Evolution of COVID-19 pandemic: Power law growth and saturation

Soumyadeep Chatterjee,[1,*] B. Shayak,[2,†] Ali Asad,[1,‡] Shashwat Bhattacharya,[3,§] Shadab Alam,[3,¶] and Mahendra K. Verma[1,**]

[1] *Department of Physics, Indian Institute of Technology Kanpur, Kanpur 208016, India*
[2] *Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, USA*
[3] *Department of Mechanical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India*

(Dated: May 5, 2020)

In this paper, we analyze the real-time infection data of COVID-19 epidemic for 21 nations up to April 20, 2020. We observe that China, South Korea, Australia, Hong Kong, and Switzerland have flattened their infection curves. For these nations, the total number of infected individuals ($I(t)$) exhibits a succession of exponential growth and power-law growth ($t^2, t, \sqrt{t}$ in sequence) before flattening of the curve. USA, Italy, UK, France, Spain, Germany, Belgium, Israel, Netherlands, India, and Sri Lanka have reached up to the linear growth ($I(t) \sim t$), but they are yet to flatten their curves. Russia and Singapore are still in the exponential growth regime. Such features of $I(t)$ curves could be used for understanding and forecast of the epidemic evolution. Besides these detailed analysis, we compare the predictions of an extended SEIR model and a delay differential equation-based model with the reported infection data, and observed a general agreement between them.

I. INTRODUCTION

As of May 1, 2020, COVID-19 pandemic has infected more than 3.3 million of human population, and caused 0.23 million deaths. The world economy is in tatters. Therefore, cure as well as understanding of the pandemic are extremely crucial. In the present paper, we analyze the publicly-available national COVID-19 infection data [1] up to 20th April, 2020. We observe that the COVID-19 infection curves for many nations exhibit power-law growth after an exponential growth. We compare the reported data with model predictions and observe a good agreement between them.

To undestand and forecast, epidemiologists have made many models [2–4]. One of the first models is called SIR model, where the variables S and I describe respectively the numbers of susceptible and infected individuals. The third variable R represents the removed individuals who have either recovered or died. An advanced model, called SEIR model, includes *exposed* individuals, E, who are infected but not yet infectious [3, 4].

SARS-CoV-2 is a flu virus. However, it extremely dangerous because of its highly infectious nature and its adverse effects on the lungs. *Asymptomatic carriers*, individuals who do not exhibit any symptoms for one to two weeks, have carried the virus to far off places where it has spread rapidly [5]. To stop the spread of the deadly virus, various nations have employed lockdowns, mandatory social distancing, quarantines for the affected, etc. These features make it difficult to make models for COVID-19 epidemic.

In spite of the difficulties stated above, there are many models that are able to describe the COVID-19 pandemic data quite well. Peng et al. [6] constructed a generalized SEIR model with seven-variables (including quarantined and death variables) for epidemic spread in China. Their predictions are in good agreement with the present data. López and Rodo [7] formulated an extended version of this model to analyse the spread of the pandemic in Spain and Italy. Hellewell et al. [8] studied the effects of isolation on controlling the COVID-19 epidemic. Chinazzi et al. [9] analyzed the effects of travel restrictions on the spread of COVID-19 in China and in the world using the global metapopulation disease transmission model. Mandal et al. [10] constructed model for devising intervention strategies in India. Using an age-structured SIR model, Singh and Adhikari [11] studied the impact of social distancing on the COVID-19 epidemic in India. Shayak et al. [12] and Shayak and Rand [13] have constructed delay differential equation (DDE) model for the spread of COVID-19; these models takes into account the pre-symptomatic period and the effects of testing delays. They have also constructed a model to predict the end of the epidemic.

Due to the above complex issues in the epidemic models of COVID-19, many researchers have consciously focussed on the data and attempted to extract useful information from them. It has been observed that the analysis of the pandemic provides important clues that may be useful for its forecast. In particular, Ziff and Ziff [14], Komarova and Wodarz [15], Manchein et al. [16], Blasius [17], Marsland and Mehta [18], Li et al. [19], Singer [20], Beare and Toda [21], and Cherednik and Hill [22] analyzed the reported count of total infections ($I(t)$) in various nations and observed power-law growth after the exponential regime. Verma et al. [23] analyzed the data of 9 nations up to April 7 and showed that the $I(t)$ goes through power laws, t^3, t^2, t and $\sqrt{t}$,

* soumyade@iitk.ac.in
† shayak.2015@iitkalumni.org
‡ aliasad@iitk.ac.in
§ shabhatt@iitk.ac.in
¶ shadab@iitk.ac.in
** mkv@iitk.ac.in

in temporal sequence before flattening out. By April 7, China and South Korea had flattened the $I(t)$ curves, but other nations were either in the exponential regime or in power-law regime. However, by April 20, more nations are close to the flattening of the curve. Besides the above results, Prakash et al. [24] reported a linear growth of $I(t)$ after early exponential growth.

Schüttler et al. [25] analyzed the daily death counts for various nations and observed that their probability distributions appear to follow Gauss profile. They argue that the Gauss model could be useful for the predictions of the duration and extent of death cases. Marsland and Mehta [18] observed that the error function provides best fit to the total count, $I(t)$; this observation follows from Schüttler et al. [25]'s analysis.

