Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hawkes process model for the propagation of COVID-19: simple analytical results

Juan Valentin Escobar (escobar@fisica.unam.mx)
Instituto de Física, Universidad Nacional Autónoma de Mexico https://orcid.org/0000-0001-6048-2011

Research Article

Keywords: Hawkes Process, COVID-19, Exponential Bare Memory Kernel

Posted Date: February 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-43940/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published at EPL (Europhysics Letters) on November 3rd, 2020. See the published version at https://doi.org/10.1209/0295-5075/131/68005.

A Hawkes process model for the propagation of COVID-19: simple analytical results

JUAN V ESCOBAR [1(a)],

[1] *Instituto de Física, Departamento de Física Química, Universidad Nacional Autónoma de México, 04510 Coyoacán, Mexico City, Mexico*

received and accepted dates provided by the publisher
other relevant dates provided by the publisher

PACS 89.75.Hc – Networks and genealogical trees
PACS 89.75.-k – Complex systems
PACS 02.50.Ga – Markov processes

Abstract – We present a model for the COVID-19 epidemic that offers analytical expressions for the newly registered and latent cases. This model is based on an epidemic branching process with latency that is greatly simplified when the bare memory kernel is given by an exponential function as observed in this pandemic. We expose the futility of the concept of "bending the curve" of the epidemic as long as the number of latent cases is not depleted. Our model offers the possibility of laying out different scenarios for the evolution of the epidemic in different countries based on the most recent observations and in terms of only two constants obtained from clinical trials.

Introduction. – For the past seven months the COVID-19 pandemic has ravaged the entire world bringing with it devastating human and economic losses. Enormous efforts have been undertaken to predict the spread of this disease [1], in particular to quantify how different degrees of social distancing and travel restrictions may be affecting it [2-5]. However, partly because it is extraordinarily difficult to predict people's behavior, many models that can excel at *explaining* what has happened, tend to have a harder time *predicting* the evolution of the epidemic [6] except over short times. For example, SIR-type models [7]—that compartmentalize a population and lets its components evolve through a group of coupled differential equations— are highly sensitive to both initial conditions and parameters which makes them unreliable as tools to predict, for instance, when the peak of the pandemic will occur [8]. Failing to predict this turning point has become a common source of frustration because public opinion in some countries seems to have equated reaching the peak of the epidemic to having "tamed" it. This evidences a general lack of feeling of how safe it is to ease social distancing restrictions necessary to reopen the economy given the information available to the public. In most countries, this information consists of the daily number of new cases (herein called "registered" cases) as well as the daily new deaths. Besides SIR-type models, other valuable and effective approaches consist of agent-based simulations that incorporate census and mobility data [2,9]. These models lay out different scenarios depending on the degree of compliance of social distancing measures. But in either case, the exact data being used as well as the assumptions and algorithms are not always available to the public or straightforward to understand, let alone reproduce [6].

In this work we develop a simple epidemiologic model for the propagation of the current COVID-19 pandemic that is amenable to analytical solutions easy to implement. Our model evidences that bending the epidemiological curve is not as relevant as depleting the number of latent cases (contagious, but not registered yet), while providing simple expressions for this quantity from publicly available data. Finally, we lay out three different scenarios for the evolution of the epidemic in the USA. Other studies that lay out scenarios for the evolution of the pandemic have also used publicly available data [10].

The model. – Figure 1 shows the time series of the registered weekly new cases (red dots) for four different countries that have handled the epidemic very differently along with the number of latent cases (solid blue traces) calculated with our model. To construct it, we make use of the fact that in those countries in which the peak of a first outbreak has been reached, the relaxation of this dynamics is well described by a decaying exponential (insets of Figs. 1a and 1b). Epidemic processes for which the analytic form of the relaxation dynamics is known can be conveniently modeled by the so-called self-excited Hawkes conditional Poisson processes [11].

