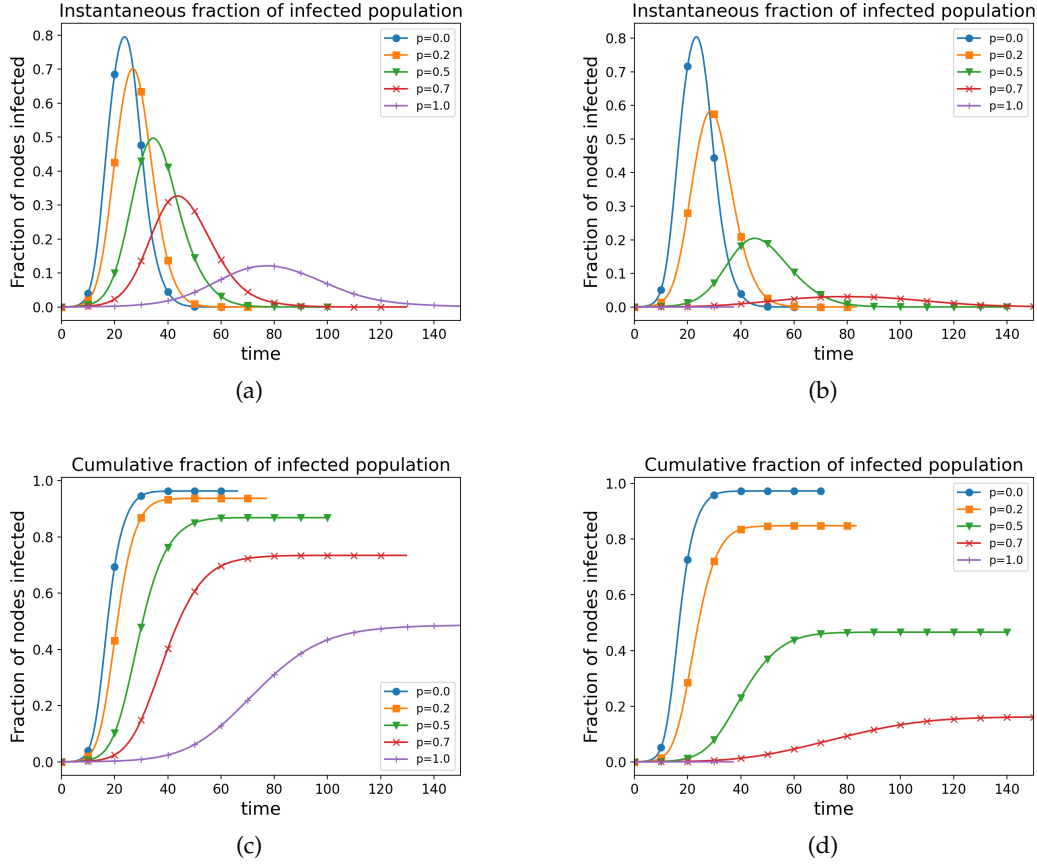


Figure 3. Time-varying infection curves for various values of p , when cloth masks or surgical masks are used. Plots (a) and (c) compare the instantaneous and cumulative infection curves for various p when $\epsilon = 0.5$ (cloth masks), while plots (b) and (d) compare the instantaneous and cumulative infection curves for various p when $\epsilon = 0.2$ (surgical masks). We see that surgical masks significantly impede the spread of the virus compared to cloth masks as the fraction of mask-wearers increases.

all experiments.⁵ Our results are displayed in Figure 2. There are several interesting phenomena that are highlighted by these plots. While Strain 1 dominates the infection curve when $\mu_{12} = 0$, Strain 2 dominates Strain 1 by a wide margin even if μ_{12} is as small as 0.001; this margin is dramatically larger in the case where $\mu_{12} = 0.003$. Hence even if there is a very small chance that a virus could evolve into a highly contagious strain, it is likely in some cases that the more infectious strain will spread the most. Our forthcoming conference publication (Sridhar et al., 2021) includes simulations for additional transmissibilities which illustrate the same qualitative phenomenon.