There are epidemic growth models based on population growth [4, 26]. Since COVID-19 spreads via asymptomatic carriers, the epidemic forms a network. Hence, network-based epidemic growth models may be useful for modelling COVID-19 pandemic. The results of Min et al. [27], who studied the epidemic spread within a network with different mixing styles, may be useful for such modelling. They observed that the average epidemic size and speed depend critically on network parameters.

In this paper, we analyzed the COVID-19 infection data up to April 20, 2020 for 21 nations and observed that all the nations are following transition from exponential to power law growth in infection counts. Many of the 21 nations are close to flattening their curves, while some of them (for example, Russia) appear to be in the exponential regime. We also showed that three epidemics—Ebola, COVID-19, MERS—have similar evolution: exponential and power law growth, and then flattening of the curve. In addition, we compared the predictions of an extended SEIR model [7] and a delay-differential equation model [13] with the real-time data and observed good agreement among them.

The structure of the paper is as follows: in Sec. II we analyse the COVID-19 data for 21 leading nations and observe power law growth for them after the exponential growth. The evolutions of Ebola, MERS, and COVID-19 are compared in Sec. III. The predictions of two models of COVID-19 pandemic are compared with the observed data in Sections IV and V. We conclude in Sec. VI.

II. DATA ANALYSIS OF COVID-19 EPIDEMIC

In this section, we perform a comprehensive data analysis of COVID-19 cases for 21 countries listed in Table I up to April 20, 2020. The majority of the countries in our analysis include those with a large number of COVID-19 cases, such as USA, Italy, Germany, China, and India among others. For a more complete study, we also include countries with a relatively smaller number of cases such as Sri Lanka and Singapore. We used the real-time data available at *worldOmeter* [1] and chose the starting date (see Table I) as the one from which the number of infected cases increased rapidly. Johns Hopkins University, Corona Resource Center [28] too is an important repository for COVID-19 data.

We analyze the evolution of cumulative number of infected cases, which is denoted by $I(t)$, with time in days, denoted by t. For all $I(t)$, we compute the derivatives $\dot{I}(t)$ using Python's *gradient* function. These derivatives indicate the daily count of infected cases. Note that $\dot{I}(t)$ exhibit lower fluctuations than the measured daily counts due to smoothing. We plot $I(t)$ (red curves) and $\dot{I}(t)$ (blue curves) in *semi-logy* format for all the 21 countries in Fig. 1.

We find that a single function does not describe the $I(t)$ curves; hence, we compute best-fit curves for different parts of $I(t)$ by employing exponential and polynomial functions. We used Python's *polyfit* function to compute the best fit curves. These curves are listed in Table I along with the relative error between the original data and the fitted data. Note, however, we exhibit only the leading power laws of the polynomials in the plots.

Initially, all the countries exhibit exponential growth ($I(t) = A\exp(\beta t)$), which is expected. It is worth mentioning that the $I(t)$ plots for USA, UK, France, Spain, Germany, Russia, Belgium, and Brazil have two exponential functions for the fit. As an example, the $I(t)$ curve of UK is described by two exponential functions, $\sim \exp(0.26t)$ and $\sim \exp(0.23t)$. The quantity β is proportional to the growth rate. The value of β varies for different countries as it depends on various factors such as population density, immunity level of the population, climate, local policy decisions (social distancing, lockdowns, testing capacity), etc.

In the exponential regime, the daily infection count is directly proportional to the cumulative count of cases, that is, $\dot{I} \approx \beta I$. The cumulative case count doubles in time $T = (\log 2)/\beta$ in this regime, For Italy, $\beta = 0.33$, resulting in $T \approx 2$ days, which means Italy's $I(t)$ doubled every two days in the early phase (February 22 to March 01).

Next, the curves transition to the regime that is best described by polynomials and can be approximated as power laws. In Fig. 1, we report the leading terms of the best-fit polynomials as the power law of the given region (see Table I). The $I(t)$ curves for South Korea, China and Switzerland exhibit three power law regimes—t^2, t, and $\sqrt{t}$—before flattening. Similarly, $I(t)$ for Australia and Hong Kong saturate after t and $\sqrt{t}$ regimes. Note that Russia and Singapore are yet to transition to a power law regime. As predicted in our earlier work [29], the countries such as USA, France, Italy, Spain, and Germany transition to a linear regime after going through regimes of t^4, or t^3, or t^2. These nations are close to flattening their $I(t)$ curves. UK, Turkey, Israel and Netherlands exhibit similar transitions. However, $I(t)$ for countries like India, Hong Kong, Sri Lanka, Australia, and Belgium directly transitioned to the linear regime from an exponential phase. We make a cautionary remark that