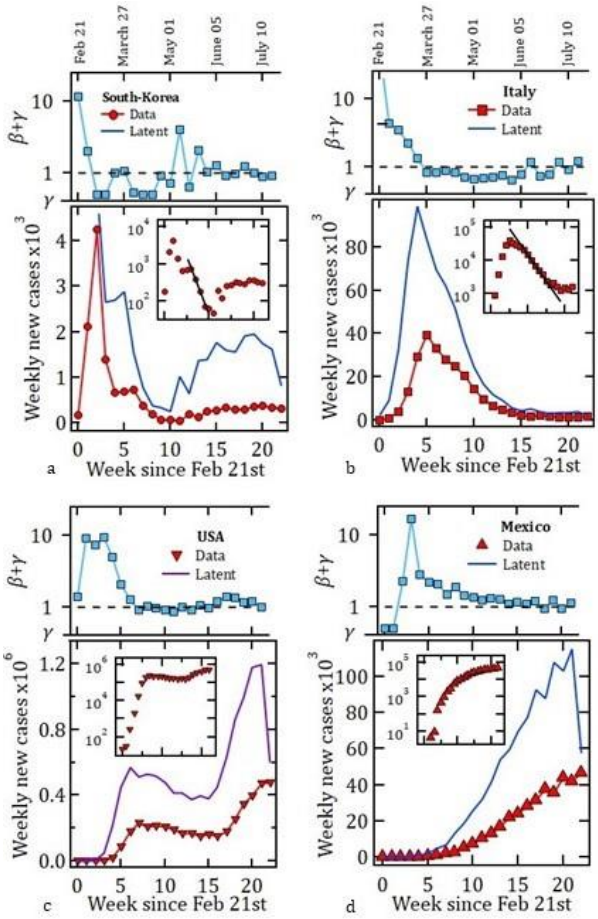

Fig. 1: Weekly registered new cases (dots) and calculated latent ones (Eq. 3), along with the corresponding reproduction numbers $R_i=\beta_i+\gamma$ (Eq. 4) for (a) South Korea, (b) Italy, (c) USA and (d) Mexico. See main text for details. Insets show the weekly new cases in log-linear scale. The solid black line in (a) depicts an exponentially decaying activity (Eq. 5 with β=0) setting γ = 0.5, as calculated from clinical studies (see Methods). This apparent agreement suggests contagion events were successfully brought to a halt in S. Korea during that relaxation period. In the inset of Fig. 1b), the solid line is the best fit to Eq. 5 (β=0.194 ±.004) assuming γ =0.5 as before.

These processes are composed of two main ingredients: the first one is the "branching ratio" that accounts for contagion. The second one is a latency function—also known as the bare memory kernel— that incorporates the fact that there is a statistical lag between contagion and the specific action whose aggregated dynamics is recorded. We propose that the weekly new observed or "registered" COVID-19 cases at time t— henceforth called the "activity" and denoted by λ_t— obeys the dynamics from a discrete-time Hawkes processes which in general takes the form:

$$\lambda_t = S_t + \sum_{i,t_i \le t}^{t} \beta_i \varphi_{t-t_i} \qquad (1)$$

where φ_t is the latency function, β_i is the number of people infected of first generation (*i.e.* directly infected by individual i at time t_i) that will become registered at any future time, and the term S_t represents the exogenous sources of new cases. In the case at hand, these sources are the number of people already infected arriving in some new city [9]. When the latency function is given by a decaying exponential with constant $(1/\tau)$ and $\beta_i = \beta$ for all i, the solution of Eq. 1 takes the form

$$\lambda_t \propto e^{-\frac{t}{\tau}(1-\theta)}, \qquad (2)$$

where the constant θ is a function of β, as we show below in more detail. Eq. 2 reveals a key feature characterizing Hawkes processes: when the latency function is given by either a power law —as in the case of many human activities [12-14] and earthquakes [15] — or by a decaying exponential [16] as in the present case, the activity during relaxation is slowed down, but it remains described by the same function. In other words, contagion renormalizes the function that describes the activity in the absence of exogenous sources. Thus, if the form of the bare memory kernel is known, the relaxation dynamics can give information about the magnitude of the contagion taking place during the relaxation period. In general, the activity derived from Hawkes processes is non-Markovian [17], but in the special case when the bare kernel is given by a decaying exponential—as is the case of the COVID-19 pandemic — the activity at time t depends only on the activity at time t-1. This feature makes our model amenable to simple analytic solutions that help shed light into the relevant mechanisms at play [16]. In a nutshell, the key to the simplicity of the model and its solution rests in the fact that for an exponential bare memory kernel, the contribution of both contagion and latency at time t can be expressed as a fraction of the previous activity at time t-1 [16]. This property is not shared by processes involving power-law bare memory kernels, where use of the Laplace transform is needed to solve a time-continuous version of Eq. 1 [13-15].

The discrete time-step process of Eq. 1 applied to the case at hand can be understood more clearly with the schematic diagram shown in Fig. 2. We consider an initial number λ_{-1} of "patients-zero" infectious at time t = -1 that are exogenously introduced to the population. At time t = 0, these individuals will follow one of three paths: they will either 1) be registered with probability α, 2) not be registered but continue being contagious with probability γ, or recover with probability $\rho = 1 - \alpha + \gamma$. The λ_{-1} patients-zero will also infect a number $\beta_0\lambda_{-1}$ of new individuals at time t = 0. In SIR-type models, the factor β is proportional to the so-called "susceptible" portion of the population that gets depleted as the epidemic advances. In the form displayed in Fig. 2, the model we propose here is valid only for the initial propagation of the pandemic since an effectively infinite source of susceptible individuals is considered, but it could be modified to incorporate a finite-sized population.

