LEVERAGING A MULTIPLE-STRAIN MODEL WITH MUTATIONS IN ANALYZING THE SPREAD OF COVID-19

Anirudh Sridhar [‡], Osman Yağan [†], Rashad Eletreby [◇], Simon A. Levin [‡], Joshua B. Plotkin [§], H. Vincent Poor [‡]

†Carnegie Mellon University, ‡Princeton University, ◇Walmart Labs, §University of Pennsylvania

ABSTRACT

The spread of COVID-19 has been among the most devastating 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. In this paper, we discuss how our recent work on a multiple-strain spreading model with mutations 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 can be used to assess the effectiveness of mask-wearing in limiting the spread of COVID-19. Finally, we present simulation results to demonstrate our ideas and the utility of the multiple-strain model in the context of COVID-19.

***Index Terms*—** COVID-19, epidemic modeling, mutations, mask-wearing

1. INTRODUCTION

For over a year, the rapid spread of the novel coronavirus has crippled economies worldwide and as of February 2021, has claimed over 2.3 million lives. Questions of how to safely reopen schools and businesses in the midst of a pandemic are still heavily debated. To provide informed guidelines for reopening, it is necessary to rely on mathematical models that take into account the ways in which the spread of COVID-19 could change in different environments (e.g., hotter vs. colder temperatures) and in response to different interventions (e.g., mask-wearing).

This work was supported in part by the National Science Foundation through grants RAPID-2026985, RAPID-2026982, RAPID-2027908, CCF-1813637, CCF-1917819 and 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.

While there are numerous ways to model an epidemic, we focus on an approach that captures the effects of evolutionary adaptations, or mutations, in viral spread. In the so-called *multi-strain model with mutations*, several strains of a virus spread through a contact network, and one strain may mutate into another when a host is infected [1]. Recent work by Eletreby et al [2] on this model derived analytical predictions for the epidemic threshold and the final fraction of infected individuals. Furthermore, they showed that epidemic models that do not account for mutations may yield incorrect predictions of the spread of an epidemic.

In this paper, we discuss how one may utilize the multi-strain model with mutations as well as the analytical results of Eletreby et al. to model the spread of COVID-19. We first review the critical role of evolutionary adaptations in zoonotic outbreaks and discuss recent reports on mutations undergone by SARS-CoV-2. Through simulations, we study how the emergence of highly virulent strains as a result of mutations can significantly affect the spread of an epidemic. We then demonstrate how the altered spread of a single-strain epidemic due to mask-wearing can also be modeled by the multi-strain model with mutations. In summary, our work emphasizes the importance of mutation in viral spread and offers a modeling framework for predicting the spread of COVID-19. We remark that our work supplements our forthcoming journal publication [3], in which we elaborate on the discussions on epidemic models and evolutionary adaptations, supply further insights on modeling mask-wearing and mutations, as well as additional simulations based on epidemiological properties of COVID-19.

2. MODELING THE SPREAD OF A VIRUS

Modeling the spread of an epidemic. Various epidemic models have been proposed and studied over the past several decades. The earliest and most widely utilized class of models takes a population-level approach, describing the dynamics of the fraction of susceptible and infected individuals by a system of ordinary differential equations (see e.g., [4]). A common criticism is that such models are only mathematically justified if every pair of individuals are equally likely to interact with each other, regardless of geographic

location or other factors. Metapopulation models [5, 6] provide a more realistic approach to population-level models: a pathogen may have different spreading characteristics in subpopulations, and could also spread from one sub-population to another. On the other hand, *network epidemics* offers a fundamentally different modeling perspective, focusing on the complex interactions between individuals on an underlying contact network rather than population-level effects (see [7] and references therein). The goal of the network epidemics is to understand how the spreading characteristics of the virus and the structure of the contact network influence the emergence and final reach of an epidemic.

Evolution of infectious diseases. An unrealistic aspect of many epidemic models is that the pathogen's spreading characteristics are static. Mutations are known to occur when a pathogen is exposed to new environments [8] and can sometimes cause the onset of an epidemic. For instance, while the introduction of zoonotic disease[1] into the human population typically causes limited outbreaks, mutations in new hosts can create highly virulent strains; examples include severe acute respiratory syndrome (SARS), Ebola and influenza [10]. Possible evolutionary adaptions within hosts include gene capture from other organisms (e.g., *Salmonelli enterica* and *E. Coli*) and recombination or reassortment (e.g., H5N1 influenza) [11]. There is evidence that multiple such events can further increase the virulence of a pathogen [10, 12].

