

RESEARCH ARTICLE

Models for Digitally Contact-Traced Epidemics

CHIARA BOLDRINI¹, ANDREA PASSARELLA¹, AND MARCO CONTI

CNR-IIT, 56124 Pisa, Italy

Corresponding author: Chiara Boldrini (chiara.boldrini@iit.cnr.it)

This work was supported in part by the SoBigData++, HumaneAI-Net, and SAI Projects funded by the European Union's Horizon 2020 Research and Innovation Program under Grant 871042 and Grant 952026; in part by the SAI Project through CHIST-ERA under Grant CHIST-ERA-19-XAI-010; in part by the Ministero dell'Università e della Ricerca (MUR); in part by the Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under Grant I 5205; in part by the Engineering and Physical Sciences Research Council (EPSRC) under Grant EP/V055712/1; in part by the Narodowe Centrum Nauki (NCN) under Grant 2020/02/Y/ST6/00064; in part by the Eesti Teadusagentuur (ETAg) under Grant SLTAT21096; and in part by the Bulgarian National Science Fund (BNSF) under Grant KP-06-DOO2/5.

ABSTRACT Contacts between people are the main drivers of contagious respiratory infections. For this reason, limiting and tracking contacts is a key strategy for controlling the COVID-19 epidemic. Digital contact tracing has been proposed as an automated solution to scale up traditional contact tracing. However, the required penetration of contact tracing apps within a population to achieve a desired target in controlling the epidemic is currently under discussion within the research community. In order to understand the effects of digital contact tracing, several mathematical models have been studied. In this article, we propose a novel compartmental SEIR model with which it is possible, differently from the models in the related literature, to derive closed-form conditions regarding the control of the epidemic. These conditions are a function of the penetration of contact tracing applications and testing efficiency. Closed-form conditions are crucial for the understandability of models, and thus for decision makers (including digital contact tracing designers) to correctly assess the dependencies within the epidemic. Feeding COVID-19 data to our model, we find that digital contact tracing alone can rarely tame the epidemic: for unrestrained COVID-19, this would require a testing turnaround of around 1 day and app uptake above 80% of the population, which are very difficult to achieve in practice. However, digital contact tracing can still be effective if complemented with other mitigation strategies, such as social distancing and mask-wearing.

INDEX TERMS COVID-19, analytical models, digital contact tracing, testing efficiency, SEIR model, closed-form solutions, app uptake, epidemic controllability.

I. INTRODUCTION

Since April 2020, the WHO has been recommending two main and complementary strategies to curb the COVID-19 epidemic: social distancing on the one end, *test, trace, treat* (the famous 3 T's) on the other. As for any respiratory viral infection, the sooner we are able to "remove" contagious people from interacting with others, the sooner the epidemic will be restrained. Indeed, if, on average, each infected person infects less than one susceptible person, rather than, for example, two or three, the epidemic will die naturally [1]. Of course, for this to be effective, all three T's must be carried out swiftly. Contact tracing without testing is impossible: first, you have to know that a person is potentially contagious before being able to track down their contacts.

The associate editor coordinating the review of this manuscript and approving it for publication was Chaitanya U. Kshirsagar.

Similarly, tracing must be completed as quickly and as thoroughly as possible: the longer it takes to identify past contacts, the longer the time a potentially contagious person spends unknowingly infecting other people. Then, contagious people must be immediately isolated, and treated if necessary.

Contact tracing can be performed manually or digitally. Manual contact tracing involves reconstructing the history of past contacts with the infected person in the days before being detected as contagious. This is typically done through interviews with the infected person. Manual contact tracing suffers from two main problems: i) it is labor-intensive, hence it struggles to keep up when the number of daily new cases is high, and ii) the contagious person might not be able to recall precisely their past contacts (simply because they forget some or because some chance contacts with strangers are not noticed in the first place). Digital contact tracing, performed

by means of smartphone apps that – typically via Bluetooth – automatically detect and register contacts, have the potential to overcome the two limitations of manual contact tracing described above. The research community has already identified convincing solutions that provide reasonable trade-offs between privacy and tracing accuracy [2], [3], [4], [5]. Specifically, decentralized Bluetooth-based contact tracing has emerged as the solution of choice and privacy-preserving apps based on this approach have been deployed¹ in many countries [6], [7]. Recent research proposals involve leveraging blockchains and IoT for a more efficient and privacy-preserving tracking [8]. The vast majority of deployed apps leverage the Exposure Notification protocol, jointly rolled out by Apple and Google in Spring 2020. However, digital contact tracing comes with its own problems. The main one is that, for it to be effective, a significant percentage of the population must have the app installed [9]. Bumping up this percentage may not be as easy as it seems [10]. For example, people with old smartphones (typically not supporting Bluetooth Low Energy or for which an updated operating system is not available) cannot enjoy the tracking functionality. Fear of government intrusion into privacy turns off other potential participants.

Due to the limitations discussed above, the percentage of people with an installed and fully-functioning contact tracing app will be far from 100%. Thus, key questions are, among others: how large should this percentage be for digital contact tracing to be effective in containing the epidemic? How does this percentage depend on the contact patterns between people? How does it depend on the implementation of other mitigation measures (such as social distancing and mask-wearing)? Digital contact tracing has yet to be properly evaluated as a public health measure through a large-scale assessment [11]. Thus, the answers to the above questions must then necessarily come from mathematical models and simulations. The network of people with the contact tracing app installed is just another instance of a mobile social network [12]: people interact with each other socially and these interactions are mirrored in the anonymous data collected from the contact tracking app. Thus, by taking advantage of the properties of this mobile social network, we will investigate the above questions.

Ferretti *et al.* [9] adapted a model introduced by Fraser *et al.* [1] (and based on the popular Von Foerster equation) in order to tackle the same research problem. However, the model in [9] can only be solved numerically, and hence it is unable to yield a closed-form condition under which the epidemic can be controlled based on the characteristics of the digital contact tracing in place. In this article, to complement the model in [9], our goal is to propose a deterministic compartmental model for digital contact tracing that provides closed-form conditions for the control of an epidemic. *Closed-form control conditions are crucial for the understandability of models and are instrumental for decision makers and computer scientists working on digital contact tracing.*

¹For an extensive list: https://en.wikipedia.org/wiki/COVID-19_apps.

The contributions of this paper are the following:

- To complement the model in [9], we propose a deterministic compartmental model for digital contact tracing. The advantage of this modeling approach is that closed-form solutions can be obtained, and hence analytical conditions on the control of the epidemic can be derived. Vice versa, the Von Foerster equation on which [9] is based, is more accurate than simple compartmental models, but can only be solved numerically.
- Alongside the compartmental model, we also introduce a standalone model that captures how the testing delay affects the efficacy of the detection of infected people, depending on the duration of the latent window (the period during which an infected person is not yet contagious) and the contagious window (during which the infected person is contagious but has yet to develop symptoms). This model is general, and can be solved in closed form for some common distributions describing these time intervals.
- We apply the model to realistic COVID-19 epidemic scenarios, showing that, even with high penetration of digital contact tracing, the control of the epidemic is extremely difficult without additional mitigation strategies.

The rest of the paper is organized as follows. In Sections II-III, we overview the main results in the related literature regarding COVID-19 modeling and we summarise the properties of the disease itself that are important from the modeling standpoint. Our deterministic compartmental model is presented in Section IV, together with the model on the efficacy of detection of infected persons. The proposed model is then applied to a set of realistic epidemiological scenarios in Section V. Finally, Section VI concludes the paper.

II. A BRIEF OVERVIEW ON COVID-19

From the modeling standpoint, a crucial aspect is to understand when infected people become contagious. For any viral disease, the typical timeline is the following. Following contact with a contagious person, an individual may become infected. However, they do not become contagious immediately: there is a *latent period* during which the person is infected but not yet contagious (i.e., the virus is replicating but its quantity is not yet enough to infect another person). Another important stage is the *incubation period*, which goes from the infection time to the time when the person starts developing symptoms. The latent period may be shorter than the incubation period: this means that an infected person becomes contagious before developing symptoms. Clearly, this makes controlling the spread of the disease harder, since the contagious person who has not developed any symptoms is not aware of their contagiousness. Despite being the subject of hot debates in the initial phases of the COVID-19 epidemic, it is now clear that asymptomatic and pre-symptomatic carriers play a major role in the spread of the SARS-CoV-2 virus [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27].