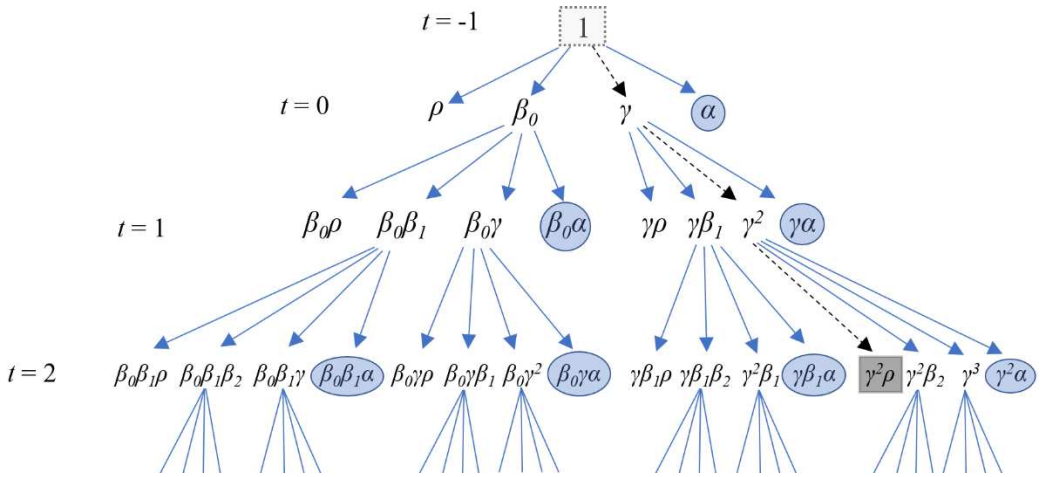

Fig. 2: Hawkes model with an exponential bare memory kernel. A normalized initial population of "patients-zero" on week -1 follows three possible outcomes on the following week given by the fractions: ρ (recovered), γ (still infectious but not registered) and α (registered and isolated). A new generation of infectious individuals is given by β_t times the population that originated it at time t. The sum of elements containing an α (elements inside circles) at a given time gives the total activity (Eq. 3), while the latent cases (Eq. 5) are the sum of the remaining terms after the recovered ones (elements with a ρ) have been subtracted. A tree is formed by replicating the four possible trajectories for each element that is still infectious at subsequent times. The dashed black arrows ending at the grey square depicts the trajectory followed by the portion of the patients-zero that were asymptomatic and recovered at time t = 2. Note these asymptomatic individuals infected others before recovering.

A second important assumption we make is that individuals that have been registered will either self-isolate or be hospitalized, but in either case, they will no longer be a source of new infections. In contrast, those that have not been registered yet can continue infecting others until they finally either recover or fall ill after which they will become registered. This contemplates the fact that asymptomatic individuals have been shown to be an important source of new COVID-19 infections [18] (see, for example, the dashed path in Fig. 2). Our model can be considered as a simplified version of other more complex epidemiological models that take into account social mobility to be able to relate the reproduction number with confinement measures [3]. The solution to Eq. 1 for the registered new cases in the absence of exogenous sources is

$$\lambda_t = \lambda_{-1}(1-\gamma-\rho)\prod_{k=0}^{t-1}(\gamma+\beta_k) = \lambda_{t-1}(\gamma+\beta_{t-1}). \quad (3)$$

The activities at times t and t+1 yield the relation:

$$\beta_t = (1-\gamma) + \frac{\lambda_{t+1}-\lambda_t}{\lambda_t} \quad (4)$$

From this last equation we can establish the condition for the contagion to either grow ($\beta_t+\gamma>1$) or die out exponentially ($\beta_t+\gamma<1$). Thus, we associate $(\beta_t+\gamma)$ with the effective reproduction number R_t at time t. While the collection of reproduction numbers $\{(\beta_0+\gamma),(\beta_1+\gamma),...,(\beta_t+\gamma)\}$ merely constitutes a parametrization of the activity in terms of exponential changes, from it we can construct the time series of the latent cases. From Fig. 2, these latent cases are given by

$$\lambda_{L,t} = \lambda_{-1}\prod_{k=0}^{t}(\gamma+\beta_k) = \lambda_{L,t-1}(\gamma+\beta_t). \quad (5)$$

Alternatively, we could have simply postulated Eq. 5 for the number of latent cases as a starting point and propose that at every time period, a fraction α of the previous latent cases is registered. However, the current exposition helps put our model in the general framework of the important class of Hawkes processes, and, in doing so, underscore the role of contagion in the renormalization of the latency function.

Knowing $\lambda_{L,t}$ allows us to calculate the number of cases that are still to be registered in the extreme optimistic scenario where one could assure no more contagion events will take place. This is readily calculated as $(1-A)\lambda_{L,t}$, where A is the fraction of asymptomatic individuals and is related to ρ (see Methods). This scenario is a particular case of $\beta_i = \beta$ for all i's. In this case, the activity λ_t can be written as

$$\lambda_t = \lambda_{-1}(1-\gamma-\rho)(\beta+\gamma)^t \propto e^{-\frac{t}{\tau}(1-\theta)}, \quad (6)$$

where $\theta \equiv \tau Ln\left(\beta e^{1/\tau}+1\right)$ and $1/\tau \equiv -Ln(\gamma)$. The form of this equation shows explicitly how contagion renormalizes the activity from a decaying exponential with constant $(1/\tau)$ in the absence of contagion (β=0), to one with constant $(1/\tau)(1-\theta)$ otherwise. Eqs. 3-6 can be used to lay out different scenarios, as we show next.

Results and discussion. – To apply our model to real data, we aggregate the registered, publicly available daily new cases of all the regions (or states) of a given country and during a whole week into the observed activity. This means that the

step-size of our model is equal to one week, thus smoothing the daily fluctuations observed in most countries. We are aware that in general the epidemic tends to evolve differently in each region of a given country since states may follow particular social distancing laws. Therefore, in its present formulation, ours can be considered as a mean field model that also ignores the topology of the social network and any mobility factors [18]. According to clinical studies, the average time from contagion to attending a hospital is ≈ 2 weeks [20] from which we obtain $\gamma = 0.5$, or, in terms of the decay constant, $(1/\tau) \equiv -Ln(\gamma)=0.693$ (see Methods). The inset of Fig. 1a shows that an exponentially decaying activity with this same constant is in apparent agreement with the dynamics observed for South-Korea from week #6 to week #10 (as counted from the week ending on Feb. 21st/2020), suggesting the intense cluster tracing, mass testing, and the imposition of quarantines on people travelling from abroad brought the number of new infections close to zero in this country during that period [21,22]. Similarly, we calculate ρ=0.1 after assuming a conservative estimate of 20% of asymptomatic cases [23] (see Methods). Other relevant studies that also take into account asymptomatic cases can be found in Refs. [24-25].