⁵Specifically, the model is solved in continuous time, and the continuous-time curves are sampled at every 0.01 units of time (days, in this case) to generate the instantaneous and cumulative curves for each simulation. These curves are then averaged over all the independent simulations.

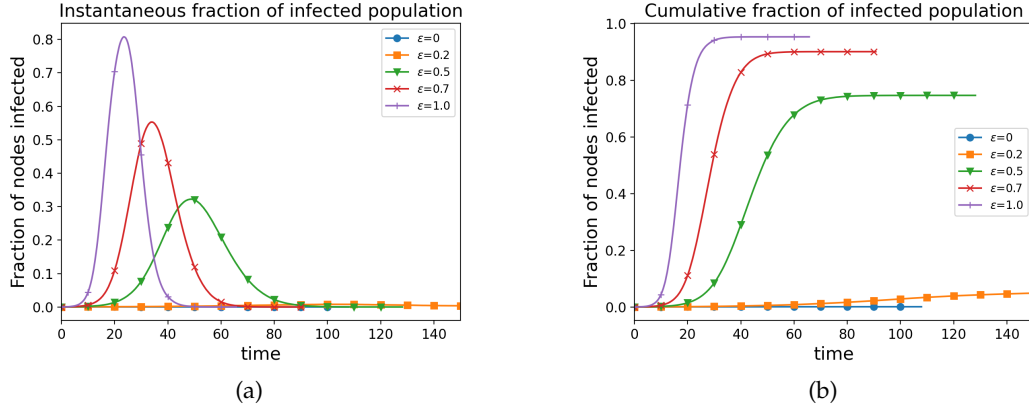


Figure 4. Time-varying infection curves for various ϵ when $p = 0.77$. Plot (a) shows the instantaneous infection curves, and plot (b) shows that cumulative infection curves. The choice of ϵ significantly affects the peak height and location, with the epidemic essentially dying out on its own when $\epsilon = 0, 0.2$.

The impact of mask-wearing. As before, we assume that the contact rate between two non-mask-wearing individuals is $r = 1/7.5 \text{ days}^{-1}$. When one or both of the individuals are wearing a mask, the contact rate decreases since it is harder to transmit the virus. To make this idea more precise, let ϵ_i denote the *inward efficiency* of a mask, defined to be the probability that the mask fails to block the pathogen from coming inside the mask. Similarly, let ϵ_o denote the *outward efficiency* of a mask, defined to be the probability that a mask-wearing individual transmits a pathogen to others. An inward efficiency of $\epsilon_i = 0$ implies that no pathogen can pass from the outside to the inside of a mask, while $\epsilon_i = 1$ means that the mask is useless at blocking pathogens from outside. Similar interpretations hold for $\epsilon_o = 0, 1$. Let r_{MM} denote the contact rate between two mask-wearers, let r_{MN} denote the contact rate between mask-wearing infective and a non-mask-wearing neighbor, with similar interpretations for r_{NM} and r_{NN} . We can then write

$$\begin{aligned} r_{MM} &= \epsilon_i \epsilon_o r, \\ r_{MN} &= \epsilon_o r, \\ r_{NM} &= \epsilon_i r, \\ r_{NN} &= r. \end{aligned}$$

Using the estimate $IP = 13 \text{ days}$, we can compute the transmissibility T_{MM} as

$$T_{MM} = 1 - e^{-\epsilon_i \epsilon_o \frac{13}{7.5}}.$$

We can derive T_{MN}, T_{NM}, T_{NN} through similar computations.