SARS-CoV-2 is an airborne² virus, i.e., it travels through the air. The typical transmission pathway occurs when a contagious person talks, sneezes, or coughs, producing infectious droplets that are inhaled by people in close proximity. Less frequently, these droplets may fall on the surfaces in close proximity and then contribute to transmitting the disease when the contaminated surface is touched by a susceptible person and then this person touches his/her face (eyes, mouth, etc.). The latter transmission pathway is known as environmental transmission. It is not known to play a major role in the COVID-19 epidemic,³ hence we will not consider it in the modeling. A third transmission pathway is that of aerosol [28], [29], [30], [31]: when a contagious person talks, sneezes, or coughs they also produce some smaller droplets (known as aerosol) that evaporate faster than they fall on the ground [32]. This means that with aerosol transmission, the dry virus lingers in the air for a considerable time and travels long distances. The bad news is that common face masks (such as surgical and cloth ones) are not well equipped to contain such small droplets. Thus, aerosol transmission is much more challenging than droplet transmission, which can be easily contained by relying on widespread social distancing and lower-grade mask-wearing. However, model-wise, they can both be captured by appropriately tuning the probability of infection upon contact.

III. MODELS OF EPIDEMICS

There are two main modeling approaches in the related literature: mathematical models and agent-based models. Mathematical models of epidemics typically lay out a system of Ordinary Differential Equations (ODE)/Partial Differential Equations (PDE) that describes how the number of susceptibles, infected, etc., varies over time. Sometimes these systems can be solved in closed form and provide very useful trends describing what-if scenarios. Otherwise, numerical solutions can be obtained. Due to their nature, these models are based on several simplifying initial assumptions to make the mathematical representation of the phenomenon tractable. On the opposite side of the modeling spectrum, there are agent-based models. Agent-based models are computational models where agents (corresponding to people) interact, in simulation, according to some predefined rules, which can be arbitrarily complex [33], [34], [35], [36], [37], [38], [39]. They are conceptually very similar to the models used in transportation simulation. They recreate synthetic populations in terms of demographics, traffic flows, etc., and then an epidemic is simulated. Recently, machine learning approaches have gained popularity. Tomy et al. [40] exploit Graph Neural Networks (GNN) to solve the problem of inferring the state of the entire population by observing just a few individuals. GNN have also been used to forecast pandemic evolution [41], [42], [43], for the temporal reconstruction of epidemic spreading [44], and for identifying patient-zero in an epidemic [45]. More traditional approaches have also been

used, e.g., in [46], for epidemic forecasting. Since neither agent-based nor machine learning approaches are the focus of this work, we will not discuss them further.

By far, the most used mathematical model is the classical SIR model and its many variations [47]. In the basic version, people are divided into three compartments (denoted with S, I, R, hence the name of the model). In S, people are susceptible to the disease, i.e., they can become infected upon contact with an infectious person. Infectious people are in compartment I. After a certain time spent in compartment I, infected people recover and move to compartment R. Transitions between compartments are then modeled as follows:

$$\begin{aligned}\frac{dS}{dt} &= -\beta I \\ \frac{dI}{dt} &= \beta I - \gamma I \\ \frac{dR}{dt} &= \gamma I.\end{aligned}$$

Parameters β and γ describe the rate at which susceptibles become infected and infected recover. The SIR model can be described by a system of ordinary differential equations that can be solved in a closed form. This representation of an epidemic is referred to as *deterministic*, because the above equations are an approximation, holding for very large populations, of the stochastic version⁴ of the SIR model [47]. The simple SIR model has been extended in several directions, adding the *exposed* compartment (where people infected but not yet contagious reside), which we also use in this work, and many more (see [47] for a general discussion and [48] for an application to COVID-19). It has also been used to study the two-pathogen case [49], when two pathogens insist on the same population (we do not consider this case in our work).

The deterministic compartmental models discussed above are based on the simple assumption that the time spent in each compartment can be reasonably approximated with an exponential random variable (the Markovian assumption). When this is not the case, other types of models must be considered. An important class of non-Markovian models are those based on the McKendrick-VonFoerster equation, which incorporates a so-called age structure to the model [47]. Originally, this model was designed to capture births and deaths in the dynamics of population growth in cellular biology: offspring are generated at a young age, and death occurs typically at an old age. Hence, keeping track of the population

⁴Stochastic SIR models have been extensively studied in the complex system community, as they are amenable to capture the effect of network topology. At the beginning of a disease outbreak, individual variability (such as whether a node is a “hub” in the contact network) plays a major role in determining whether an epidemic will occur or not, and deterministic models (which treat all nodes as equal) are not an appropriate choice. In fact, deterministic models assume a fully-mixed population, meaning that an infected individual has the same probability of infecting any susceptible node in the network. This assumption is needed to write down the ODE system, but it is only considered reasonable once an epidemic has started, and several nodes are already contributing to it. Our model, similarly to other ODE-base models, should be used when a disease has already achieved its epidemic phase, in which accurately capturing the outbreak stage is no longer important while having tools for studying the controllability becomes essential. Please refer to [47], and reference therein, for a detailed discussion.

²We use the layman’s definition of airborne here. Technical use implies only transmission through aerosol.

³<https://www.ecdc.europa.eu/en/covid-19/latest-evidence/transmission>

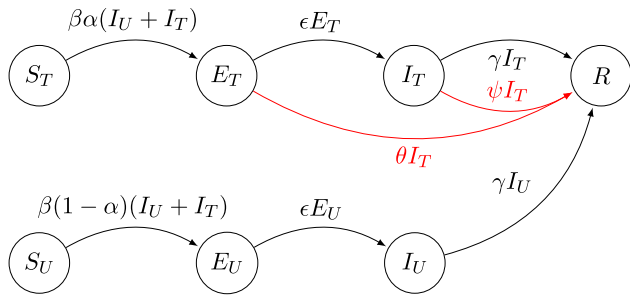


FIGURE 1. How people move across SEIR compartments. In red are the arcs associated with containment measures: quarantine of exposed people plus isolation of those infected and contagious. α denotes the fraction of the digitally tracked population.

age over time was essential to predict the evolution of the population size. When applied to epidemiology, *age* is seen from the point of view of infection, i.e., it corresponds to the time since the individual became infected. And the *birth rate* at infection age τ becomes the rate at which a person infected τ days ago produces offspring, i.e., newly infected people. Thus, the infection rate is no longer constant over time and depends on the current age profile of the population.

Finally, a related active area of study for COVID-19 is the correct estimation of the parameters that describe the dynamics of infection [50], [51], [52], [53], [54], [55]. This is important both for purely mathematical models and for agent-based models, because a correct estimation of the epidemic parameters allows researchers to correctly set up their assumptions and simulations.

IV. FACTORING IN DIGITAL CONTACT TRACING

The McKendrick-VonFoerster model introduced in Sec. III has been used in [9], a seminal paper dedicated to evaluating the efficacy of contact tracing for COVID-19. The model by Ferretti et al. [9] is based on the one proposed in [1], with parameter values customized to the COVID-19 setting. This model is the de-facto reference for digital contact-tracing effectiveness estimation and the vast majority of forecasts, coming both from within the scientific communities and news outlets, have been based on its results. By its own nature, the McKendrick-VonFoerster model can only be solved numerically. Hence, it is unable to yield a closed-form condition under which the epidemic can be controlled based on the characteristics of the digital contact tracing in place. Thus, in this work, we complement the results of Ferretti et al. [9] showing how a simpler model, whose control condition is solvable in closed form, can be obtained. The notation we use in the paper is summarized in Table 1 at the end of the section.