Fig. 1 shows the latent cases *vs.* t using Eq. 4 and the values for γ and ρ calculated above for four countries that have responded in very different ways to the pandemic. This figure shows how South-Korea (Fig. 1a) managed to stop a first wave of contagion by week #6 (April 3rd/2020), a fact reflected in the reproduction number being closer to γ. Consequently, the number of latent cases also decreased sharply reaching almost the same number of registered cases by week #8 (mid. April/2020). By then, our model shows there were practically no unregistered cases propagating the infection. In the case of Italy (Fig. 1b), where it took longer for the government to impose confinement measures, the number of latent cases reached almost 140,000 by week #4 (March 20th/2020), which amounts to close to 60% of the total 230,000 new cases this country has registered since then. It then took almost 10 weeks since the peak of activity was reached for these latent cases to decay, despite a strict lockdown being enforced around week #4. Using Eq. 5, we estimate the reproduction number in Italy in the period between weeks #9 and #15 to be $(\beta+\gamma) = 0.694 \pm 0.004$ (see inset of Fig. 1b). After reaching this stage, relaxing some social distance measures necessary to reopen their respective economies has been done with some confidence. Having just a few latent cases is why a second wave of infections has been successfully contained in South-Korea (from week #14, see inset of Fig. 1a). In contrast, for both the USA and Mexico, the number of latent cases as of July 23rd is enormous. Even if a strict nationwide lockdown were to be imposed and assuming the conservative estimate of 20% for the asymptomatic cases, we can still expect at least $\approx 0.115 \times 10^6 \times (1 - 0.2) \approx 92{,}000$ new cases in Mexico and $\approx 0.5 \times 10^6 \times (1 - 0.2) \approx 400{,}000$ new cases in the USA. A more realistic scenario for the USA in the case of a lockdown can be laid out by assuming a similar reproduction number as the one calculated in the case of Italy after it underwent lockdown.

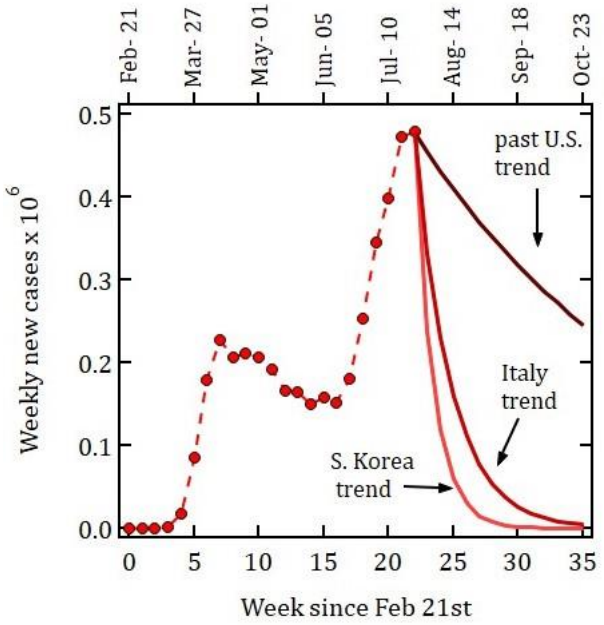

Fig. 3: Evolution of the number of weekly new cases in the USA after week #22 (July 24th) using Eq. 6 assuming the relaxation follows a) the same decay constant from the previous relaxation in the USA $(\gamma + \beta) = 0.95$ ("worst-case scenario"), b) the trend in Italy, $(\gamma + \beta) = 0.694$ and c) the trend in South-Korea $(\gamma + \beta) = 0.5$, ("best-case scenario"). In the best and worst-case scenarios, our model estimates, respectively, 400,000 and 9 million new cases from July 23rd on.

Using Eq. 6, the total sum of the activity after a peak λ_{max} during an infinite number of weeks is $\lambda_{max}\,(\beta + \gamma) / (1 - \beta - \gamma)$. After plugging the reproduction number for Italy during relaxation $((\beta+\gamma) = 0.694)$ and $\lambda_{max} \approx 480{,}000$, see last point in fig. 1c) in the last expression, the USA can still expect about 1.1 million new cases in the course of the following months. In a much grimmer (but not at all unrealistic) scenario, if the USA goes back to the slow decay in activity it experienced after reaching the initial peak on week #7 (mid-April, $(\beta + \gamma) \approx 0.95$, see top panel of Fig. 1c), then we can expect 9.12 million new COVID-19 cases. The number of cases to be expected in the worst and best-case scenarios explored here (see Fig. 3) differ by a factor of over 20, evidencing the enormous difference that distinct approaches to re-opening the economy and returning to some sort of normality can make [26]. This cannot be understated, specially considering that the death-toll is proportional to the number of cases. In this respect, we can estimate the number of new deaths from COVID-19 that may occur in the USA in the following months. The current mortality rate (for registered patients) is likely to be much smaller than its present time-average of 7% due to some of the following factors: 1) earlier infection identification, 2) better equipped hospitals, 3) younger average infected population (down from 65 to 35 years old), 4) and better knowledge about how to treat critically ill patients [27]. The current mortality rate can be roughly estimated to be about 1.36% by assuming a

one-week lag between new cases and new deaths (6426 new deaths on week #22 from 472,842 cases on week #21). This would imply 124,000 new COVID-19 deaths in the following months in the worst-case scenario presented above for the USA, bringing the total casualties to about 270,000 if no new treatment is developed that can reduce further the mortality rate.