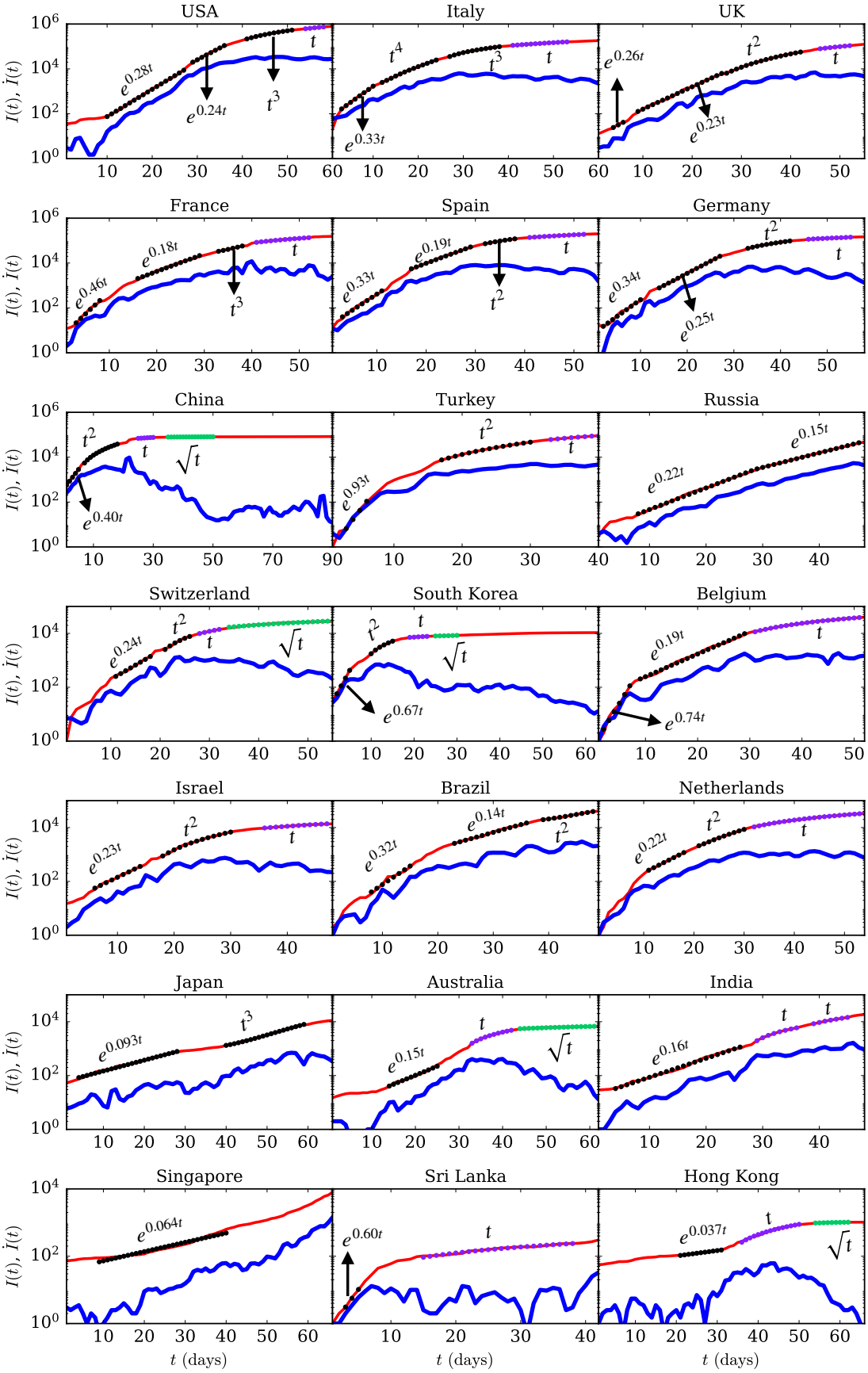

FIG. 1. (color online) For the COVID-19 epidemic, the *semi-logy* plots of total infected individuals ($I(t)$) vs. time (t) (red curves) for 21 countries. We also plot $\dot{I}(t)$ vs. t (blue curves). The dotted curves represent the best-fit curves, which are exponential or polynomial functions.

TABLE I. The best-fit curves along with their relative errors for the COVID-19 data for 21 countries.. The best-fit curves are shown in Figs. 1. The start date for the plots of Fig. 1 are listed below the names of the countries.