We start with a simplified version of the model, presented in Figure 1, for illustrative purposes. As usual for deterministic compartmental models, we start with a population of constant size N , i.e., the sum of people in all states must add up to N . When looking at a large population, the short-term variation in its size is small and can be neglected. The goal of the model is to capture how infection spreads through the population and to assess whether the spread can be stopped

or not by means of digital contact tracing and the resulting quarantine of contacts. We assume that the epidemic is faster than the long-term dynamics of births and deaths in the population, so we ignore the latter. People can be in one of four states: *S* (susceptible), *E* (exposed), *I* (infectious), *R* (removed). Since people can be either tracked (with a contact tracing app) or untracked, we duplicate these states to account for this difference (thus, each state is marked with subscript T or U for tracked and untracked, respectively). We do not need to duplicate the *removed* state because people in *R* do not contribute to the epidemic anymore. While we use the common letters *S*, *E*, *I*, *R* to denote the states, we slightly adjust the default meaning of the states to take into account asymptomatic and presymptomatic transmission. In this model, then, *exposed* means infected but not yet contagious, while *infectious* means infected and contagious but with no symptoms (this includes the pre-symptomatic and the asymptomatic phase of the disease). The *removed* state includes all infected (whether contagious or not) that have been isolated and/or have recovered. In this simplified model, a person is isolated either because she is infected and has been tracked by the contact tracing app or because she is contagious and has started developing symptoms. We do not include a dedicated state where people are both contagious and symptomatic because it is reasonable to assume that people with symptoms will isolate themselves (and therefore join the *removed* state).

The fraction of tracked people is indicated by α , where α represents the percentage of the population that subscribed to the considered contact tracing app. Thus, at time $t = 0$, we have αN people in state S_T (corresponding to people that are susceptible and tracked) and $(1 - \alpha)N$ people in S_U (susceptible but not tracked). From the *susceptible* state, people can only move to the *exposed* state.⁵ Recall that “exposed”, in this case, means infected but not yet contagious. Therefore, the time spent in the *exposed* state corresponds, without containment measures in place, to the latent period of the disease. However, there is a crucial difference between untracked and tracked exposed: tracked exposed will be notified by health authorities about their previous contact with a tracked contagious person and they will be isolated, i.e., they will move to the *removed* state. The same happens for tracked infectious. Removed people are isolated, and hence cannot infect anyone. This is the crucial contribution of contact tracing. Below, we summarize how the transitions from each state can be modeled.

$S_U \rightarrow E_U$ The rate at which people leave the S_U state is given by the effective contact rate β (rate at which there is an encounter between an *S* and an *I* and this encounter generates an infection) times the number of possible encounters between people in S_U and those in *I*

⁵Technically, also some susceptibles can be removed (i.e., asked to isolate) due to, e.g., a faulty detection from the contact tracing app or because the app detected a contact that did not generate an infection. In practice, these notified people will take a test in a matter of days and the test will come up negative, joining again the *S* compartment. Thus, since SEIR models are based on the assumption $S \sim N$, with N large, this removal can be effectively ignored.

(we don't care whether the encounter is with a tracked or untracked infected).

$S_T \rightarrow E_T$ The same holds true for the rate at which tracked susceptibles leave S_T , with the appropriate change from U to T of S's subscript with respect to the previous case.

$E_U \rightarrow I_U$ Untracked exposed (people in E_U state) stay there until they become contagious, and this happens at a rate ϵ .

$E_T \rightarrow I_T$ Instead, *tracked* exposed will either become contagious and move to state I_T (this happens with the rate ϵ) or be warned of having had contact with an infected and told to self-isolate (this happens with rate θ , discussed in detail in Sec. IV-A).

$I_U \rightarrow R$ Untracked contagious (I_U) are isolated when they begin to develop symptoms, and this happens at a rate γ .

$I_T \rightarrow R$ Instead, tracked contagious (I_T) can be either isolated when they start developing symptoms (similarly to the untracked case) or be informed that they have had contact with an infected and told to self-isolate (this happens with rate ψ , discussed in detail in Sec. IV-A).

The key point in being able to add the effect of contact tracing to the SEIR equations is to adequately model the red transitions in Figure 1. The rates θ and ψ capture how effective digital contact tracing is in removing infected people, and factor in testing delay as well as epidemic features (latent period, contagious period, etc.). We will discuss this aspect in Section IV-A and provide a methodology to derive θ and ψ . For now, let us assume that we can assign proper values to all the parameters of the model. In Theorem 1 below, we discuss how to solve the model and how to assess whether the epidemic can be controlled or not depending on the efficacy of contact tracing. The proof of the theorem can be found in Appendix.

Theorem 1: The epidemic described by the SEIR model in Figure 1 can be controlled when the following condition is true:

$$C1: \alpha - \left(1 + \frac{\gamma}{\theta + \psi}\right) \left(1 - \frac{\gamma}{\beta}\right) > 0. \quad (1)$$

Remark: The closed-form condition in Theorem 1 could not have been obtained with the model used by Ferretti et al. in [9], which can only be solved numerically. Closed-form conditions are crucial for the understandability of models, and thus for decision makers (including digital contact tracing designers) to correctly assess the dependencies within the epidemic. Note also that the complexity of this solution is $O(1)$, hence it does not depend on the size of the population like, e.g., for agent-based models. Thus, C1 provides an easily interpretable and fast answer to the controllability problem, albeit approximate.

To illustrate the intuition behind condition C1 in Theorem 1, let us consider two ideal cases separately: i) instantaneous tracing ($\theta + \psi \rightarrow \infty$) and ii) perfect app uptake ($\alpha = 1$). These correspond to the two dimensions of digital contact tracing: how good we are at detecting infections of the tracked people and how many people we are able to track. When tracing is instantaneous (corresponding to the

first case above), the threshold on α (derived from Equation 1) converges to $1 - \frac{\gamma}{\beta}$. For SIR models, the ratio $\frac{\beta}{\gamma}$ corresponds to the basic reproduction number R_0 [56], hence the threshold on α , interestingly, is equivalent to the herd immunity threshold $1 - \frac{1}{R_0}$. Note that instantaneous tracing alone is not sufficient to control the epidemic: α must be high enough for tracking to cover a large fraction of the population. A superfast tracking that only follows just a tiny fraction of the population is basically useless. In the second case (perfect app uptake, i.e. $\alpha = 1$), condition C1 reduces to $\gamma + \theta + \psi > \beta$. Therefore, $\theta + \psi$ must be large enough to compensate for a high β (effective contact rate). This means that even under the ideal situation where everyone has the app ($\alpha = 1$), control of the epidemic may not be attainable if the tracing process is slow. The efficiency of contact tracing is captured by θ and ψ and, in Section IV-A, we discuss how to derive them.

A. ESTIMATING PARAMETER θ AND ψ FROM CONTACT HISTORY

Now, we step back and discuss how to model θ and ψ , which are the rates at which exposed and infectious people are removed, respectively. As discussed above, they capture the effectiveness of testing. To derive them, we have to reconstruct the process from contagion (encounter with an infectious person that yields to infection) until removal.

Exposure notifications are triggered by tracked people becoming symptomatic and, therefore, being tested. We know that, since SEIR models assume homogeneity in encounters (which boils down to a single β describing the entire contact process, with no distinction between high vs. low social interactions), the contact rate at which the newly symptomatic tracked person met with tracked susceptibles is $\alpha\beta$ (i.e., the baseline rate scaled by the fraction of tracked people). This rate must be split across the different states in which the previous contact might currently be in. Specifically, a past contact can be still exposed, already infectious, or removed. We neglect the removal of susceptibles because the population of susceptibles is very large (by assumption, $S \sim N$), hence removing them would not impact the epidemic. Thus, Definition 1 below follows.

Definition 1 (Alertable Contacts): The alertable contacts of a positive person i can be a) in state E_T (no symptoms, not contagious), b) in state I_T (contagious, no symptoms), c) in state R (symptomatic or recovered, hence already "removed" from the epidemic). We denote the probabilities associated with each of these conditions as p_E , p_I , and p_R , respectively (note that they add up to 1).

Intuitively, health authorities should strive to increase as much as possible p_E , because people in the *exposed* state have yet to infect someone. Of course, this might not be possible (e.g., due to testing delays) so the next best thing is to increase p_I . Instead, notifying people who are already in the *removed* state is completely useless from the point of view of epidemic containment.