The novel coronavirus is an unfortunate example of the power of evolutionary adaptations in creating highly contagious and devastating pathogens. Beyond the initial mutation that led to high inter-human transmission rates, there is evidence that the novel coronavirus has undergone several subsequent mutations [13] and can be classified under two major lineages with functional differences [14]. Further genomic studies have identified mutations in the encoding of the spike proteins of SARS-CoV-2, which facilitated infections in host cells [15]. As the pandemic continues, the novel coronavirus may evolve again, possibly leading to strains that interact with each other and the population in different ways.

A multiple-strain model with mutations. Mutations have only been recently studied in the setting of network epidemic models. Alexander and Day [1] proposed a multi-strain model with mutations and derived the probability of emergence and epidemic threshold for random contact networks with a prescribed degree distribution, generated via the configuration model [16, 17]. Subsequently, Eletreby et al [2] provided an analysis of the expected fraction of individuals infected by each strain, and validated the theoretical predictions on contact networks from an American high school [18] and from a hospital in Lyon, France [19]. We believe that the multi-strain model with mutations can reasonably approximate the spread of RNA viruses [20] such as COVID-19 which have short infectious periods and high mutability. In the following section, we formally describe the multi-strain model with mutations and illustrate through simulations the importance of incorporating evolutionary adaptations into models of viral spread. Further discussion on epidemic models and results obtained by prior authors can be found in [3].

[1] A zoonosis is any disease or infection that is naturally transmissible from vertebrate animals to humans [9].

3. IMPACT OF MUTATIONS ON VIRAL SPREAD

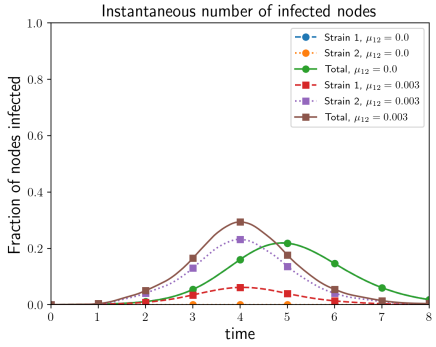
In this section, we illustrate through the multiple-strain model with mutations studied in [2] how the presence of mutations could dramatically affect the trajectory of the COVID-19 epidemic. While prior work on this model [1, 2] considered only long-term characterizations of the multiple-strain model with mutations (such as the final fraction of infections), we simulate a continuous-time model which also captures intermediate-term effects, such as the location and height of the epidemic peak. In [3] we model additional parameter sets based on epidemiological data of the spread of COVID-19.

Model overview. Assume that there are number of strains of a pathogen, and initially, there is a single vertex infected by some strain i. The time it takes for the vertex to infect a susceptible neighbor is assumed to be an independent $\mathrm{Exp}(r_i)$ random variable, where $r_i \in (0, \infty)$ is the *rate of transmission* when an infected individual has strain i. If an individual is infected with strain i, then it mutates into strain j within the host with probability μ_{ij}. Subsequently, the time it takes for the vertex to infect a susceptible neighbor is an independent $\mathrm{Exp}(r_j)$ random variable, where r_j is the rate of transmission for strain j. Once an individual is infected, they recover after one time period[2] and can no longer infect their neighbors. The process terminates when no more infections are possible. An important quantity in the analysis of this process is the *transmissibility* T_i of strain i, which is the probability that a host carrying strain i will infect a susceptible neighbor before recovering. The quantities T_i and r_i are related by the equation $T_i = 1 - e^{-r_i}$ (see e.g., [21]). The underlying contact network is generated by the configuration model [17, 16], which can generate random graphs of any specified degree distribution. To generate the data in Figure 1, we created a random contact network with $\mathrm{Poisson}(5)$ degree distribution on 5000 nodes and simulated the epidemic process to generate instantaneous and cumulative infection curves. We repeated this process 1000 times and averaged over all the infection curves.

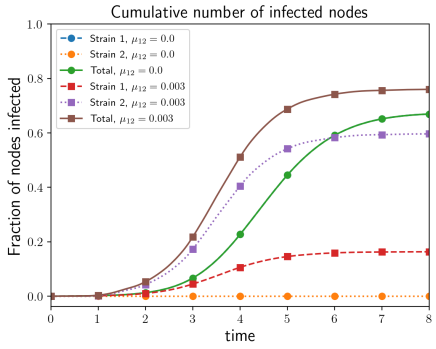
The emergence of highly virulent strains. Suppose that initially a single strain of the virus, but with some very small probability it may mutate into a highly contagious strain. In our simulations, we set $T_1 = 0.4$ and $T_2 = 0.95$.[3] We ran

[2] The timescale of this model is relative to how often individuals have close interactions with each other that could enable the virus to spread from one person to another.