Country (Start Date)	Best-fit functions and errors	Country (Start Date)	Best-fit functions and errors
USA (February 21)	1) $4.6e^{0.28t}$ ($\pm 4.1\%$) 2) $17e^{0.24t}$ ($\pm 3.5\%$) 3) $8.5t^3 - 970t^2 + 67000t - 15 \times 10^5$ ($\pm 0.46\%$) 4) $32000t - 11 \times 10^5$ ($\pm 0.11\%$)	Italy (February 21)	1) $60e^{0.33t}$ ($\pm 3.1\%$) 2) $0.51t^4 - 30t^3 + 770t^2 - 8500t - 35000$ ($\pm 1.5\%$) 3) $7.8t^3 - 760t^2 + 30000t - 38 \times 10^4$ ($\pm 0.61\%$) 4) $4000t - 54000$ ($\pm 0.41\%$)
UK (February 26)	1) $8.5e^{0.26t}$ ($\pm 8.1\%$) 2) $16e^{0.23t}$ ($\pm 5.1\%$) 3) $130t^2 - 5400t + 60000$ ($\pm 1.8\%$) 4) $4900t - 15 \times 10^4$ ($\pm 0.24\%$)	France (February 24)	1) $5.4e^{0.46t}$ ($\pm 11\%$) 2) $110e^{0.18t}$ ($\pm 6.5\%$) 3) $6.9t^3 - 360t^2 + 4200t - 40000$ ($\pm 1.6\%$) 4) $5400t - 13 \times 10^4$ ($\pm 0.24\%$)
Spain (February 26)	1) $15e^{0.33t}$ ($\pm 8.1\%$) 2) $230e^{0.19t}$ ($\pm 3.6\%$) 3) $27t^2 + 5900t - 14 \times 10^4$ ($\pm 1.8\%$) 4) $4700t - 54000$ ($\pm 0.24\%$)	Germany (23 February)	1) $7.6e^{0.34t}$ ($\pm 11\%$) 2) $24e^{0.25t}$ ($\pm 2.9\%$) 3) $22t^2 + 4100t - 110000$ ($\pm 1.2\%$) 4) $2900t - 16000$ ($\pm 0.24\%$)
China (January 22)	1) $380e^{0.40t}$ ($\pm 2.2\%$) 2) $89t^2 + 750t - 4100$ ($\pm 2.8\%$) 3) $1400t - 34000$ ($\pm 0.64\%$) 4) $2200\sqrt{t} + 65000$ ($\pm 0.085\%$)	Turkey (March 12)	1) $0.40e^{0.93t}$ ($\pm 9.1\%$) 2) $110t^2 - 2300t + 14000$ ($\pm 0.85\%$) 3) $4300t - 80000$ ($\pm 0.30\%$)
Russia (March 04)	1) $5.1e^{0.22t}$ ($\pm 5.8\%$) 2) $40e^{0.15t}$ ($\pm 1.9\%$)	Switzerland (February 26)	1) $17e^{0.24t}$ ($\pm 4.2\%$) 2) $31t^2 - 440t - 2200$ ($\pm 5.6\%$) 3) $1000t - 19000$ ($\pm 0.47\%$) 4) $7700\sqrt{t} - 28000$ ($\pm 2.3\%$)
South Korea (February 18)	1) $15e^{0.67t}$ ($\pm 2.2\%$) 2) $9.7t^2 + 430t - 3500$ ($\pm 1.5\%$) 3) $160t - 4000$ ($\pm 0.67\%$) 4) $880\sqrt{t} + 3600$ ($\pm 0.085\%$)	Belgium (February 29)	1) $0.63e^{0.74t}$ ($\pm 18\%$) 2) $36e^{0.19t}$ ($\pm 7.6\%$) 3) $1400t - 31000$ ($\pm 1.0\%$)
Israel (March 04)	1) $14e^{0.23t}$ ($\pm 6.0\%$) 2) $19t^2 - 440t - 2500$ ($\pm 2.5\%$) 3) $380t - 3900$ ($\pm 0.67\%$)	Brazil (March 03)	1) $3.2e^{0.32t}$ ($\pm 11\%$) 2) $120e^{0.14t}$ ($\pm 3.6\%$) 3) $120t^2 - 8500t - 16000$ ($\pm 1.4\%$)
Netherlands (February 27)	1) $24e^{0.22t}$ ($\pm 2.3\%$) 2) $41t^2 - 1400t + 13000$ ($\pm 1.7\%$) 3) $1050t - 23000$ ($\pm 0.85\%$)	Japan (February 15)	1) $58e^{0.093t}$ ($\pm 4.6\%$) 2) $0.38t^3 - 39t^2 + 1400t - 16000$ ($\pm 0.46\%$)
Australia (February 19)	1) $4.7e^{0.15t}$ ($\pm 6.1\%$) 2) $360t - 10000$ ($\pm 1.3\%$) 3) $1000\sqrt{t} - 1400$ ($\pm 1.4\%$)	India (March 04)	1) $17e^{0.16t}$ ($\pm 7.1\%$) 2) $560t - 14000$ ($\pm 0.86\%$) 3) $1000t - 31000$ ($\pm 0.71\%$)
Singapore (February 15)	1) $38e^{0.064t}$($\pm 12\%$)	Sri Lanka (March 10)	1) $0.51e^{0.60t}$ ($\pm 4.0\%$) 2) $6.3t - 1.0$ ($\pm 3.1\%$)
Hong Kong (February 15)	1) $49e^{0.037t}$ ($\pm 1.5\%$) 2) $45t - 10000$ ($\pm 3.0\%$) 3) $100\sqrt{t} + 200$ ($\pm 0.55\%$)		

the coefficients of the polynomials depend quite critically on the choice of end points of the fit. Our observations of power-law growth are consistent with earlier results [14–21, 23].

In Fig. 1, in the exponential regime, the $\dot{I}(t)$ curves (daily counts) are nearly parallel to the $I(t)$ curves. It means $\dot{I}$ increases exponentially in the beginning, similar to $I(t)$. Subsequently, the curves transition to power-law regimes. As discussed by Verma et al. [23], the power law can be represented as $I(t) \sim At^n$ and $\dot{I}(t) \sim I^{1-1/n}$, which is slower than $\dot{I}$ for the exponential regime. We also remark that for large n, $\dot{I}(t) \sim I$, similar to exponential function.

We estimate the doubling time using the best-fit functions. For example, for India, $I(t) \sim 1000t - 31000$ from the $t_0 = 38$th day onward. Therefore, the doubling time T satisfies the relation: $I(t_0 + T) = 2I(t_0)$, which yields

$$T \approx t_0 - 31 \approx 7 \text{ days} \tag{1}$$

for this phase. The linear growth regime has another

TABLE II. The best-fit curves along with their relative errors for the COVID-19 data for the world. The best-fit curves are shown in 2. Part 1 is from January 22 to March 01, while Part 2 is from March 02 to April 20; they correspond to Figures 2(a) and 2(b) respectively. For part 2, $\bar{t} = t - 40$ and $\bar{I}(\bar{t}) = I(t) - I(t_0) = I(t) - 88000$, that is, time is measured from March 02.