As illustrated in Figure 2, we can model the conversion to symptomatic of a previous contact considering: the length of the latent period L (which, as discussed in Section II, goes

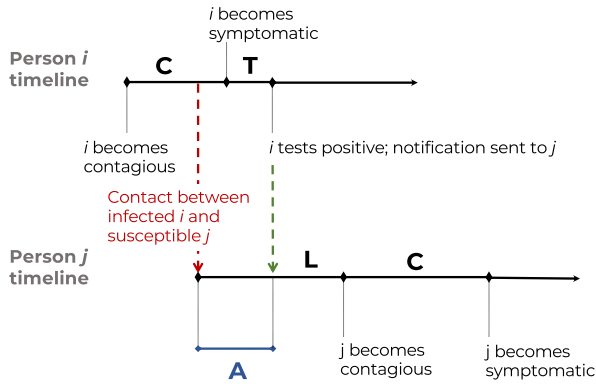


FIGURE 2. The timeline of the conversion to symptomatic.

from the time of infection to the time a person becomes contagious), the length of the infectious but asymptomatic period C , and the testing delay T (the time it takes for a test result to be available after the person has developed symptoms). Note that L and C are only determined by the properties of the specific disease. On the contrary, T is totally dependent on the efficacy of the testing system in place, hence it can be shortened by human interventions (e.g., using rapid tests rather than molecular ones or by scaling up testing facilities). The probability distribution of L , C , and T can be obtained from real data, when available (at the end of the section, we will discuss an example based on a realistic duration of the latent and infectious windows). Using their distribution, we can characterize the only missing time interval in Figure 2: A , which represented the time it takes for the app notification to pop up after contact. In Lemma 1 we derive interval A 's distribution.

Lemma 1: The random variable A describing the time interval between the at-risk contact and the time when the notification from the contact tracing app arrives is distributed as $C' + T$, i.e., as the sum (between random variables) of the residual infectious-but-asymptomatic period C' and the testing delay T .

Proof: As illustrated in Figure 2, A describes the time at which the contact tracing app notification arrives. This time corresponds to the interval between contact with an infectious person and the notification time, hence it includes a residual contagious period (which we denote with C') and the testing delay T . Thus, A is distributed as $C' + T$ (hence its PDF is given by the convolution of the PDF of C' and T [57]). Mathematically, C' can be obtained assuming that the contact between the susceptible and the contagious individuals appears during C as a random observer: as a result, C' can be derived as the residual duration [58] of C for the infectious person i (i.e., the time left before the person becomes symptomatic, hence they are discovered to be positive). Denoting by F_X the CCDF of a generic random variable X , the formula to calculate the residual time C' is the following [58]:

$$F_{C'}(t) = \frac{1}{E[C]} \int_t^\infty F_C(u) du. \quad (2)$$

Since C can be derived from real epidemic data, C' can also be calculated. \square

Now that A is fully characterized, by deriving its interplay with L and C we obtain p_E , p_I , and p_R in Lemma 2 below.

Lemma 2: The probabilities p_E , p_I , and p_R (associated, respectively, with catching a person in state E_T , I_T , and R) are given by the following:

$$\begin{aligned} p_E &= P(W < 0) \\ p_I &= \int_0^\infty P(W = w)P(C > w)dw \\ p_R &= P(W - C > 0), \end{aligned}$$

where we denote by W the difference $A - L$.

Proof: From Figure 2, we can see that p_E is equivalent to the probability of the notification arriving within the latent period (corresponding to $P(A < L)$). The probability that the notification arrives during the contagious and asymptomatic period ($P(L < A < L + C)$) yields p_I . The value of p_R can then be obtained complementing to 1 (or computing $P(A > L + C)$, i.e., the probability that the notification arrives when the individual is already symptomatic). Operatively, this results in the thesis. \square

Not for all distributions the above algebra of random variables yields closed-form solutions, but for some significant ones, it does, at least approximately. This happens, e.g., in the Normal case discussed in the next section (Sec. IV-A1). Closed-form solutions can also be obtained with exponential random variables. Once the probabilities p_E and p_I are derived, it is straightforward to obtain rates θ and ψ .

Theorem 2: The rates at which exposed and contagious people are removed (θ and ψ , respectively) are given by the following:

$$\begin{aligned} \theta &= p_E \alpha \beta, \\ \psi &= p_I \alpha \beta, \end{aligned} \quad (3)$$

where p_E and p_I are obtained as in Lemma 2.

Proof: The thesis simply follows from scaling the overall tracked contact rate $\alpha\beta$ by the probability that the exposed person is notified when still in the *exposed* state or in the *infectious* state. \square

1) EXAMPLE WITH NORMALLY DISTRIBUTED CHARACTERISTIC TIMES

For the sake of example, we can now obtain p_E , p_I , and p_R leveraging the typical average duration of the latent and contagious periods for the original COVID-19 epidemic. From [59], we obtain the average duration of the latent period ($\mathbb{E}[L] = \epsilon^{-1} = 3$ days) and that of the period before an infected becomes contagious ($\mathbb{E}[C] = \gamma^{-1} = 2$ days). Note that the expectations of L and C correspond to the inverse of ϵ and γ in the SEIR model of Section IV. For example, assume that L and C are normally distributed, each with a standard deviation 0.5 (the occurrence of negative values with this configuration is negligible). We also assume that T is normally distributed, with rate μ_T and standard deviation σ_T . It is easy to verify that $A = C' + T$ can be approximated as normally distributed as well, specifically

TABLE 1. Summary of notation.

NOTATION	DESCRIPTION
N	population size
S	susceptible population
E	people that have the disease but not yet symptoms
I	contagious & symptomatic people
R	removed people (either recovered or isolated)
$(\)_T$	subscript T denotes compartments that are tracked
$(\)_U$	subscript U denotes compartments that are <i>not</i> tracked
β	effective contact rate
ϵ^{-1}	time from infected to contagious
γ^{-1}	time from contagious to removed (recovery/isolation)
θ	rate at which tracked exposed are isolated
ψ	rate at which tracked infectious are isolated
α	fraction of population with the app installed and running
L	latent period
C	infectious-but-asymptomatic period
T	testing delay
A	time between at-risk contact and app notification
W	time to contagious after app notification
p_E	probability of alerting a person in state E_T
p_I	probability of alerting a person in state I_T
p_R	probability of alerting a person in state R

$A \sim \mathcal{N}(\mathbb{E}[C'] + \mu_T, \text{Var}(C') + \sigma_T^2)$. Since we are dealing with normally distributed variables, it is easy to obtain their difference and sum using the algebra of normally distributed random variables.

Leveraging the formulas we have obtained, we can now better understand the impact of testing delays on the ability to intercept infected people in each stage using contact tracing. In the following, we focus on a tagged pair of people (one tracked infectious i and one tracked susceptible j infected by i , analogously to Figure 3) and we study the probability that j is notified when in the *exposed*, *infectious*, or *removed* state, respectively, as we vary the testing delay. Note that, since we focus on a tagged pair of tracked people, this result does not depend on α , which is a population-level parameter. As Figure 3 shows, as long as the test turnaround is less than 2 days, the infected person is most likely caught while they are still not contagious. Vice versa, beyond a 4-day turnaround, we basically intercept only people that are already contagious or even symptomatic. As discussed above, the earlier we intercept infected people, the better. Small testing delays are thus a key ingredient of a containment plan.

V. CAN DIGITAL CONTACT TRACING TAME AN EPIDEMIC?

Now we use the model defined in the previous section and apply it to study the effectiveness of digital contact tracing in controlling the COVID-19 epidemic. To this end, we need COVID-19-specific estimates for β (effective contact rate), γ (transition rate to symptomatic), and ϵ (transition rate to contagious), which are the baseline (i.e., not depending on

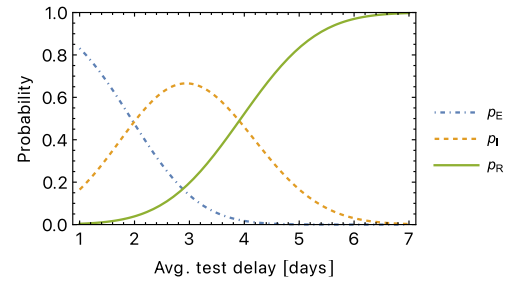


FIGURE 3. p_E , p_I , and p_R as the average testing delay increases.

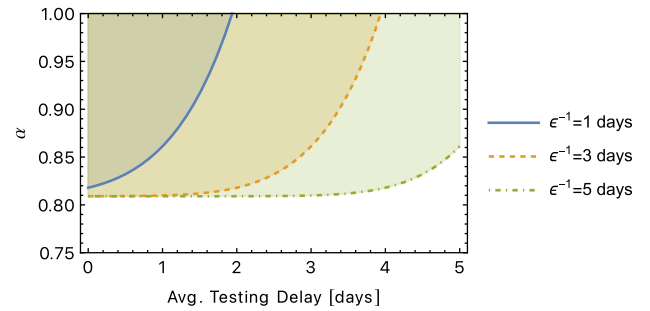


FIGURE 4. Controllability when the average latent period increases (corresponding to the decrease of ϵ). We consider $\epsilon^{-1} = \{1, 3, 5\}$ days. Control is attained in the shadowed regions.

the contact tracing process) transition rates in Figure 1. Then, θ and ψ (which instead are also determined by the contact tracing and the testing process) can be derived as discussed in Section IV-A. Note that the digital contact tracing process itself can be described with two parameters: α , the app uptake, and the test delay μ_T . In the following, we will assess their impact on the control condition C1 (Theorem 1) for different epidemic scenarios. Specifically, we consider three cases. First, we test the effect of an increasing latent period. Then, we study the effect of a longer infectious period with constant transmissibility (see below for details). Finally, we consider what happens when transmissibility increases.