[3] We chose these parameters to illustrate this scenario. In [3], we consider parameters based on epidemiological properties of COVID-19.



(a)



(b)

Fig. 1. Comparison of a baseline example without mutation ($\mu_{12} = 0$) to a contagious but unlikely mutation ($\mu_{12} = 0.003$). Note that the curve marked by blue dots is not visible since it completely coincides with the green dots.

simulations when $(\mu_{11}, \mu_{12}) = (1.0, 0.0)$ and $(0.997, 0.003)$. We additionally set $\mu_{22} = 1.0$ so that the more contagious strain does not mutate. Our results are displayed in Figure 1. Even if Strain 1 mutates into the highly virulent Strain 2 with probability 0.003, the epidemic peak is significantly earlier and higher than the baseline example where $\mu_{12} = 0$; furthermore, the bulk of the infections comes from Strain 2. In other words, if a strain is highly virulent then it tends to be the cause of most of the infections, even if such a mutation may be unlikely. We show in [3] that this phenomenon holds for additional transmissibilities and mutation probabilities μ_{12}.

4. EFFECTIVENESS OF MASK-WEARING

Here we describe how our multiple-strain model with mutations can be used to assess the effectiveness of certain mitigation strategies, focusing on the specific case of mask-wearing. The high-level idea is that if an individual wears a mask, it is harder for them to transmit the virus to others. To phrase things in terms of our multiple-strain model, one can imagine that infected mask-wearers carry a less-contagious strain of the virus compared to infected non-mask-wearers. We formally describe the analogy between mask-wearing and a multi-strain model with mutations below.

Define the transmissibilities T_{MM}, T_{MN}, T_{NM} and T_{NN}, where T_{MM} is the transmissibility between an infected mask-wearer and a susceptible mask-wearer, T_{MN} is the transmissibility between an infected mask-wearer and a susceptible non-mask-wearer, with similar interpretations for T_{NM} and T_{MM}.[4] We assume $T_{MM} < T_{MN} < T_{NM} < T_{NN}$. Clearly, T_{MM} (resp. T_{NN}) is the smallest (resp. largest) of the transmissibilities. We further assume that $T_{MN} < T_{NM}$ since masks help limit respiratory droplets from an infected person, but are less effective in limiting transmission from an infected non-mask-wearer [22]. Let p denote the fraction of mask-wearers in the population and assume that each person wears a mask with probability p, independently across all individuals. Starting with a single infected individual, the virus propagates in an analogous manner to the multi-strain model with mutations described in Section 3. For brevity, we refer to this model as the *mask model*.

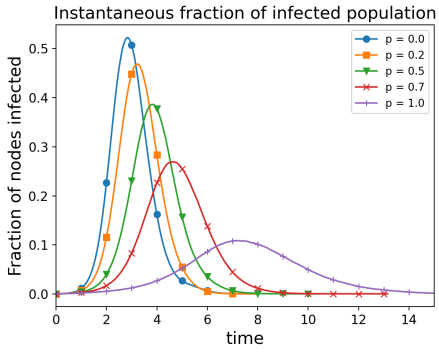
In Figure 2, we simulate the spread of a single strain on random contact networks for various values of p. The contact network was generated by the configuration model with $\text{Poisson}(5)$ degree distribution and 5000 vertices. We set $T_{MM} = 0.3, T_{MN} = 0.4, T_{NM} = 0.5, T_{NN} = 0.6$ and simulated the continuous-time epidemic spread in the same manner as the simulations of Section 3, e.g., we have a rate of transmission r_{MM} related to T_{MM} by $T_{MM} = 1 - e^{-r_{MM}}$. We averaged the epidemic curves over 100 simulations. Figure 2(a) indicates that as more of the population wears a mask, the epidemic peak occurs later and is less intense, thereby "flattening the curve". In Figure 2(b), the epidemic still manages to reach over half of the population even if everyone wears a mask, but at a slower rate. Similar behavior is observed for other network parameters and transmissibilities [3].