Several papers have assessed the effectiveness of different types of masks in preventing the transmission of respiratory droplets; we refer the reader to the review in Eikenberry et al. (2020, Section 2.3). For simplicity, we shall assume that the inward and outward efficiencies are the same: $\epsilon = \epsilon_i = \epsilon_o$. Based on the discussion by Eikenberry et al. (2020), a reasonable estimate of ϵ is 0.5 if cloth masks are used, and 0.2 if surgical masks are used. Our first set of results can be found in Figure 3, which shows how the infection curves change as the fraction of mask-wearers, p , varies between 0 and 1 when cloth masks or surgical masks are used. In Figure 4, we display

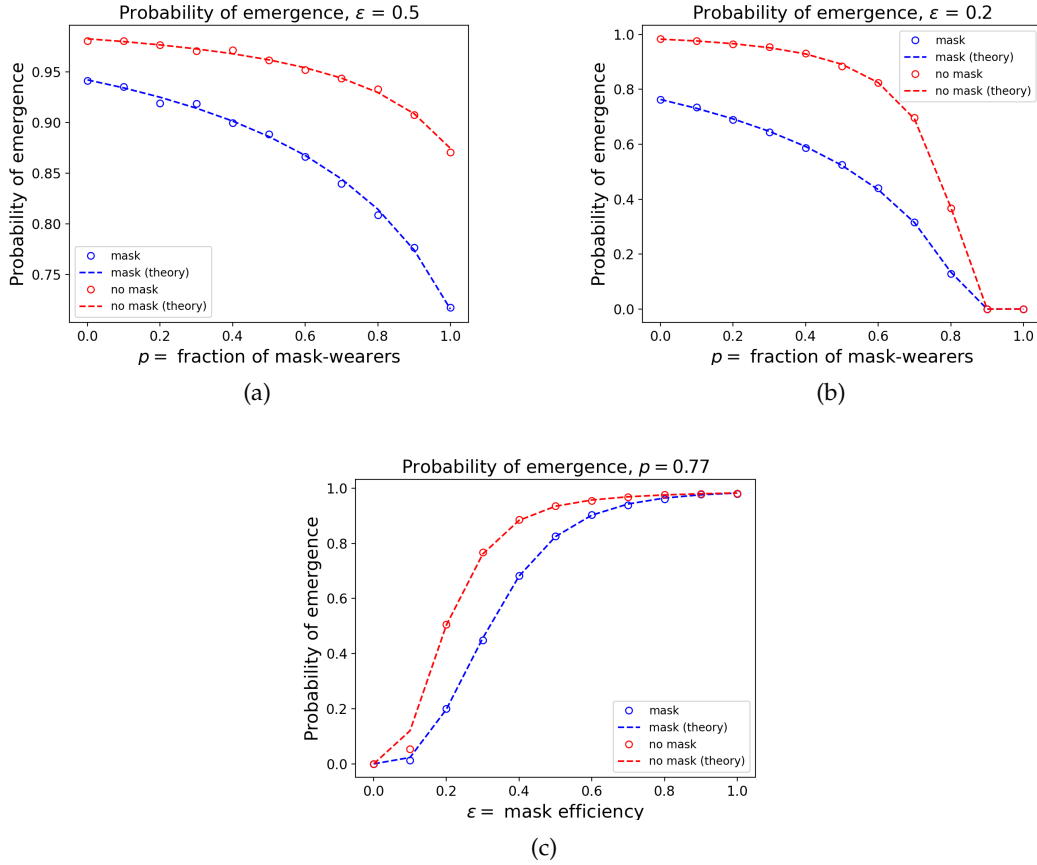


Figure 5. The probability of emergence as a function of p and ϵ . Lines (marked as "theory") correspond to analytical results from Eleteby et al. (2020) and Alexander and Day (2010) corresponding to the mutation model, while the symbols are obtained by empirical simulation of the mask model. Plot (a) studies the probability of emergence as a function of p when $\epsilon = 0.5$ (cloth masks) and plot (b) does the same when $\epsilon = 0.2$ (surgical masks). Plot (c) considers the probability of emergence as a function of ϵ when $p = 0.77$. The blue points/curves correspond to the case where patient zero wears a mask, and the red points/curves correspond to the case where patient zero does not wear a mask. For the most part, the empirical data matches the theoretical curves very well; we expect that the few fluctuations are due to finite-sample effects.

the infection curves when we vary ϵ between 0 and 1 and fix p to be a constant. We set $p = 0.77$, which was the estimated fraction of mask-wearers in New Jersey as of November 20, 2020 ("IHME: COVID-19 Projections", 2020). For all the plots in Figures 3 and 4, the curves shown are generated by averaging over the infection curves of 100 independent simulations run on contact networks with 5000 nodes and a Poisson(5) degree distribution. Additional simulations regarding the time-varying behavior of the mask model can be found in (Sridhar et al., 2021).