Regarding the baseline epidemic parameters, generally γ and the ratio $\frac{\beta}{\gamma}$, corresponding to the basic reproduction number R_0 , are estimated early in an epidemic. Hence, in the following, β will be set to the value yielding the COVID-19 R_0 for the chosen γ . Intuitively, the basic reproduction number R_0 captures the transmissibility of a disease, i.e., the average number of cases directly generated by one infectious person in a population with a very large number of susceptibles. As we also see in the following, the larger R_0 , the more difficult it is to contain the epidemic with digital contact tracing (and in general).

We start (Figure 4) with a scenario with an average contagious but asymptomatic window length (γ^{-1}) equal to 2 days (typical of COVID-19) and fixed $R_0 = 2$ (this value correspond to the initial 2020 estimate for COVID-19 in [9]). We test the effect of an increasing latent period length ($\epsilon^{-1} \in \{1, 3, 5\}$ days) on the controllability. Specifically, we plot condition C1 in Equation 1 as a function of app uptake α and testing delay (the two directions along which digital contact tracing can be improved): C1 holds true in the

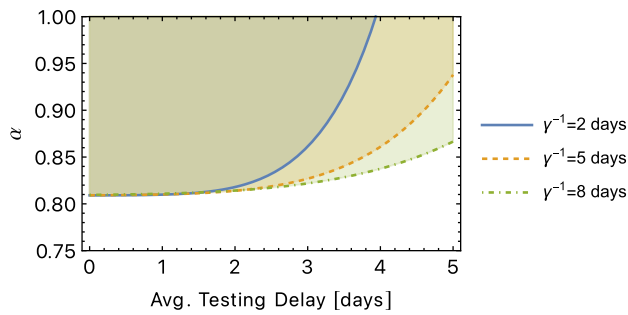


FIGURE 5. Controllability when the average length of the contagious period increases (corresponding to γ decreasing). We consider $\gamma^{-1} = \{2, 5, 8\}$ days. Control is attained in the shadowed regions.

shadowed areas of the plot (therefore, the epidemic is under control). Intuitively, the longer the latent period, the easier is the control of the epidemic because we have more time to intercept tracked people before they become contagious. Figure 4 confirms this: as ϵ^{-1} (the latent window) increases, the importance of small testing delays is reduced, and an app uptake above 80% may be sufficient to control the outbreak. Looking again at Figure 4, we also note that an increase in ϵ^{-1} induces a temporal shift on the controllability boundary (the curves delimiting the shadowed areas in the figure): in other words, the controllability boundary is affected by a shift along the x-axis as the latent window increases.

Next, we consider the effect of the infectious period when the transmissibility is kept constant. Without contact tracing, two epidemics with the same transmissibility evolve in the same way (since the controllability condition reduces to $R_0 < 1$). However, the length of the infectious period affects the effectiveness of contact tracing. Therefore, the digital controllability of epidemics with the same R_0 is different if their infectious periods are different. In Figure 5, we set the latent window ϵ^{-1} at 3 days (its average value for COVID-19) and vary γ (the duration of the infectious period) in $\{2, 5, 8\}$ days. Note that we want R_0 to remain fixed (no change in transmissibility), so we vary β accordingly. This implies that each person, on average, infects the same number of people in all three cases captured by Figure 5. We observe that the longer the contagious period, the easier the containment (since the shadowed area expands). This was expected because a longer contagious period increases the probability of catching infected people before they become symptomatic. With respect to the controllability boundary (the curves delimiting the shadowed areas in the figure), a change in γ induces a change in the convexity of the boundary but no horizontal shift.

Finally, in Figure 6 we fix the latent period ϵ^{-1} and the contagious period γ^{-1} to their typical COVID-19 values of 3 and 2 days, respectively, and we change the R_0 of the epidemic by varying β . Note that this analysis is especially important, given the rise of new variants with increased transmissibility (and therefore greater R_0). We study $R_0 \in \{2, 3, 4, 5, 6\}$. $R_0 = 2$ is the initial estimate for COVID-19 (original strain), then revised to be much higher in some areas (e.g., the estimate in [60] is $R_0 \sim 4$).

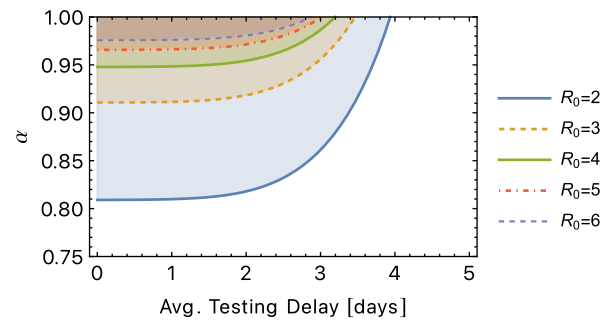


FIGURE 6. Controllability when the $R_0 = \frac{\beta}{\gamma}$ increases (we fix γ and we increase β). We study $R_0 \in \{2, 3, 4, 5, 6\}$. Control is attained in the shadowed regions.

The Alpha variant (B.1.1.7 lineage) is estimated to feature at least 40% higher R [61] with respect to the original strain, while the Delta variant (B.1.617.2) has transmissibility estimated between 6 and 7 [62], [63]. Note that the apparent even higher transmissibility of the recent Omicron variant (B.1.1.529) seems to be due to immune evasion (e.g., vaccines not as effective as for previous variants) rather than to an actual increase in basic transmissibility [64]. Figure 6 shows that, as expected, the impact of increasing R_0 is much more disruptive than that of different latent/contagious windows. Specifically, even with instantaneous testing, the minimum uptake α needed to control the epidemic increases as R_0 increases. This means that even a small fraction of untracked people can wreak havoc on the containment measures. In practice, however, with $R_0 = 4$, the control of the epidemic through digital contact tracing becomes impossible: an uptake above 95% is unrealistic for all the reasons discussed in Section I (e.g., technical problems with old smartphones, distrust by a fraction of the population). In this case, R_0 must also be reduced by exploiting mitigation measures (social distancing, masks), in order to reduce the probability of infection upon contact, hence β .

VI. CONCLUSION

In this work, we have discussed the modeling efforts for COVID-19 and we have proposed a SEIR model that factors in digital contact tracing and is capable of producing a closed-form condition on the controllability of the epidemic. Leveraging this model, we have studied how the penetration of digital contact tracing apps within the population impacts the control of the epidemic. We have found that the penetration must be in general high, hence digital contact tracing may not be sufficient to contain an epidemic, even with a fast turnaround of tests. Additional mitigation strategies must be implemented, such as social distancing and mask wearing. Additionally, the impact of digital contact tracing is highest when the testing delay is low. If the test turnaround is greater than 4 days, digital contact tracing has zero impact on containment. In future work, we plan to extend the model to account for more variability among nodes, e.g., in terms of contact patterns, (along the lines of [65]), while still striving for closed-form controllability conditions.