We can relate the mask-model to a two-strain model with mutations. Let T_1 (resp. T_2) be the average transmissibility of a mask-wearer (resp. non-mask-wearer). Since there is a p fraction of mask-wearers and a $(1-p)$ fraction of non-mask-wearers in the neighborhood of a vertex, we have

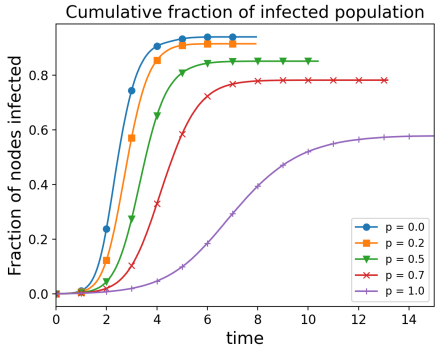
$$\begin{aligned} T_1 &:= T_{MM} \cdot p + T_{MN} \cdot (1-p) \\ T_2 &:= T_{NM} \cdot p + T_{NN} \cdot (1-p). \end{aligned} \tag{1}$$

In this analogy, if an individual is infected with Strain 1 (resp. Strain 2), then they are a mask-wearer (resp. non-mask-wearer). Hence μ_{11} can be interpreted as the probability that a newly-infected individual wears a mask given that they were infected by a mask-wearer. Similar interpretations hold

[4] These transmissibilities represent *average* values, which is evident when we describe the corresponding continuous-time model.

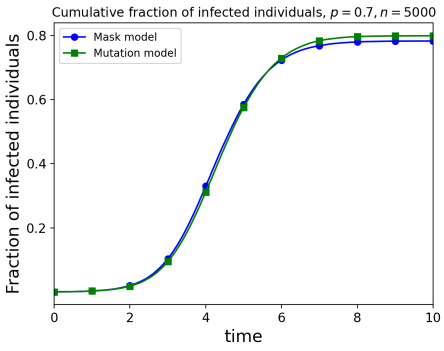


(a)

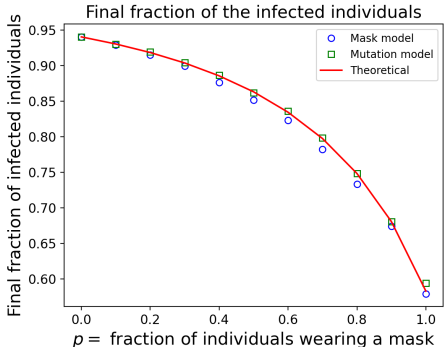


(b)

Fig. 2. Instantaneous and cumulative infection curves as a function of p, the fraction of the population wearing a mask.



(a)



(b)

Fig. 3. Comparison of the evolution of the mask model and mutation model epidemic curves (a) and a comparison of the final fraction of infected nodes in the mask model, mutation model and theoretical predictions for the mutation model (b).

for $\mu_{12}, \mu_{21}, \mu_{22}$. By Bayes' formula,

$$\begin{aligned}
\mu_{11} &= \frac{T_{MM} * p}{T_{MM} * p + T_{MN} * (1-p)} \\
\mu_{12} &= \frac{T_{MN} * (1-p)}{T_{MM} * p + T_{MN} * (1-p)} \\
\mu_{21} &= \frac{T_{NM} * p}{T_{NM} * p + T_{NN} * (1-p)} \\
\mu_{22} &= \frac{T_{NN} * (1-p)}{T_{NM} * p + T_{NN} * (1-p)}.
\end{aligned} \tag{2}$$

In (1) and (2), we assumed that the fraction of susceptible mask-wearers (resp. non-mask-wearers) in a vertex's neighborhood is exactly p (resp. $1-p$). Once the epidemic spreads to a large number of individuals, it is possible that the fraction of susceptible mask-wearers in a neighborhood increases, since the virus has a higher transmissibility among non-mask-wearers. We therefore expect that the analogy is accurate in the early stages of the viral spread and may diverge once a large fraction of the population has been infected. Figure 3 confirms this intuition. Figure 3(a) shows that the infection curves for the mask model and mutation model are nearly identical until a large fraction of the population has been infected, and Figure 3(b) shows that the analytical predictions for the mutation model are slightly inaccurate for the mask model. This inaccuracy does not seem to be caused by finite-network effects or approximation errors [3]. We believe that the analogy between mask-wearing and mutation is a useful tool for understanding the early and intermediate-term spread of an epidemic. Deriving corrections to predict the long-term spread will be a focus for future work.