Next, we compare the mask model to the analytic results of the corresponding mutation model. In Figure 5, we study the probability of *emergence* as a function of p when cloth masks or surgical masks are used, as well as the probability of emergence as a function of ϵ . To generate the

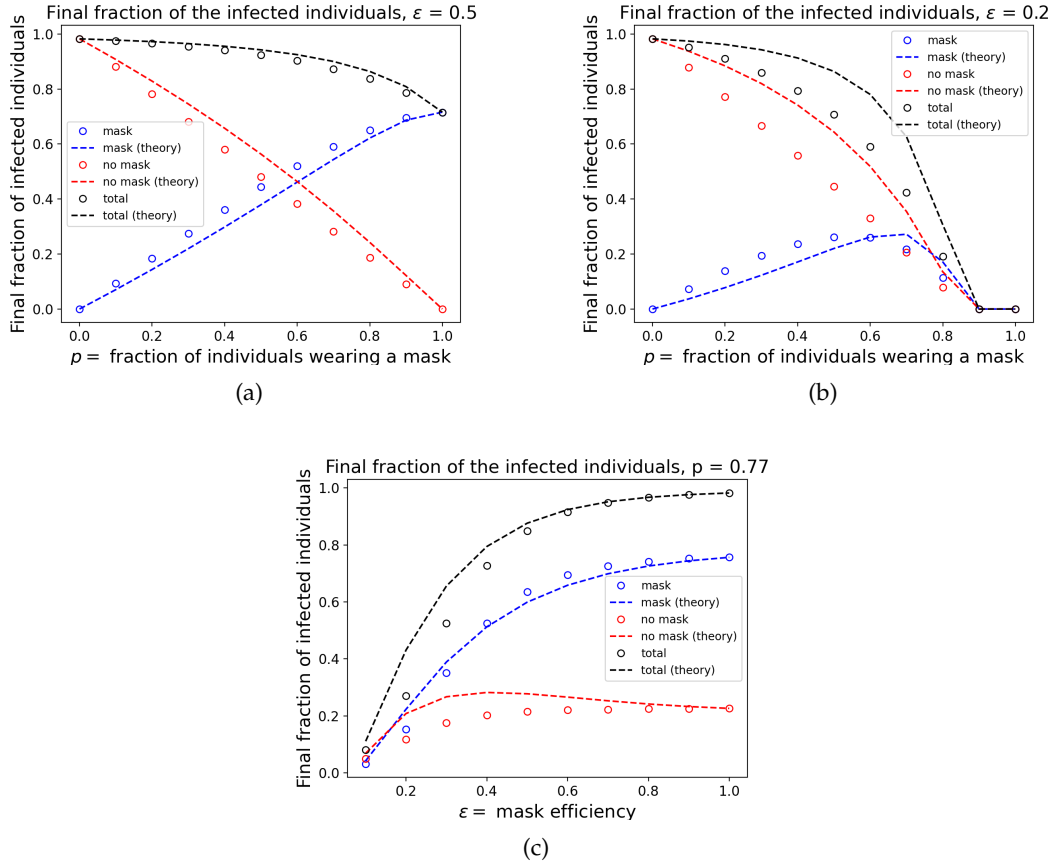


Figure 6. Expected epidemic size as a function of p and ϵ . Lines (marked as "theory") correspond to analytical results from Eleteby et al. (2020) corresponding to the mutation model, while the symbols are obtained by empirical simulation of the mask model. Plot (a) studies the expected epidemic size as a function of p when $\epsilon = 0.5$ (cloth masks), and plot (b) does the same for $\epsilon = 0.2$ (surgical masks). Plot (c) considers the expected epidemic size as a function of ϵ when $p = 0.77$. The blue points/curves are the fraction of individuals who wear a mask and are eventually infected, the red points/curves are the fraction of individuals who do not wear a mask and are eventually infected and the black points/curves are the total fraction of infected individuals (the sum of the red and blue). Although the theoretical results of Eleteby et al. (2020) do not match the expected epidemic size in the mask model, they have the same general trends. The most significant differences between the empirical data and the theoretical predictions are in plot (b).