Part (Start Date)	Best-fit functions and errors
Part 1 (January 22)	1) $390e^{0.40t}$ ($\pm 2.3\%$) 2) $90t^2 + 750t - 4100$ ($\pm 2.7\%$) 3) $1800t + 25000$ ($\pm 0.89\%$) 4) $13000\sqrt{t} + 6500$ ($\pm 0.29\%$)
Part 2 (March 02)	1) $4000e^{0.23\bar{t}}$ ($\pm 6.9\%$) 2) $1700\bar{t}^2 - 29000\bar{t} + 95000$ ($\pm 0.83\%$) 3) $79000\bar{t} - 16 \times 10^5$ ($\pm 0.40\%$)

interesting property. In this regime, $\dot{I} \approx$ const., that is, constant daily infection count. The daily infection count starts to decrease after the linear regime, hence linear regime is the transition point.

We also analyze the data of cumulative infected individuals in the entire world. In Fig. 2, we plot $I(t)$ and $\dot{I}(t)$ versus time in *semi-logy* format. Note that the initial epicenter of the COVID-19 outbreak was in China which later shifted to Europe and then to USA. Therefore, we divide the plot in two parts. In the first part [Fig. 2(a)], we illustrate cases that belong mostly to China. After approximately thirty days of outbreak (around February 20), $I(t)$ for China starts to saturate. In Fig. 2(b), we exhibit the $\bar{I}(\bar{t})$ curve after $t = 41$ (March 02) when China had achieved flattening of the curve. In Fig. 2(b), $\bar{t} = t - 40$ and $\bar{I}(\bar{t}) = I(t) - I(t_0) = I(t) - 88000$ due to the coordinate shifts. Both the plots exhibit exponential and power law regimes, but $\bar{I}(\bar{t})$ is yet to flatten. We hope that there is no third part to this curve, which is possible if the unaffected countries remain so.

The transition from the exponential to power law behaviour is expected from the nature of the $I(t)$ curve. The $I(t)$ curve is convex during the exponential growth phase, that is, its center of curvature is upward. However, the curve must turn concave for it to flatten. This transformation occurs via a sequence of growth phases: power-law, linear, square-root, and then flat. In Sec. VI we argue that the power law behaviour is possibly due to lockdown and social distancing.

We remark that the death count due to COVID-19 also exhibits similar behaviour as the infection count $I(t)$. It is expected because a fraction of infected individuals unfortunately die. However, we expect a small time delay difference between the death time series and the infection time series. Some researchers have attempted to fit the $I(t)$ and death counts with error functions [18, 25].

The above analysis shows that we can track the development of the epidemic locally in time. The best-fit

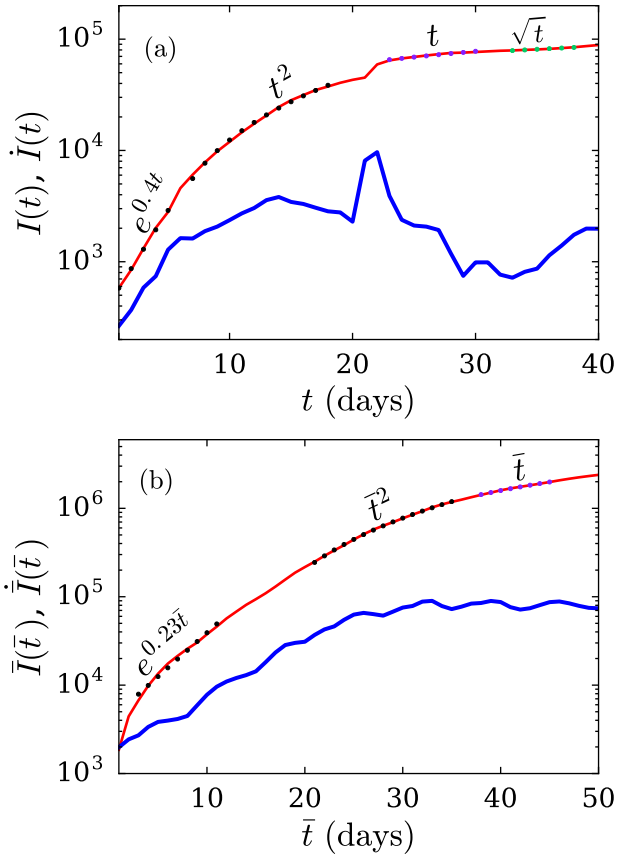

FIG. 2. (color online) For the COVID-19 epidemic, the *semi-logy* plots of total infected individuals ($I(t)$) vs. time (t) (red curves) for (a) Part 1: from January 22 to March 01 (comprising mostly of China), (b) Part 2: from March 02 to April 20 (world other than China). In Part 2, we subtract the counts of China, i.e., $\bar{t} = t - 40$ and $\bar{I}(\bar{t}) = I(t) - I(t_0) = I(t) - 88000$. The blue curves in the plot describe the derivatives of $I(t)$. The dotted curves represent the best-fit curves.

curves in particular segments provide the status of the epidemic. For example, if we have reached the linear regime, then we are not far from flattening the curve. Thus, simple data analytics described above has significant predictive power.