The analysis we have presented illustrates that bending the epidemic curve and even having a downwards trend is meaningless in terms of public safety unless such trend lasts long enough as to allow for most of the latent cases to either recover or become symptomatic. Furthermore, given that the dynamics of Eq. 3 and 5 is Markovian, a sudden change in confinement practices is bound to have strong effects regardless of any apparent trend. It is worth pointing out that trends reflect social behavior and not any kind of inertia or "weakening" of the pathogen itself.

As stated in the introduction, our model is a special case of SIR-type models and is therefore not devoid of the same issues related to the sensitivity to the choice of parameters present in this class of models. However, we emphasize that our model constitutes a reparameterization of the weekly activity in terms of exponential changes. Consequently, it is particularly simple to implement and allows for the formulation of predictions by assuming previous reproduction numbers that can be calculated easily. Furthermore, it evidences the Markovian nature of this epidemic process.

The model we have proposed in this work is only valid for the initial stages of the epidemic. To account for the finite size of a given population, the β_k factors could be made proportional to the product of the susceptible and infected populations as is done in SIR-type models. Also, more precise scenarios than the ones we have laid out in this work could be obtained by atomizing our mean field model to the state or county level if, for example, knowledge of whether people are following social distancing measures is available.

The authors declare that there is no conflict of interest regarding the publication of this article.

***

We thank Raul Esquivel and Gabriel Caballero for reviewing the manuscript and acknowledge funding from DGAPA-PAPIIT project IT101820.

Methods. – *M.1 Data collection.* Daily new cases and new deaths were obtained from the website https://www.worldometers.info/coronavirus/. Daily counts were aggregated weekly starting from February 15th 2020, so that, for example, week #0 comprises the daily data from Feb 15th to Feb. 21st 2020.

In constructing our model, we have made the important assumption that the number of registered cases consists only of symptomatic individuals. But it is worth pointing out that this is not necessarily the case. For instance, during the early stages of the pandemic in a given country, the number of registered cases likely consisted of the number of individuals that having developed symptoms were be admitted into a hospital. As the number of tests performed increased, it is almost certain that the registered cases included also asymptomatic individuals. The validity of our results hinges on the assumption of a somewhat stable percentage of positive tests. In the USA, this percentage stabilized to around 20% from mid-February to the end of April (see https://covidtracking.com/data/us-daily), the period over which the reproduction number was calculated to predict the dynamics of future relaxations for this country. For our prediction to be valid, this percentage of positive tests must remain somewhat constant in the following months. This important caveat needs be kept in mind when considering the validity of our results.

M.2 Estimation of γ and ρ from clinical trials. First, consider the evolution of the patients-zero in the absence of contagion. The probability of being registered after a number t of weeks has elapsed since this population was introduced, is given by $\lambda\left(t|\beta_k=0\right)=\alpha\gamma^t=(1-\gamma-\rho)\gamma^t$. From this equation, the percentage of asymptomatic cases is $A\equiv 1-\sum_{t=0}^{\infty}\left(1-\gamma-\rho\right)\gamma^t=\rho/(1-\gamma)$. Then, if the possibility of having asymptomatic cases is precluded (as in the case of clinical tests with infected patients), the naturally normalized bare memory kernel is $\varphi_t\equiv\lambda\left(t|\beta_k=0,\rho=0\right)$, or $\varphi_t=(1-\gamma)\gamma^t$. Then, the average time elapsed from contagion to being registered is then $T\equiv\sum_{t=0}^{\infty}\varphi_t=\gamma/\left(1-\gamma\right)$. Independent clinical studies on isolated infected patients have found that $T=13$ days $\simeq 2$ weeks [20], which contemplates the fact that, on average, it takes people 2 days after developing dyspnea to go to a hospital. Since in our model $t=0$ is the first week after exposure (see Fig. 2), this period corresponds to $T\simeq 1$, yielding $\gamma=0.5$ and $(1/\tau)\equiv$-$Ln(\gamma)$=0.6931 Regarding the rate of asymptomatic cases, this is a topic of current debate [28,29]. But investigations on isolated groups of people over long periods of time have found $A\simeq$20% to be a conservative estimate [23]. We use this value in the present study, but other scenarios could be easily explored by modifying it. Along with the estimation for γ made above, this implies ρ=0.1. Using clinical data to fix γ and ρ as we have just done implies that the only free parameter of the model is the number of patients-zero, λ_{-1}. The simplicity of our model carries with it another important condition that must be consistent with independent clinical results for COVID-19. Given that the γ and ρ branches are symmetric in the diagram of Fig. 2, the latency function for new registered cases is identical to the latency function for recovered ones. Consequently, the average time from infection to recovery (T_ρ) must equal T. A recent study [15] has found that the probability of being infectious at time t is given by a gamma distribution function with an average close to 2 weeks. While other studies have found that the period during which virus shedding lasts is longer than this value, its magnitude is not enough for contagion to continue [30]. Once again, this average of two weeks corresponds to T_ρ=1 in our model, which

equals T, as required. Note that we are requiring these average values to coincide and not the full probability functions themselves, which would be a much stronger condition. The only distribution function we do require to equal the observed data is the activity λ_t during the relaxation period with constant β (Eq. 5, insets of Figs. 1a and 1b). Here we have made the simplifying assumption that symptomatic and asymptomatic clinical features are similar, even though some differences have been found for the characteristics of viral shedding in these populations [31]. We emphasize the fact that the condition $T = T_\rho$ is a consequence of the simplicity of our model that may not be satisfied by other infectious diseases.