plots in Figure 5, we conducted 10,000 independent simulations on contact networks with 50,000 nodes and a Poisson(5) degree distribution. To obtain the empirical estimates of the probability of emergence, we counted the fraction of simulations where the epidemic infected at least 5% of the population. We see a good fit between the empirical and theoretical values; while there are a few deviations, we expect that these are due to finite-sample effects. Finally, we study the expected size of the epidemic as a function of p as well as ϵ in Figure 6. To generate the empirical data in this figure, we ran 100 independent simulations on contact networks with 5000 nodes and a Poisson(5)

degree distribution. If the epidemic did not infect more than 5% of the population, we threw out the simulation result and re-ran the simulation until it infected more than 5% of the population. This allowed us to estimate the expected epidemic size conditioned on the ultimate survival of the epidemic.

While the analytical predictions from the mutation model do not match the empirical results from the mask model perfectly concerning the expected epidemic size, they still provide a ballpark estimate and can potentially be useful. As explained before, the mismatch can be attributed to the fact that the fraction p of the mask-wearers among the susceptible population changes over time as the pathogen spreads to a significant fraction of the population, leading to a change in the *mutation probabilities* from the values calculated via (3.3). Further simulations in Sridhar et al. (2021) show that, in some cases, even though the expected size of the epidemic differs in the mask and mutation models, the time-varying behaviors of the two models are very close in the earlier stages of the epidemic propagation.

5. CONCLUSION

The COVID-19 pandemic has claimed hundreds of thousands of lives along with disrupting the lives of billions. A key scientific goal towards helping with the fight against COVID-19 is to develop mathematical models to understand and predict its spreading behavior, as well as to assess the effectiveness of mitigation strategies implemented to limit its spread.

In this paper, we discuss how the multiple-strain epidemic model with mutations studied in Eletreby et al. (2020) can help address key questions concerning the spread of COVID-19. Highlighting the recent reports on a mutation of SARS-CoV-2 that is believed to be more transmissible than the original strain, we discuss the importance of incorporating mutation and evolutionary adaptations in epidemic models. We also demonstrate how the results of Eletreby et al. (2020) can be used to assess the effectiveness of mask-wearing in limiting the spread of COVID-19. We present simulation results on a few sample cases to demonstrate our ideas and the utility of the findings of Eletreby et al. (2020) in the context of COVID-19.

We believe that this paper may stimulate more research efforts on incorporating virus mutation and evolutionary adaptations in epidemic models. We also expect that the analogy established here between the mutation model and the mask model can help facilitate sound assessment of the effectiveness of masks in mitigating the spread of COVID-19.

Disclosure statement. The authors have no conflicts of interest to declare.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation through grants RAPID-2026985, RAPID-2026982, RAPID-2027908, CCF-1813637, CCF-1917819, DMS-1811724; in part by the Army Research Office through grants # W911NF-20-1-0204, # W911NF-17-1-0587, and # W911NF-18-1-0325; and in part by the C3.ai Digital Transformation Institute.