In the next section we compare the functional behaviour of COVID-19's $I(t)$ curve with those of other major epidemics.

III. COMPARISON WITH OTHER EPIDEMICS

A natural question is whether the epidemic spread of COVID-19 differs from the spread of Ebola and MERS (Middle Eastern Respiratory Syndrome). In this section we perform a comparative study of Ebola, MERS, and Covid-19 epidemics. We digitized data for these epidemics for their respective time periods: MERS [30] from

TABLE III. For the cumulative infected cases data for MERS, Ebola, and COVID-19, the best-fit functions and the respective relative errors for various stages of evolution shown in Fig. 3.

Epidemic	Best fit functions and errors
MERS $t_{max} = 720$	1) $60e^{0.005t}$ ($\pm 3.2\%$) 2) $0.03t^2 - 24t + 5100$ ($\pm 3.4\%$) 3) $1.5t + 38$ ($\pm 1.2\%$)
Ebola $t_{max} = 360$	1) $14e^{0.03t}$ ($\pm 7.5\%$) 2) $0.7t^3 - 570t^2 + 14\times 10^4 t - 12\times 10^6$ ($\pm 1.4\%$) 3) $135t - 24000$ ($\pm 0.71\%$) 4) $2500\sqrt{t} - 24000$ ($\pm 0.48\%$)
Covid-19 $t_{max} = 90$	See Table II

May 01, 2013 to April 30, 2015; Ebola [31] from May 01, 2014 to April 30, 2015; and Covid-19 [1] from January 22, 2020 to April 20, 2020. In Fig. 3, we plot $I(t)$ vs. normalized time, t/t_{max}, in a *semi-logy* format for all three epidemics. Here, t_{max} (see Table III) is the time span of the epidemic, except for COVID-19 for which t_{max} is taken up to April 20.

In Fig. 3 we present the best-fit curves as dotted lines. Clearly, the three curves look similar, with regimes exhibiting exponential, power-law, and linear growth before flattening. A major difference however is that COVID-19 has two subparts, which is due to the spread of COVID-19 by asymptomatic carriers. As shown in Table II and Fig. 2, the epidemic first spread in China, and then in rest of the world. In contrast, the other two epidemics, Ebola and MERS, were somewhat confined.

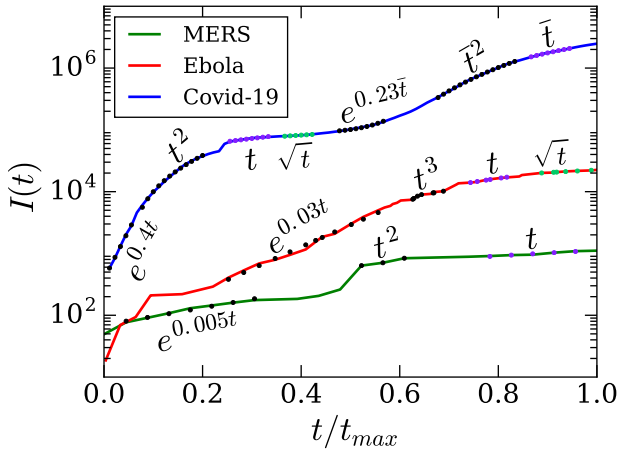

FIG. 3. (color online) For the COVID-19, Ebola and, MERS epidemics, the *semi-logy* plots of total infected individuals ($I(t)$) vs. normalized time (t/t_{max}). The dotted curves represent the best-fit curves using the exponential and polynomial functions. For COVID-19 curve, refer to Fig. 2.

In the next section, we present a model for COVID-19 whose predictions match with the data of several countries.

IV. SEIR MODEL FOR COVID-19 EPIDEMIC

In Sec. II, we analyzed the COVID-19 infection data and observed a power-law growth (followed by a linear regime near saturation) after an exponential growth. In this section, we attempt to get some insights about this transition using SEIR model [3, 4, 7, 32, 33]. Note the other important epidemic models are regression models [34, 35], ARIMA forcasting model [36–38], SIR model [2, 39], etc. All these models have been frequently and successfuly used to analyse the transmission dynamics of COVID-19. For example, Labadin and Hong [32] used this model to predict the second confirmed case in Malaysia.

Recently, Peng et al. [6] constructed a generalised SEIR model for the spread of SARS-Cov-2 virus in China. López and Rodo [7] modified Peng et al. [6]'s model to analyze the data of Spain and Italy up to the end of March. In this section, we will discuss a simplified version of López and Rodo [7]'s SEIR model and fit it with the real-time data of USA, Italy, Spain and Japan till 20th April.