**APPENDIX
PROOF OF THEOREM 1**

Proof: We start by writing the ODE system corresponding to Figure 1:

$$\begin{aligned} \frac{dS_U}{dt} &= -\beta(1-\alpha)(I_U + I_T) \\ \frac{dE_U}{dt} &= \beta(1-\alpha)(I_U + I_T) - \epsilon E_U \\ \frac{dI_U}{dt} &= \epsilon E_U - \gamma I_U \\ \frac{dS_T}{dt} &= -\beta\alpha(I_U + I_T) \\ \frac{dE_T}{dt} &= \beta\alpha I_U + (\beta\alpha - \theta)I_T - \epsilon E_T \\ \frac{dI_T}{dt} &= \epsilon E_T - (\gamma + \psi)I_T \\ \frac{dR}{dt} &= \gamma I_U + (\gamma + \theta + \psi)I_T. \end{aligned} \tag{5}$$

The corresponding system of ODE can be rewritten in matrix form as $\mathbf{y}' = \mathbf{A}\mathbf{y}$, where $\mathbf{y} = [E_U, I_U, E_T, I_T]^T$ and \mathbf{A} is given by the following:

$$\mathbf{A} = \begin{bmatrix} -\epsilon & \beta(1-\alpha) & 0 & \beta(1-\alpha) \\ \epsilon & -\gamma & 0 & 0 \\ 0 & \beta\alpha & -\epsilon & \beta\alpha - \theta \\ 0 & 0 & \epsilon & -\gamma - \psi \end{bmatrix}. \tag{6}$$

The system in (5) describes a dynamic system. Its stability (corresponding to the epidemic being under control or not) is assessed by studying its eigenvalues (see, e.g., [60]), which correspond to the roots of the characteristic polynomial $p_A(x)$ of matrix \mathbf{A} . In fact, since the solutions to a system of linear ODE $\mathbf{y}' = \mathbf{A}\mathbf{y}$ are of the form $y(t) = \sum_i c_i * e^{r_i t}$ [66] (where c_i 's are constants and r_i 's the eigenvalues/roots), it is clear that a positive root introduces instability into the system, because there would be an exponential function with a positive argument, hence an exponential growth in the epidemic. Therefore, we can study the roots of the characteristic polynomial $p_A(x)$ to assess under which conditions only negative roots exist. In order to avoid a trivial case, we assume $\beta > \gamma$ (i.e., the epidemic is not under control without contact tracing). Using Descartes' rule of signs, we can derive the number of positive and negative roots of $p_A(x)$ without actually having to solve the polynomial (finding a closed form for the roots would not be feasible in this case). Starting with positive roots, we observe the following signs:

$$\{+, +, \text{sgn}(k_2), \text{sgn}(k_1), \text{sgn}(k_0)\}, \tag{7}$$

where we have expressed $p_A(x)$ as $\sum_i k_i x^i$, sgn is the sign function (where $\text{sgn}(\cdot) = 1$ corresponds to sign +, $\text{sgn}(\cdot) = -1$ corresponds to -), and k_2, k_1, k_0 are given by the following:

$$\begin{aligned} k_2 &= -\beta\epsilon + \epsilon^2 + \gamma(\gamma + \psi) + \epsilon(4\gamma + 2\psi + \theta), \\ k_1 &= \epsilon(2\gamma + \psi + \theta) \\ &\quad + \gamma[2(\gamma + \psi) + \theta] - \beta(\epsilon + \gamma + \psi - \psi\alpha), \\ k_0 &= -[(\beta - \gamma)(\gamma + \psi + \theta)] + \beta(\psi + \theta)\alpha. \end{aligned} \tag{8}$$

TABLE 2. All possible sign permutations for the coefficients of the characteristic polynomial of \mathbf{A} . The shadowed areas correspond to unfeasible permutations.

k_4	k_3	k_2	k_1	k_0	sign changes
+	+	+	+	+	0
+	+	+	+	-	1
+	+	+	-	+	2
+	+	+	-	-	1
+	+	-	+	+	2
+	+	-	+	-	3
+	+	-	-	+	2
+	+	-	-	-	1

By studying the functions $\text{sgn}(k_2), \text{sgn}(k_1), \text{sgn}(k_0)$, we obtain the following relationships between the coefficients' signs:

$$\text{sgn}(k_0) = 1 \Rightarrow \text{sgn}(k_1) = 1 \Rightarrow \text{sgn}(k_2) = 1. \tag{9}$$

In other words, since the rates must be all positive and $\alpha \in [0, 1]$, when coefficient k_0 is positive, k_2 and k_1 must also be positive. This implies that not all possible sign permutations in Equation 7 are attainable, as illustrated in Table 2. Discarding unattainable permutations, we can have at most one sign change across the coefficients of the polynomial, which implies at most one positive root. It thus follows that the condition under which we observe no sign changes is also the condition under which the epidemic can be controlled (zero positive roots). Thanks to Equation 9, we know that $\text{sgn}(k_0) = 1$ is a sufficient condition for this to happen. Then, solving for $k_0 > 0$ (with k_0 defined Equation 8), we obtain condition C1 in Equation 1.

To conclude the proof, we only need to verify that there are no complex roots. This is easy to do by applying again Descartes' rule, this time to $p_A(-x)$. To this aim, we need to change the coefficient sign of odd-power terms (i.e., k_3, k_1) in Table 2 and count the sign changes. By summing the sign changes for positive and negative roots corresponding to the same permutation (equivalently, by summing the sign changes per corresponding row in Table 2 and Table 2 with the sign of odd-power terms changed), we obtain the total number of real roots. If we do the math, we discover that the total sign changes are at most 4, hence $p_A(x)$ features four real roots. Then, the number of complex roots is given by the difference between the degree of the polynomial and the maximum number of real roots. Since $p_A(x)$ is a polynomial of degree 4, we know that there are no complex roots. □

REFERENCES

- [1] C. Fraser, S. Riley, R. M. Anderson, and N. M. Ferguson, "Factors that make an infectious disease outbreak controllable," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. 16, pp. 6146–6151, Apr. 2004.
- [2] H. Cho, D. Ippolito, and Y. William, "Contact tracing mobile apps for COVID-19: Privacy considerations and related trade-offs," 2020, *arXiv:2003.11511*.
- [3] R. Raskar, I. Schunemann, R. Barbar, K. Vilcans, J. Gray, P. Vepakomma, S. Kapa, A. Nuzzo, R. Gupta, A. Berke, and D. Greenwood, "Apps gone rogue: Maintaining personal privacy in an epidemic," 2020, *arXiv:2003.08567*.