REFERENCES

- Alexander, H., & Day, T. (2010). Risk factors for the evolutionary emergence of pathogens. *Journal of The Royal Society Interface*, 7(51), 1455–1474.
- Allard, A., Noël, P.-A., Dubé, L. J., & Pourbohloul, B. (2009). Heterogeneous bond percolation on multitype networks with an application to epidemic dynamics. *Physical Review E*, 79(3), 036113.
- Anderson, R. M., May, R. M., & Anderson, B. (1992). *Infectious Diseases of Humans: Dynamics and Control* (Vol. 28). Wiley Online Library.
- Antia, R., Regoes, R. R., Koella, J. C., & Bergstrom, C. T. (2003). The role of evolution in the emergence of infectious diseases. *Nature*, 426(6967), 658–661.
- Balmer, O., & Tanner, M. (2011). Prevalence and implications of multiple-strain infections. *The Lancet Infectious Diseases*, 11(11), 868–878.
- Balthrop, J., Forrest, S., Newman, M. E., & Williamson, M. M. (2004). Technological networks and the spread of computer viruses. *Science*, 304(5670), 527–529.
- Barabási, A.-L. (2016). *Network science*. Cambridge University Press.
- Brauer, F., Castillo-Chavez, C., & Castillo-Chavez, C. (2012). *Mathematical Models in Population Biology and Epidemiology*. Springer.
- Bullard, J., Dust, K., Funk, D., Strong, J. E., Alexander, D., Garnett, L., Boodman, C., Bello, A., Hedley, A., Schiffman, Z., Doan, K., Bastien, N., Li, Y., Van Caesele, P. G., & Poliquin, G. (2020). Predicting infectious SARS-CoV-2 from diagnostic samples. *Clinical infectious diseases : An official publication of the Infectious Diseases Society of America*. <https://doi.org/10.1093/cid/ciaa638>
- Daszak, P., Berger, L., Cunningham, A. A., Hyatt, A. D., Green, D. E., & Speare, R. (1999). Emerging infectious diseases and amphibian population declines. *Emerging Infectious Diseases*, 5(6), 735–748.
- Diekmann, O., & Heesterbeek, J. A. P. (2000). *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Sons.
- Dodds, P. S., & Watts, D. J. (2004). Universal behavior in a generalized model of contagion. *Physical Review Letters*, 92(21), 218701.
- Eikenberry, S. E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., Kostelich, E., & Gumel, A. B. (2020). 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*, 5, 293–308.
- Eletreby, R., Zhuang, Y., Carley, K. M., Yağan, O., & Poor, H. V. (2020). The effects of evolutionary adaptations on spreading processes in complex networks. *Proceedings of the National Academy of Sciences of the U.S.A.*, 117(11), 5664–5670.
- Fraser, C., Riley, S., Anderson, R. M., & Ferguson, N. M. (2004). Factors that make an infectious disease outbreak controllable. *Proceedings of the National Academy of Sciences of the U.S.A.*, 101(16), 6146–6151.
- Granell, C., Gómez, S., & Arenas, A. (2014). Competing spreading processes on multiplex networks: Awareness and epidemics. *Physical Review E*, 90(1), 012808.
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656), 327–332.
- Hanski, I., & Hanski, P. (1999). *Metapopulation Ecology*. Oxford University Press.