In the model, we assume the disease transmission to take place only among humans. Further, the natural birth and death rates are assumed to be negligible. We divide the total population (N) at a certain place at time t into seven categories: Susceptible ($S(t)$), Exposed ($E(t)$), Infected ($I(t)$), Recovered ($R(t)$), Insusceptible ($P(t)$), Quarantined ($Q(t)$) and Dead ($D(t)$). Here, $Q(t)$ is the number of confirmed infected cases at time t. Time evolution equations of the seven categories are:

$$\frac{dS(t)}{dt} = -\beta\frac{S(t)I(t)}{N} - \alpha S(t), \tag{2}$$

$$\frac{dP(t)}{dt} = \alpha S(t), \tag{3}$$

$$\frac{dE(t)}{dt} = \beta\frac{S(t)I(t)}{N} - \gamma E(t), \tag{4}$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t), \tag{5}$$

$$\frac{dQ(t)}{dt} = \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t), \tag{6}$$

$$\frac{dR(t)}{dt} = \lambda Q(t), \tag{7}$$

$$\frac{dD(t)}{dt} = \kappa D(t), \tag{8}$$

where $\beta, \alpha, \delta, \lambda(t)$, and $\kappa(t)$ are the infection, protection, average quarantine, recovery and mortality rates respectively; and γ^{-1} is the average latency period for COVID-19. The protection rate α is governed by the intensity of contact tracing, lockdown policies, and improvement of health facilities. The time-dependent parameters $\lambda(t)$

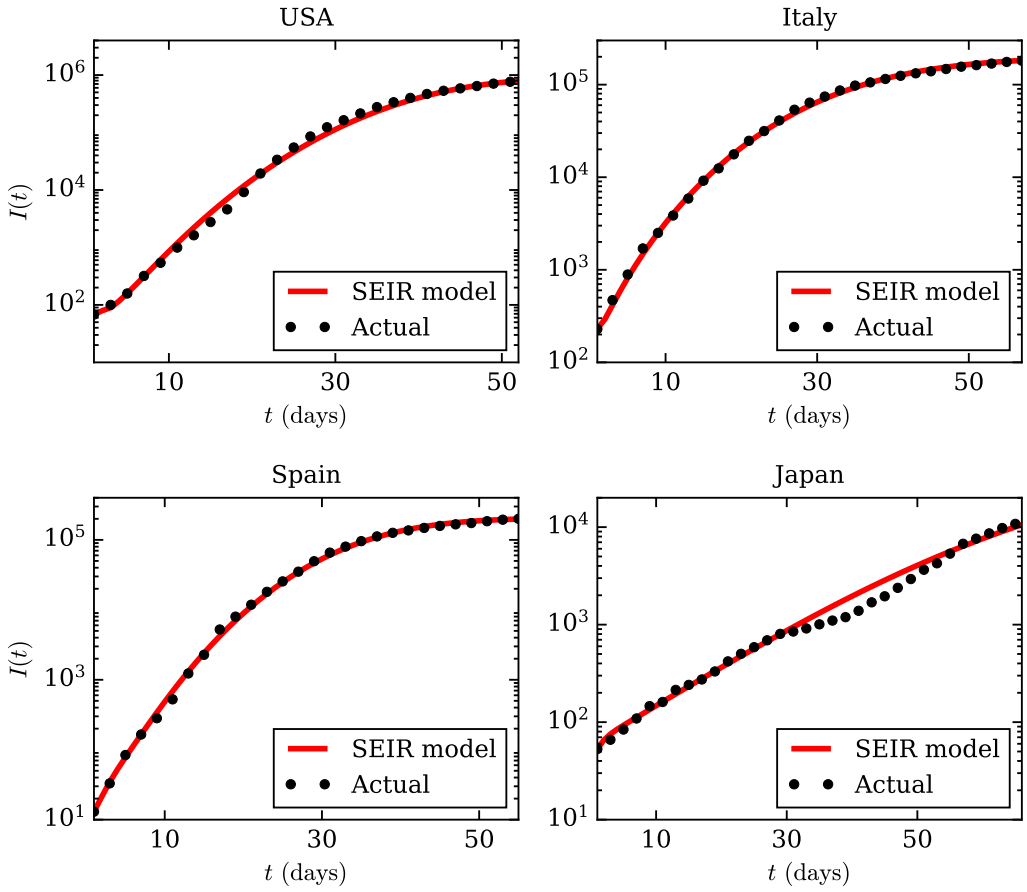

FIG. 4. (color online) For the COVID-19 pandemic, *semi-logy* plots of total infected individuals ($I(t)$) vs. time (t) (black dotted curves) for USA, Italy, Spain and Japan from the available data at *worldOmeter* [1] till April 20. The red curves are the best-fit curves using the SEIR model described in Sec. IV.

and $\kappa(t)$ are modeled as follows [7]:

$$\lambda(t) = \lambda_0(1 - exp(-\lambda_1 t)), \quad (9)$$

$$\kappa(t) = \kappa_0 exp(-\kappa_1 t), \quad (10)$$

where λ_0, κ_0, λ_1, and κ_1 are constants. The functional forms in Eqs. (9-10) are chosen in such a way that the recovery rate saturates and the death rate vanishes with time. Note that the cumulative number of reported infected cases (denoted by $I(t)$ in Sec. II) is the sum of $Q(t)$, $R(t)$, and $D(t)$. For further details, refer to Peng et al. [6] and López and Rodo [7].

We compare the model predictions [Eqs. (2-10)] with the available data [1] for USA, Italy, Spain and Japan. For Spain and Japan, $t = 0$ is taken to be the starting date shown in Table I. For USA and Italy, $t = 0$ corresponds to 29th and 24th February respectively. The end date for all the four countries is 20th April. The initial values of Q, R, and D are taken to be the total active cases, recovered cases and deaths respectively at $t = 0$ for each country. The number of initial insusceptible cases ($P(t = 0)$) is assumed to be zero. We adjust the parameters $\{\beta, \alpha, \gamma, \delta, \{\lambda_0, \lambda_1\}, \{\kappa_0, \kappa_1\}\}$, $E(t = 0)$ and $I(t = 0)$ such that the relative error between the model and actual data is minimized. Note that the initial condition satisfies the relation

$$S(t = 0) = N - E(t = 0) - I(t = 0) - Q(t = 0) - R(t = 0) - D(t = 0) - P(t = 0), \quad (11)$$

where N is the total population of the country.