M.3 Exponential form of the activity. The binomial expression for the activity of Eq. 5 can be recast in exponential form as follows. Starting from Eq. 5, the term involving the time evolution is:

$$(\beta+\gamma)^t = e^{t\ Ln(\beta+\gamma)}\ , \qquad \text{(M1)}$$

Factorizing γ inside the logarithm, and in turn factorizing $Ln(\gamma)$, one arrives at the following expression:

$$e^{t\ Ln(\gamma)\{1+Ln(1+\beta/\gamma)/Ln(\gamma)\}} = e^{-t/\tau\ \left\{1-\tau Ln(1+\beta e^{1/\tau})\right\}}. \qquad \text{(M2)}$$

In the last step we defined the decaying constant $1/\tau \equiv -Ln(\gamma)$. Finally, by defining $\theta \equiv \tau Ln\left(\beta e^{1/\tau}+1\right)$, we arrive at

$$e^{-t/\tau\ \{1-\theta\}}. \qquad \text{(M3)}$$

REFERENCES

[1] HELBING D., et al. Saving human lives: What complexity science and information systems can contribute. *Journal of statistical physics,* **158**, 3 (2015): 735-781. https://doi.org/10.1007/s10955-014-1024-9

[2] KESKINOCAK P., et al. The Impact of Social Distancing on COVID19 Spread: State of Georgia Case Study. *medRxiv*, (2020).; https://doi.org/10.1101/2020.04.29.20084764

[3] ARENAS A., et al. Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. *medRxiv*, (2020). https://doi.org/10.1101/2020.04.06.20054320

[4] CHINAZZI M., et al. Science **368**, 395 (2020). https://doi.org/10.1126/science.aba9757

[5] PREM, K., et al. "The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study." *The Lancet Public Health* (2020). https://doi.org/10.1016/S2468-2667(20)30073-6

[6] KRAMER D. Physics Today **73**, 25 (2020). https://doi.org/10.1063/PT.3.4493

[7] KERMACK W.O. and MCKENDRICK A. G., Journal of Hygiene **39**, 271 (1939). https://doi.org/10.1098/rspa.1927.0118

[8] CASTRO M., et al. "Predictability: Can the turning point and end of an expanding epidemic be precisely forecast?." arXiv preprint arXiv:2004.08842 (2020).

[9] TIAN H. Y. et al., Science **368**, 638 (2020). https://doi.org/10.1126/science.abb6105

[10] PERC M., et al. Forecasting Covid-19. *Front. Phys.* **8**, 127 (2020) https://doi.org/10.3389/fphy.2020.00127

[11] HAWKES A. G., Biometrika **58**, 83 (1971). https://doi.org/10.1093/biomet/58.1.83

[12] VAZQUEZ A., et al, Physical Review E **73**, 036127 (2006). https://doi.org/10.1103/PhysRevE.73.036127

[13] CRANE R. and SORNETTE D., Proceedings of the National Academy of Sciences of the United States of America **105**, 15649 (2008). https://doi.org/10.1073/pnas.0803685105

[14] SORNETTE D., DESCHATRES F., GILBERT T., and AGEON Y., Physical Review Letters **93**, 228701 (2004). https://doi.org/10.1103/PhysRevLett.93.228701

[15] HELMSTETTER A. and SORNETTE D., Journal of Geophysical Research-Solid Earth **107**, 2237 (2002). https://doi.org/10.1029/2001JB001580

[16] ESCOBAR J. V. and SORNETTE D, Plos One **10**, e0116811 (2015). https://doi.org/10.1371/journal.pone.0116811

[17] SAICHEV A. I. and SORNETTE D., *European Physical Journal* B **75**, 343 (2010). https://doi.org/10.1140/epjb/e2010-00121-7

[18] LONG Q. X. *et al.*, Nature Medicine. https://doi.org/10.1038/s41591-020-0869-5

[19]GOMEZ S., ARENAS A., BORGE-HOLTHOEFER J., MELONI, S. and MORENO Y., Epl **89**, 38009 (2010). https://doi.org/10.1209/0295-5075/89/38009

[20] WANG D. W. *et al.*, *Jama-Journal of the American Medical Association* **323**, 1061 (2020). https://doi.org/doi:10.1001/jama.2020.1585

[21] SHAW R. KIM Y. and HUA J. *Progress in disaster science* **6** (2020): 100090. https://doi.org/10.1016/j.pdisas.2020.100090

[22] KUPFERSCHMIDT K., Science **368**, 808 (2020). https://doi.org/10.1126/science.368.6493.808

[23] MIZUMOTO K., et al, Eurosurveillance **25**, 2, 2000180 (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000180

[24] DE LA SEN M., et al. " On a new epidemic model with asymptomatic and dead-infective subpopulations with feedback controls useful for Ebola disease." Discrete dynamics in Nature and society 2017, 4232971 (2017). https://doi.org/10.1155/2017/4232971