- [4] M. Nanni, "Give more data, awareness and control to individual citizens, and they will help COVID-19 containment," *Trans. Data Privacy*, vol. 13, no. 1, pp. 61–66, 2020.
- [5] N. Oliver, E. Letouzé, H. Sterly, S. Delataille, M. De Nadai, B. Lepri, R. Lambiotte, R. Benjamins, C. Cattuto, V. Colizza, and N. de Cordes, "Mobile phone data and COVID-19: Missing an opportunity?" 2020, *arXiv:2003.12347*.
- [6] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha, "A survey of COVID-19 contact tracing apps," *IEEE Access*, vol. 8, pp. 134577–134601, 2020.
- [7] C. Troncoso, D. Bogdanov, E. Bugnion, S. Chatel, C. Cremers, S. Gürses, J.-P. Hubaux, D. Jackson, J. R. Larus, W. Lueks, R. Oliveira, M. Payer, B. Preneel, A. Pyrgelis, M. Salathé, T. Stadler, and M. Veale, "Deploying decentralized, privacy-preserving proximity tracing," *Commun. ACM*, vol. 65, no. 9, pp. 48–57, Sep. 2022, doi: [10.1145/3524107](https://doi.org/10.1145/3524107).
- [8] L. Garg, E. Chukwu, N. Nasser, C. Chakraborty, and G. Garg, "Anonymity preserving IoT-based COVID-19 and other infectious disease contact tracing model," *IEEE Access*, vol. 8, pp. 159402–159414, 2020.
- [9] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing," *Science*, vol. 368, no. 6491, May 2020, Art. no. eabb6936.
- [10] T. Li, C. Cobb, D. Jackie, Yang, S. Baviskar, Y. Agarwal, B. Li, L. Bauer, and J. I. Hong, "What makes people install a COVID-19 contact-tracing app? Understanding the influence of app design and individual difference on contact-tracing app adoption intention," 2020, *arXiv:2012.12415*.
- [11] V. Colizza, E. Grill, R. Mikolajczyk, C. Cattuto, A. Kucharski, S. Riley, M. Kendall, K. Lythgoe, D. Bonsall, C. Wymant, L. Abeler-Dörner, L. Ferretti, and C. Fraser, "Time to evaluate COVID-19 contact-tracing apps," *Nature Med.*, vol. 27, no. 3, pp. 361–362, Mar. 2021, doi: [10.1038/s41591-021-01236-6](https://doi.org/10.1038/s41591-021-01236-6).
- [12] M. Conti, S. K. Das, C. Bisdikian, M. Kumar, L. M. Ni, A. Passarella, G. Roussos, G. Tröster, G. Tsudik, and F. Zambonelli, "Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber-physical convergence," *Pervasive Mobile Comput.*, vol. 8, no. 1, pp. 2–21, 2012.
- [13] M. M. Böhmer, "Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: A case series," *Lancet Infectious Diseases*, vol. 20, no. 8, pp. 920–928, Aug. 2020.
- [14] Z. Hu, C. Song, C. Xu, G. Jin, Y. Chen, X. Xu, H. Ma, W. Chen, Y. Lin, Y. Zheng, J. Wang, Z. Hu, Y. Yi, and H. Shen, "Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China," *Sci. China Life Sci.*, vol. 63, no. 5, pp. 706–711, May 2020.
- [15] G. Qian, N. Yang, A. H. Y. Ma, L. Wang, G. Li, X. Chen, and X. Chen, "COVID-19 transmission within a family cluster by presymptomatic carriers in China," *Clin. Infectious Diseases*, vol. 71, no. 15, pp. 861–862, Jul. 2020.
- [16] C. Rothe, M. Schunk, P. Sothmann, G. Bretzel, G. Froeschl, C. Wallrauch, T. Zimmer, V. Thiel, C. Janke, W. Guggemos, M. Seilmaier, C. Drosten, P. Vollmar, K. Zwirgmaier, S. Zange, R. Wölfel, and M. Hoelscher, "Transmission of 2019-nCoV infection from an asymptomatic contact in Germany," *New England J. Med.*, vol. 382, no. 10, pp. 970–971, Mar. 2020.
- [17] Y. Wang, J. Tong, Y. Qin, T. Xie, J. Li, J. Li, J. Xiang, Y. Cui, E. S. Higgs, J. Xiang, and Y. He, "Characterization of an asymptomatic cohort of SARS-CoV-2 infected individuals outside of Wuhan, China," *Clin. Infectious Diseases*, vol. 71, no. 16, pp. 2132–2138, Oct. 2020.
- [18] P. Yu, J. Zhu, Z. Zhang, and Y. Han, "A familial cluster of infection associated with the 2019 novel coronavirus indicating possible person-to-person transmission during the incubation period," *J. Infectious Diseases*, vol. 221, no. 11, pp. 1757–1761, May 2020.
- [19] Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, and M. Wang, "Presumed asymptomatic carrier transmission of COVID-19," *JAMA*, vol. 323, no. 14, p. 1406, Apr. 2020.
- [20] W. E. Wei, Z. Li, C. J. Chiew, S. E. Yong, M. P. Toh, and V. J. Lee, "Presymptomatic transmission of SARS-CoV-2—Singapore, January 23–March 16, 2020," *MMWR. Morbidity Mortality Weekly Rep.*, vol. 69, no. 14, pp. 411–415, Apr. 2020.
- [21] R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirgmaier, C. Drosten, and C. Wendtner, "Virological assessment of hospitalized patients with COVID-2019," *Nature*, vol. 25, pp. 1–10, Apr. 2020.
- [22] A. Kimball, "Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King county, Washington, Mar. 2020," *MMWR. Morbidity Mortality Weekly Rep.*, vol. 69, no. 13, pp. 377–381, Apr. 2020. [Online]. Available: http://www.cdc.gov/mmwr/volumes/69/wr/mm6913e1.htm?s_cid=mm6913e1_w
- [23] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)," *Science*, vol. 368, no. 6490, pp. 489–493, May 2020.
- [24] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, and X. Xing, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *New England J. Med.*, vol. 382, no. 13, pp. 1199–1207, Jan. 2020.
- [25] D. Sutton, K. Fuchs, M. D'Alton, and D. Goffman, "Universal screening for SARS-CoV-2 in women admitted for delivery," *Obstetric Anesthesia Dig.*, vol. 40, no. 4, p. 199, 2020.
- [26] X. He, "Temporal dynamics in viral shedding and transmissibility of COVID-19," *Nature Med.*, vol. 26, no. 5, pp. 672–675, May 2020, doi: [10.1038/s41591-020-0869-5](https://doi.org/10.1038/s41591-020-0869-5).
- [27] L. Luo, "Modes of contact and risk of transmission in COVID-19 among close contacts," *MedRxiv*, pp. 1–37, Mar. 2020, doi: [10.1101/2020.03.24.20042606](https://doi.org/10.1101/2020.03.24.20042606).
- [28] L. Morawska and D. K. Milton, "It is time to address airborne transmission of COVID-19," *Clin. Infectious Diseases*, vol. 73, no. 11, pp. e3891–e3982, Aug. 2020.
- [29] Z. Chagla, S. Hota, S. Khan, and D. Mertz, "Airborne transmission of COVID-19," *Clin. Infectious Diseases*, vol. 73, no. 11, pp. e3981–e3982, May 2020.
- [30] M. Klompas, M. A. Baker, and C. Rhee, "Airborne transmission of SARS-CoV-2: Theoretical considerations and available evidence," *Jama*, vol. 324, no. 5, p. 441, Aug. 2020.
- [31] M. Z. Bazant and J. W. M. Bush, "A guideline to limit indoor airborne transmission of COVID-19," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 17, Apr. 2021, Art. no. e2018995118, doi: [10.1073/pnas.2018995118](https://doi.org/10.1073/pnas.2018995118).
- [32] L. Bourouiba, "Turbulent gas clouds and respiratory pathogen emissions: Potential implications for reducing transmission of COVID-19," *Jama*, vol. 323, no. 18, pp. 1837–1838, May 2020.
- [33] E. Hunter, B. M. Namee, and J. Kelleher, "An open-data-driven agent-based model to simulate infectious disease outbreaks," *PLoS ONE*, vol. 13, no. 12, Dec. 2018, Art. no. e0208775.
- [34] S. Venkatramanan, B. Lewis, J. Chen, D. Higdon, A. Vullikanti, and M. Marathe, "Using data-driven agent-based models for forecasting emerging infectious diseases," *Epidemics*, vol. 22, pp. 43–49, Mar. 2018.
- [35] J. Hackl and T. Dubernet, "Epidemic spreading in urban areas using agent-based transportation models," *Future Internet*, vol. 11, no. 4, p. 92, Apr. 2019.
- [36] J. Parker and J. M. Epstein, "A distributed platform for global-scale agent-based models of disease transmission," *ACM Trans. Model. Comput. Simul.*, vol. 22, no. 1, pp. 1–25, Dec. 2011.
- [37] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 3, pp. 7280–7287, May 2002.
- [38] J. J. Grefenstette, S. T. Brown, R. Rosenfeld, J. DePasse, N. T. Stone, P. C. Cooley, W. D. Wheaton, A. Fyshe, D. D. Galloway, A. Sriram, H. Guclu, T. Abraham, and D. S. Burke, "FRED (a framework for reconstructing epidemic dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations," *BMC Public Health*, vol. 13, no. 1, pp. 1–14, Dec. 2013.
- [39] F. Liu, W. T. A. Enanoria, J. Zipprich, S. Blumberg, K. Harriman, S. F. Ackley, W. D. Wheaton, J. L. Allpress, and T. C. Porco, "The role of vaccination coverage, individual behaviors, and the public health response in the control of measles epidemics: An agent-based simulation for California," *BMC Public Health*, vol. 15, no. 1, pp. 1–16, Dec. 2015.
- [40] A. Tomy, M. Razzanelli, F. Di Lauro, D. Rus, and C. Della Santina, "Estimating the state of epidemics spreading with graph neural networks," *Nonlinear Dyn.*, vol. 109, no. 1, pp. 249–263, Jul. 2022, doi: [10.1007/s11071-021-07160-1](https://doi.org/10.1007/s11071-021-07160-1).
- [41] C. Fritz, E. Dorigatti, and D. Rügamer, "Combining graph neural networks and spatio-temporal disease models to improve the prediction of weekly COVID-19 cases in Germany," *Sci. Rep.*, vol. 12, no. 1, pp. 1–18, 2022, doi: [10.1038/s41598-022-07757-5](https://doi.org/10.1038/s41598-022-07757-5).
- [42] J. Gao, R. Sharma, C. Qian, L. M. Glass, J. Spaeder, J. Romberg, J. Sun, and C. Xiao, "STAN: Spatio-temporal attention network for pandemic prediction using real-world evidence," *J. Amer. Med. Inform. Assoc.*, vol. 28, no. 4, pp. 733–743, Mar. 2021.