5. CONCLUSION

In this paper, we have emphasized the importance of incorporating mutation into models of viral spread especially in light of the ongoing mutations of COVID-19. Through simulations, we have illustrated the potential impact of an unlikely mutation into a virulent strain. We have also shown how the mutation model can be used to model mitigation strategies, such as mask-wearing. Future directions include investigating the relation between the mask and mutation models and fitting our model to data from the COVID-19 pandemic.

6. REFERENCES

[1] H. K. Alexander and T. Day, "Risk factors for the evolutionary emergence of pathogens," *J. R. Soc. Interface*, vol. 7, no. 51, pp. 1455–1474, 2010.

[2] R. Eletreby, Y. Zhuang, K. M. Carley, O. Yağan, and H V. Poor, "The effects of evolutionary adaptations on spreading processes in complex networks," *Proc. Natl. Acad. Sci. USA*, March 2020.

[3] O. Yağan, A. Sridhar, R. Eletreby, S. A. Levin, J. B. Plotkin, and H. V. Poor, "Modeling and Analysis of the Spread of COVID-19 under a Multiple-strain Model with Mutations," 2020.

[4] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*, Springer, 2012.

[5] I. Hanski and P.D.E.S.I. Hanski, *Metapopulation Ecology*, Oxford University Press, 1999.

[6] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds, "Multiscale, resurgent epidemics in a hierarchical metapopulation model," *Proc. Natl. Acad. Sci. USA*, vol. 102, no. 32, pp. 11157–11162, 2005.

[7] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.*, vol. 87, no. 3, pp. 925, 2015.

[8] R. Antia, R. R. Regoes, J. C. Koella, and C. T Bergstrom, "The role of evolution in the emergence of infectious diseases," *Nature*, vol. 426, no. 6967, pp. 658, 2003.

[9] World Health Organization, ," http://www.who.int/topics/zoonoses/en/.

[10] C. R. Parrish, E. C. Holmes, D. M. Morens, E. Park, D. S Burke, C. H. Calisher, C. A. Laughlin, L. J. Saif, and P. Daszak, "Cross-species virus transmission and the emergence of new epidemic diseases," *Microbiol. Mol. Biol.*, vol. 72, no. 3, pp. 457–470, 2008.

[11] M. E. J. Woolhouse, D. T. Haydon, and R. Antia, "Emerging pathogens: the epidemiology and evolution of species jumps," *Trends Ecol. Evol.*, vol. 20, no. 5, pp. 238–244, 2005.

[12] S. E. Lindstrom, N. J. Cox, and A. Klimov, "Genetic analysis of human h2n2 and early h3n2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events," *Virology*, vol. 328, no. 1, pp. 101 – 119, 2004.

[13] L. van Dorp, M. Acman, D. Richard, L. P. Shaw, C. E. Ford, L. Ormond, C. J. Owen, J. Pang, C. C.S. Tan, F. A.T. Boshier, A. T. Ortiz, and F. Balloux, "Emergence of genomic diversity and recurrent mutations in sars-cov-2," *Infection, Genetics and Evolution*, vol. 83, pp. 104351, 2020.

[14] X. Tang, C. Wu, Xiang Li, Y. Song, X. Yao, X. Wu, Y. Duan, H. Zhang, Y. Wang, Z. Qian, et al., "On the origin and continuing evolution of SARS-CoV-2," *Natl. Sci. Rev.*, 2020.

[15] L. Zhang, C. Jackson, H. Mou, A. Ojha, E. Rangarajan, T. Izard, M. Farzan1, and Choe1 H., "The d614g mutation in the sars-cov-2 spike protein reduces s1 shedding and increases infectivity," *bioRxiv*, 2020.

[16] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. Algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[17] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Phys. Rev. E*, vol. 64, no. 2, pp. 026118, 2001.

[18] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 107, no. 51, pp. 22020–22025, 2010.

[19] P. Vanhems, A. Barrat, C. Cattuto, J. Pinton, N. Khanafer, C. Régis, B. Kim, B. Comte, and N. Voirin, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PLOS ONE*, vol. 8, no. 9, pp. 1–9, 09 2013.

[20] B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. N. Wood, J. M. Daly, J. A. Mumford, and E. C. Holmes, "Unifying the epidemiological and evolutionary dynamics of pathogens," *Science*, vol. 303, no. 5656, pp. 327–332, 2004.

[21] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, pp. 016128, Jul 2002.

[22] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. B. Gumel, "To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the covid-19 pandemic," *Infectious Disease Modklempner2004crossingelling*, 2020.