[25] GAETA, G. "A simple SIR model with a large set of asymptomatic infectives." *arXiv preprint arXiv:2003.08720* (2020). https://doi.org/10.3934/mine.2021013
[26] KUPFERSCHMIDT, K. "Ending coronavirus lockdowns will be a dangerous process of trial and error." *Science| AAAS* (2020). https://doi.org/doi:10.1126/science.abc2507
[27] HORBY P., et al. "Effect of Dexamethasone in Hospitalized Patients with COVID-19: Preliminary Report." *medRxiv* (2020). https://doi.org/10.1101/2020.06.22.20137273
[28] DAY M., Bmj-British Medical Journal **368**, m1165 (2020). https://doi.org/10.1136/bmj.m1165
[29] PARK M. et al, Journal of Clinical Medicine **9**, 967 (2020). https://pubmed.ncbi.nlm.nih.gov/32244365/
[30] WOLFEL R. *et al.*, Nature **581** 7809 (2020). https://doi.org/10.1038/s41586-020-2196-x
[31] LONG Q. X. *et al.*, Nature Medicine Jun 18:1-5 (2020). https://doi.org/10.1038/s41591-020-0965-6

Figures

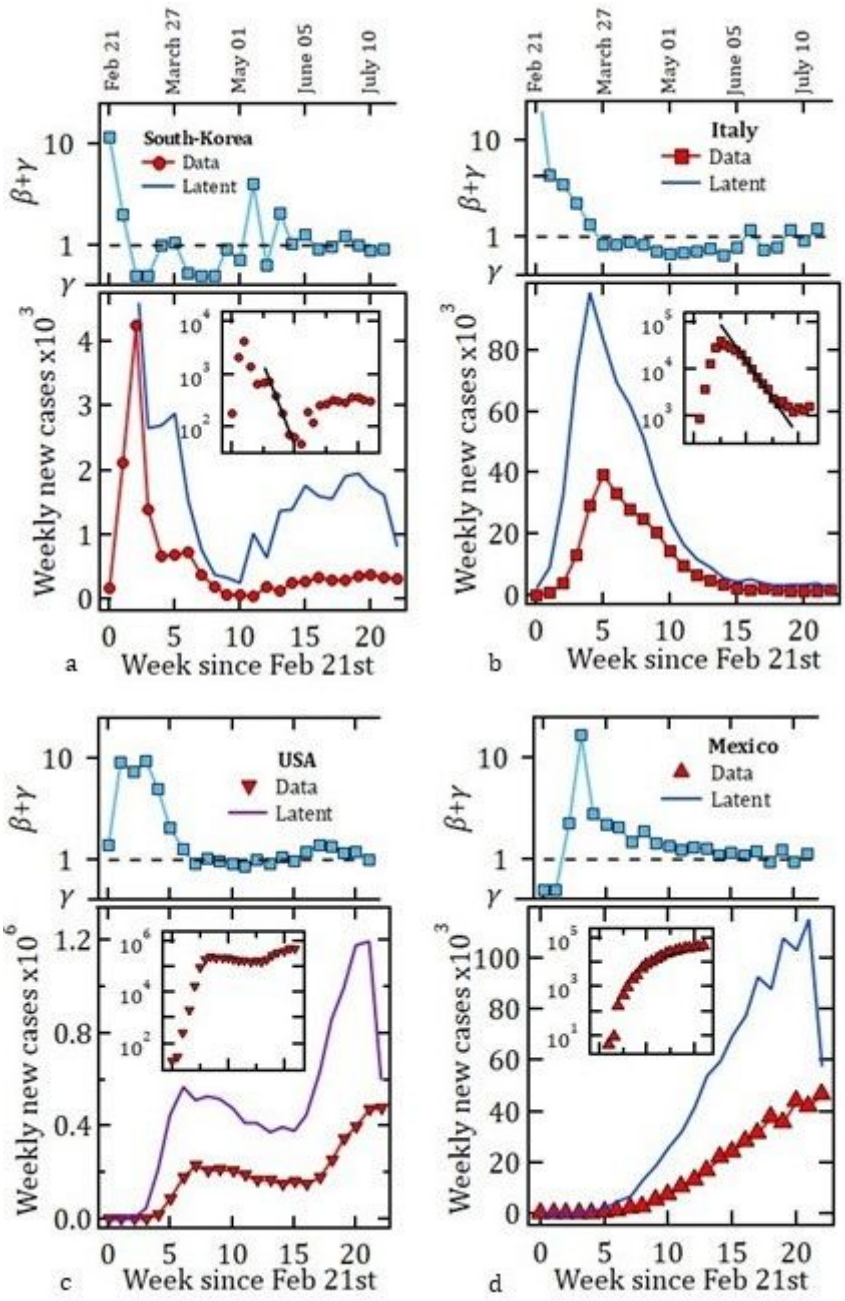

Figure 1

Weekly registered new cases (dots) and calculated latent ones (Eq. 3), along with the corresponding reproduction numbers Ri=βi+γ (Eq. 4) for (a) South Korea, (b) Italy, (c) USA and (d) Mexico. See main text for details. Insets show the weekly new cases in log-linear scale. The solid black line in (a) depicts an exponentially decaying activity (Eq. 5 with β=0) setting γ = 0.5, as calculated from clinical studies (see Methods). This apparent agreement suggests contagion events were successfully brought to a halt in S. Korea during that relaxation period. In the inset of Fig. 1b), the solid line is the best fit to Eq. 5 (β=0.194 ±.004) assuming γ =0.5 as before.