- He, X., Lau, E., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y., Wong, J. Y., Guan, Y., Tan, X., Mo, X., Chen, Y., Liao, B., Chen, W., Hu, F., Zhang, Q., Zhong, M., Wu, Y., Zhao, L., & Leung, G. (2020). Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26. <https://doi.org/10.1038/s41591-020-0869-5>
- IHME: COVID-19 projections. (2020). <https://covid19.healthdata.org>
- Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181), 990.
- Keeling, M. J., & Eames, K. T. (2005). Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4), 295–307.
- Keeling, M. J., & Rohani, P. (2002). Estimating spatial coupling in epidemiological systems: A mechanistic approach. *Ecology Letters*, 5(1), 20–29.
- Keeling, M. J., & Rohani, P. (2011). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.
- Klempner, M. S., & Shapiro, D. S. (2004). Crossing the species barrier—one small step to man, one giant leap to mankind. *New England Journal of Medicine*, 350(12), 1171–1172.
- Leventhal, G. E., Hill, A. L., Nowak, M. A., & Bonhoeffer, S. (2015). Evolution and emergence of infectious diseases in theoretical and real-world networks. *Nature Communications*, 6, 6101.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S., Lau, E. H., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., ... Feng, Z. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 382(13), 1199–1207.
- Liu, Y., Eggo, R. M., & Kucharski, A. J. (2020). Secondary attack rate and superspreading events for SARS-CoV-2. *The Lancet*, 395(10227), e47.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., & Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066), 355–359.
- Long, S. W., Olsen, R. J., Christensen, P. A., Bernard, D. W., Davis, J. J., Shukla, M., Nguyen, M., Saavedra, M. O., Yerramilli, P., Pruitt, L., Subedi, S., Kuo, H.-C., Hendrickson, H., Eskandari, G., Nyugen, H. A. T., Long, J. H., Kumaraswami, M., Goike, J., Boutz, D., ... Musser, J. M. (2020). Molecular architecture of early dissemination and massive second wave of the sars-cov-2 virus in a major metropolitan area. *Mbio*, 11(6).
- McAloon, C. G., Collins, Á. B., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., Wall, P., Green, M. J., O’Grady, L., & More, S. J. (2020). The incubation period of COVID-19 – A rapid systematic review and meta-analysis of observational research. *medRxiv*. <https://www.medrxiv.org/content/early/2020/04/28/2020.04.24.20073957>
- Miller, J. C., & Kiss, I. Z. (2014). Epidemic spread in networks: Existing methods and current challenges. *Mathematical Modelling of Natural Phenomena*, 9(2), 4–42.
- Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3), 161–180.
- Moreno, Y., Pastor-Satorras, R., & Vespignani, A. (2002). Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4), 521–529.
- Morens, D. M., Folkers, G. K., & Fauci, A. S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996), 242–249.

- Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B., Zambrana-Torrel, C., Lipkin, W. I., & Daszak, P. (2012a). Prediction and prevention of the next pandemic zoonosis. *The Lancet*, 380(9857), 1956–1965.
- Morse, S. S., Mazet, J. A., Woolhouse, M., Parrish, C. R., Carroll, D., Karesh, W. B., Zambrana-Torrel, C., Lipkin, W. I., & Daszak, P. (2012b). Prediction and prevention of the next pandemic zoonosis. *The Lancet*, 380(9857), 1956–1965.
- Newman, M. E. (2002). Spread of epidemic disease on networks. *Physical Review E*, 66(1), 016128.
- Newman, M. E., Forrest, S., & Balthrop, J. (2002). Email networks and the spread of computer viruses. *Physical Review E*, 66(3), 035101.
- Newman, M. E., Strogatz, S. H., & Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2), 026118.
- Ngonghala, C. N., Iboi, E., Eikenberry, S., Scotch, M., MacIntyre, C. R., Bonds, M. H., & Gumel, A. B. (2020). Mathematical assessment of the impact of non-pharmaceutical interventions on curtailing the 2019 novel coronavirus. *Mathematical Biosciences*, 108364.
- Parrish, C. R., Holmes, E. C., Morens, D. M., Park, E.-C., Burke, D. S., Calisher, C. H., Laughlin, C. A., Saif, L. J., & Daszak, P. (2008). Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews*, 72(3), 457–470.
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3), 925–979.
- Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6), 066117.
- Pfennig, K. S. (2001). Evolution of pathogen virulence: The role of variation in host phenotype. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1468), 755–760.
- Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, 15(8), 502–510.
- Qian, D., Yağan, O., Yang, L., & Zhang, J. (2012). Diffusion of real-time information in social-physical networks. *IEEE Conference and Exhibition on Global Telecommunications*, 2072–2077.
- Riou, J., & Althaus, C. L. (2020). Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance*, 25(4), 30.
- Sahneh, F. D., Scoglio, C., & Van Mieghem, P. (2013). Generalized epidemic mean-field model for spreading processes over multilayer complex networks. *IEEE/ACM Transactions on Networking*, 21(5), 1609–1620.
- Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the U.S.A.*, 107(51), 22020–22025.
- Siettos, C. I., & Russo, L. (2013). Mathematical modeling of infectious disease dynamics. *Virulence*, 4(4), 295–306.
- Sridhar, A., Yağan, O., Eletreby, R., Levin, S. A., Plotkin, J. B., & Poor, H. V. (2021). Leveraging a multiple-strain model with mutations in analyzing the spread of COVID-19.
- Stutt, R. O., Retkute, R., Bradley, M., Gilligan, C. A., & Colvin, J. (2020). A modelling framework to assess the likely effectiveness of facemasks in combination with ‘lock-down’ in managing the COVID-19 pandemic. *Proceedings of the Royal Society A*, 476(2238), 20200376.