- [43] A. Kapoor, X. Ben, L. Liu, B. Perozzi, M. Barnes, M. Blais, and S. O'Banion, "Examining COVID-19 forecasting using spatio-temporal graph neural networks," 2020, *arXiv:2007.03113*.
- [44] G. Cutura, B. Li, A. Swami, and S. Segarra, "Deep Demixing: Reconstructing the evolution of epidemics using graph neural networks," in *Proc. 29th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 2204–2208.
- [45] C. Shah, N. Dehmamy, N. Perra, M. Chinazzi, A.-L. Barabási, A. Vespignani, and R. Yu, "Finding patient zero: Learning contagion source with graph neural networks," 2020, *arXiv:2006.11913*.
- [46] S. Dash, C. Chakraborty, S. K. Giri, S. K. Pani, and J. Frnda, "BIFM: Big-data driven intelligent forecasting model for COVID-19," *IEEE Access*, vol. 9, pp. 97505–97517, 2021.
- [47] F. Brauer and C. Castillo-Chavez, *Mathematical Models in Population Biology and Epidemiology*. New York, NY, USA: Springer, 2012.
- [48] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. D. Filippo, A. D. Matteo, and M. Colaneri, "Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy," *Nature Med.*, vol. 35, pp. 1–6, Apr. 2020.
- [49] M. E. J. Newman, "Threshold effects for two pathogens spreading on a network," *Phys. Rev. Lett.*, vol. 95, no. 10, Sep. 2005, Art. no. 108701.
- [50] J. Riou and C. L. Althaus, "Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020," *Eurosurveillance*, vol. 25, no. 4, Jan. 2020, Art. no. 2000058.
- [51] N. Ferguson, D. Laydon, G. N. Gilani, N. Imai, K. Ainslie, M. Baguein, S. Bhatia, A. Boonyasiri, Z. Cucunuba Perez, and G. Cuomo-Dannenburg, "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand." Imperial College London, London, U.K., COVID-19 Rep. 9, 2020. [Online]. Available: <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>
- [52] N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-M. Jung, B. Yuan, R. Kinoshita, and H. Nishiura, "Epidemiological characteristics of novel coronavirus infection: A statistical analysis of publicly available case data," *J. Clin. Med.*, vol. 9, no. 2, pp. 1–9, Feb. 2020, Art. no. 538, doi: [10.1101/2020.01.26.20018754](https://doi.org/10.1101/2020.01.26.20018754).
- [53] Q. Li, "Early transmission dynamics in Wuhan, China, of novel coronavirus—Infected pneumonia," *New England J. Med.*, vol. 382, no. 13, pp. 1199–1207, Jan. 2020.
- [54] I. Dorigatti, "Report 4: Severity of 2019-novel coronavirus (nCoV)," Imperial College London, London, U.K., COVID-19 Rep. 4, 2020. [Online]. Available: <https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-16-COVID19-Report-9.pdf>, doi: [10.25561/77154](https://doi.org/10.25561/77154).
- [55] A. Arenas, W. Cota, J. Gomez-Gardenes, S. Gomez, C. Granell, J. T. Matamalas, D. Soriano-Panos, and B. Steinegger, "Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement," *MedRxiv*, pp. 1–48, Apr. 2020, doi: [10.1101/2020.04.06.20054320](https://doi.org/10.1101/2020.04.06.20054320).
- [56] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [57] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä, and L. E. Meester, *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Cham, Switzerland: Springer, 2005.
- [58] C. Boldrini, M. Conti, and A. Passarella, "From Pareto inter-contact times to residuals," *IEEE Commun. Lett.*, vol. 15, no. 11, pp. 1256–1258, Nov. 2011.
- [59] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, "SARS-CoV-2 (COVID-19) by the numbers," *eLife*, vol. 9, Apr. 2020, Art. no. e57309.
- [60] F. Casella, "Can the COVID-19 epidemic be controlled on the basis of daily test reports?" *IEEE Control Syst. Lett.*, vol. 5, no. 3, pp. 1079–1084, Jul. 2021.
- [61] N. G. Davies, "Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England," *Science*, vol. 372, no. 6538, Apr. 2021, Art. no. eabg3055, doi: [10.1126/science.abg3055](https://doi.org/10.1126/science.abg3055).
- [62] R. Pung, T. M. Mak, A. J. Kucharski, and V. J. Lee, "Serial intervals observed in SARS-CoV-2 B.1.617.2 variant cases," *Lancet*, vol. 398, no. 10303, pp. 837–838, 2021.
- [63] J. Dagpunar, "Interim estimates of increased transmissibility, growth rate, and reproduction number of the Covid-19 B.1.617.2 variant of concern in the United Kingdom," *MedRxiv*, pp. 1–7, Jun. 2021, doi: [10.1101/2021.06.03.21258293](https://doi.org/10.1101/2021.06.03.21258293).
- [64] F. P. Lyngse, L. H. Mortensen, M. J. Denwood, L. E. Christiansen, C. H. Møller, R. L. Skov, K. Spiess, A. Fomsgaard, M. M. Lassaunière, M. Rasmussen, M. Stegger, C. Nielsen, R. N. Sieber, A. S. Cohen, F. T. Møller, M. Overvad, K. Mølbak, T. G. Krause, and C. T. Kirkeby, "Sars-CoV-2 omicron VOC transmission in Danish households," *MedRxiv*, pp. 1–31, Dec. 2021, doi: [10.1101/2021.12.27.21268278](https://doi.org/10.1101/2021.12.27.21268278).
- [65] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, no. 1, pp. 1–11, Jul. 2002.
- [66] M. Braun, *Differential Equations and Their Applications An Introduction to Applied Mathematics*. New York, NY, USA: Springer, 1993.



CHIARA BOLDRINI is currently a Researcher at the IIT-CNR. She has published more than 50 papers on these topics. She is also the IIT-CNR Co-PI for H2020 SoBigData++ and HumaneE-AI-Net Projects, and was involved in several EC projects since FP7. Her research interests include computational social sciences, decentralized AI, urban data science, mobile, and ubiquitous systems. She is on the Editorial Board of *Pervasive and Mobile Computing* (Elsevier) and *Computer Communications* (Elsevier). She was the Lead Guest Editor of the PMC Special Issue on IoT for Fighting COVID-19. She has served as the TPC Chair for IEEE SmartComp'22, the TPC Vice-Chair for IEEE PerCom'21, and over the years, she has been in the Organizing Committee of several IEEE and ACM conferences/workshops, including IEEE PerCom and ACM MobiHoc.



ANDREA PASSARELLA received the Ph.D. degree, in 2005. He is currently a Research Director at the Institute for Informatics and Telematics (IIT), National Research Council of Italy (CNR). Prior to joining IIT, he was with the Computer Laboratory, University of Cambridge, U.K. He is the coauthor of the book *Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs* (Elsevier, 2015). He has published more than 200 papers on online and mobile social networks, decentralized AI, next generation internet, opportunistic, and *ad-hoc* and sensor networks, receiving the Best Paper Award at IFIP Networking 2011 and IEEE WoWMoM 2013. He has served as the General Chair for IEEE PerCom 2022. He is the Founding Associate Editor-in-Chief of *Online Social Networks* (Elsevier). He was the Guest Co-Editor of several special sections in ACM and Elsevier journals. He is the PI of the EU CHIST-ERA SAI (Social Explainable AI) project.



MARCO CONTI is currently the CNR Research Director and the Director of the IIT-CNR Institute. He has published more than 400 scientific articles related to design, modeling, and experimentation of computer and communication networks, pervasive systems, and online social networks. He has received several awards, including the Best Paper Award at IFIP Networking 2011, IEEE ISCC 2012, and IEEE WoWMoM 2013. He was included in the "2017 Highly Cited Researchers" list compiled by Web of Science for the most cited articles in computer science. He served as the General Chair/the Program Chair of several major conferences, including IFIP Networking 2002, IEEE WoWMoM 2005 and 2006, IEEE PerCom 2006 and 2010, ACM MobiHoc 2006, IEEE MASS 2007, and IEEE Smart-Comp 2021. He is the Founding Editor-in-Chief of *Online Social Networks and Media* and the Editor-in-Chief for Special Issues of Pervasive and Mobile Computing. From 2009 to 2018, he was the Editor-in-Chief of *Computer Communications*.

...