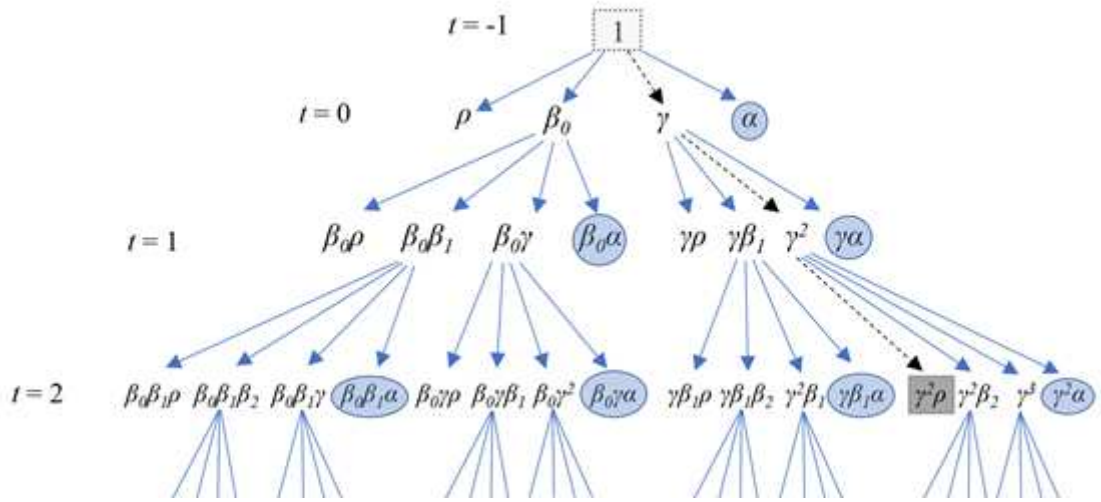

Figure 2

Hawkes model with an exponential bare memory kernel. A normalized initial population of “patients-zero” on week -1 follows three possible outcomes on the following week given by the fractions: ρ (recovered), γ (still infectious but not registered) and α (registered and isolated). A new generation of infectious individuals is given by βt times the population that originated it at time t. The sum of elements containing an α (elements inside circles) at a given time gives the total activity (Eq. 3), while the latent cases (Eq. 5) are the sum of the remaining terms after the recovered ones (elements with a ρ) have been subtracted. A tree is formed by replicating the four possible trajectories for each element that is still infectious at subsequent times. The dashed black arrows ending at the grey square depicts the trajectory followed by the portion of the patients-zero that were asymptomatic and recovered at time t = 2. Note these asymptomatic individuals infected others before recovering.

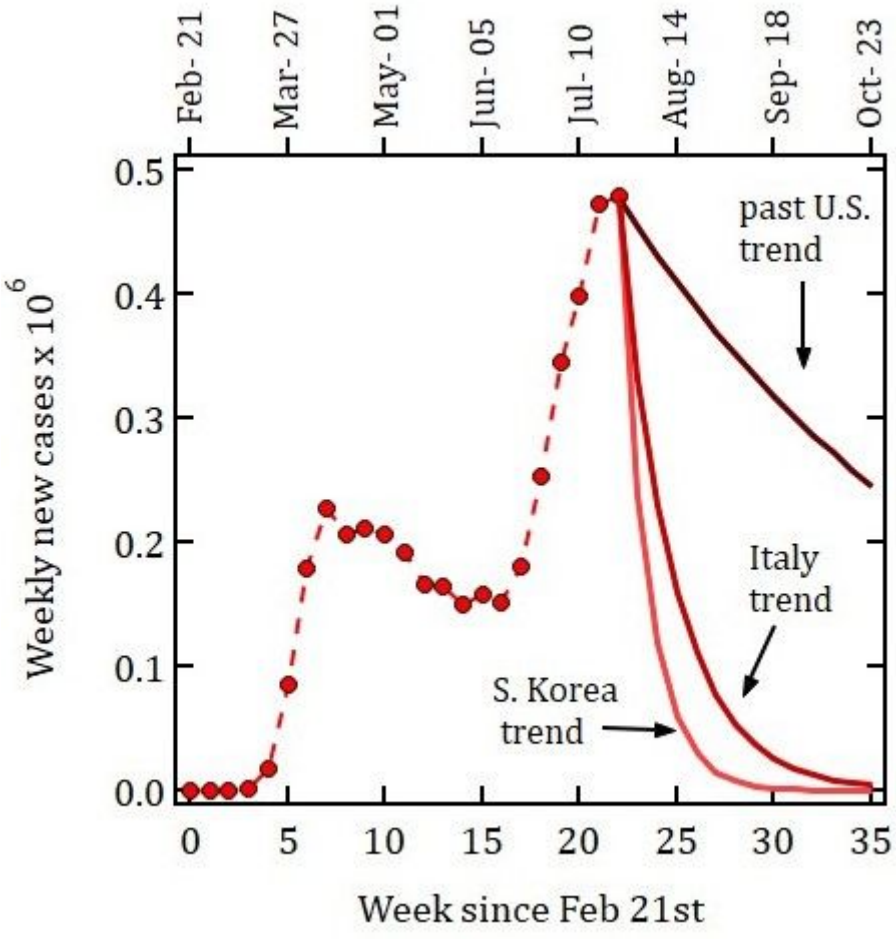

Figure 3

Evolution of the number of weekly new cases in the USA after week #22 (July 24th) using Eq. 6 assuming the relaxation follows a) the same decay constant from the previous relaxation in the USA ($\gamma + \beta$) = 0.95 (“worst-case scenario”), b) the trend in Italy, ($\gamma + \beta$) = 0.694 and c) the trend in South-Korea ($\gamma + \beta$) = 0.5, (“best-case scenario”). In the best and worst-case scenarios, our model estimates, respectively, 400,000 and 9 million new cases from July 23rd on.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementalMaterialHawkesCOVID19Escobar.pdf