- Tang, X., Wu, C., Li, X., Song, Y., Yao, X., Wu, X., Duan, Y., Zhang, H., Wang, Y., Qian, Z., Cui, J., & Lu, J. (2020). On the origin and continuing evolution of SARS-CoV-2. *National Science Review*, 7(6), 1012–1023.
- Tian, Y., Sridhar, A., Yağan, O., & Poor, H. V. (2020). Analysis of the impact of mask-wearing in viral spread: Implications for COVID-19.
- van Kampen, J. J., van de Vijver, D. A., Fraaij, P. L., Haagmans, B. L., Lamers, M. M., Okba, N., van den Akker, J. P., Endeman, H., Gommers, D. A., Cornelissen, J. J., Hoek, R. A., van der Eerden, M. M., Hesselink, D. A., Metselaar, H. J., Verbon, A., de Steenwinkel, J. E., Aron, G. I., van Gorp, E. C., van Boheemen, S., ... van der Eijk, A. A. (2020). Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (covid-19): Duration and key determinants. *medRxiv*. <https://doi.org/10.1101/2020.06.08.20125310>
- Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., Kim, B.-a., Comte, B., & Voirin, N. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLOS ONE*, 8(9), 1–9.
- Watts, D. J., Muhamad, R., Medina, D. C., & Dodds, P. S. (2005). Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, 102(32), 11157–11162.
- Wei, X., Ghosh, S. K., Taylor, M. E., Johnson, V. A., Emini, E. A., Deutsch, P., Lifson, J. D., Bonhoeffer, S., Nowak, M. A., Hahn, B. H., Et al. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(6510), 117.
- Wolfe, N. D., Dunavan, C. P., & Diamond, J. (2007). Origins of major human infectious diseases. *Nature*, 447(7142), 279–283.
- Wölfel, R., Corman, V. M., Guggemos, W., Seilmaier, M., Zange, S., Müller, M. A., Niemeyer, D., Kelly, T. C. J., Vollmar, P., Rothe, C., Hoelscher, M., Bleicker, T., Brünink, S., Schneider, J., Ehmann, R., Zwirgmaier, K., Drosten, C., & Wendtner, C. (2020). Virological assessment of hospitalized cases of coronavirus disease 2019. *medRxiv*. <https://doi.org/10.1101/2020.03.05.20030502>
- Woolhouse, M. E., Haydon, D. T., & Antia, R. (2005). Emerging pathogens: The epidemiology and evolution of species jumps. *Trends in Ecology & Evolution*, 20(5), 238–244.
- Yağan, O., & Gligor, V. (2012). Analysis of complex contagions in random multiplex networks. *Physical Review E*, 86, 036103.
- Yağan, O., Qian, D., Zhang, J., & Cochran, D. (2013). Conjoining speeds up information diffusion in overlaying social-physical networks. *IEEE Journal on Selected Areas in Communications*, 31(6), 1038–1048.
- Zhang, L., Jackson, C., Mou, H., Ojha, A., Rangarajan, E., Izard, T., Farzan, M., & H, C. (2020). The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*. <https://www.scripps.edu/news-and-events/press-room/2020/20200611-choe-farzan-sars-cov-2-spike-protein.html>
- Zhuang, Y., & Yağan, O. (2016). Information propagation in clustered multilayer networks. *IEEE Transactions on Network Science and Engineering*, 3(4), 211–224.