In Fig. 4, we present the best-fit curves from the SEIR model along with the actual real-time data. Clearly, this model fits well with the data for Italy and Spain. In Table IV we list the numerical values of the best-fit parameters and the relative errors between the predictions and data. Note that for Spain and Italy, López and Rodo [7] considered natural birth and death rates in their model and obtained fits for Q, R and D seperately. In contrast, we stick to the fundamental assumption regarding natural birth and death rates of the basic SEIR model [32] and obtain the fits for $Q(t) + R(t) + D(t)$ till April 20.

Our best-fit values of paramerters for Spain and Italy are nearly consistent with those of López and Rodo [7]. The model shows that high infection rates (β) and small

TABLE IV. Best-fit numerical values of $\{\beta, \alpha, \gamma, \delta, \{\lambda_0, \lambda_1\}, \{\kappa_0, \kappa_1\}\}$ of SEIR model for four countries with relative errors between the best-fit curves and actual data [1]. The best-fit curves are shown in Figs. 4.

	Best-fit parameters of $SEIR$ model								
Country	α	β	γ	δ	λ_0	λ_1	κ_0	κ_1	Errors
USA	0.032	7.1	0.12	1.1	0.067	0.021	2.1	1.5	13.8%
Italy	0.037	1.3	0.24	0.39	0.91	0.085	0.31	0.085	4.1%
Spain	0.018	1.7	1.40	0.92	0.81	0.079	0.96	0.51	6.8%
Japan	0.017	1.7	0.05	0.3	0.096	1.4	0.46	0.12	15.6%

average latency periods γ^{-1} try to push the cumulative number of infected cases (reported) to a large saturation value via an exponential growth. On the other hand, high protection and quarantine rates, α and δ, slow down the growth and minimize the saturation level of the cumulative infected (reported) cases. Thus, the values of the control parameter set $\{\beta, \gamma, \alpha, \delta\}$ in Table IV determine the nature of the power-law after the exponential growth. On the other hand, the linear regime (for Italy, Spain) near the saturation is determined well by the removal rate set $\{\lambda_0, \lambda_1, \kappa_0, \kappa_1\}$. Thus, the present model is consistent with the results presented in Section. II.

In the next section, we will present an another model which is based on delayed differential equations.

V. MODEL BASED ON DELAYED-DIFFERENTIAL EQUATIONS

Although the extended SEIR model [6, 7] is adequate to describe COVID-19 epidemic, it has some limitations which do not enable it to capture the full intricacies of COVID-19. For example, Coronavirus has a pre-symptomatic period, i.e., a period during which people are transmissible but do not show symptoms. There are also a significant number of asymptomatic carriers who do not manifest symptoms during the entire infection period. Finally, due to scarcity of testing equipment, an issue which was even more pronounced during the early phases of the outbreak, it sometimes takes many days for the test results to arrive. These features cannot be accounted for by a real-time model properly and need a delayed differential equation model.

Shayak et al. [12] and Shayak and Rand [13] have formulated dynamic models for COVID-19 based on delay differential equations (DDE). The first model aims to describe community transmission of the disease, while the second one describes a new route to the end of the pandemic through a combination of social distancing, sanitization, contact tracing and preventive testing. We shall briefly describe the two models below.

A. Community transmission

In the community transmission model, our four variables are x, the number of healthy and susceptible people at large (i.e. not in quarantine), y, the number of people who fall sick at large, z, the number of people in quarantine and w, the number of confirmed cases. Our variable choice is somewhat different from a traditional SIR model, since the sets y and w overlap. We have gone with this choice to highlight the discrepancies which occur between the actual progress of the disease in the society and the case history records, on account of the the latency and the testing delay. A full derivation of the model may be found in Shayak et al. [12]. The equations are

$$\begin{aligned}\frac{\mathrm{d}x}{\mathrm{d}t} &= -k_0\mu_1 x\left(y - y_{\tau 1}\right) + k_0\left(1-\mu_1\right)x\left(y - y_{\tau 2}\right) \\ &\quad -k_3 xw - k_4 x + \left(1-\mu_2\right)k_3 x_{\tau 3} w_{\tau 3} + k_4 x_{\tau 3}, \qquad (12)\end{aligned}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -k_0\mu_1 x\left(y - y_{\tau 1}\right) + k_0\left(1-\mu_1\right)x\left(y - y_{\tau 2}\right), \qquad (13)$$

$$\begin{aligned}\frac{\mathrm{d}z}{\mathrm{d}t} &= k_0 x_{\tau 2}\mu_1\left(1-\mu_1\right)\left(y_{\tau 2} - y_{\tau 1+\tau 2}\right) + k_0 x_{\tau 2}(1-\mu_1)^2 \\ &\quad \left(y_{\tau 2} - y_{2\tau 2}\right) + k_3 xw + k_4 x - k_0 x_{\tau 2+\tau 3}\mu_1\left(1-\mu_1\right) \\ &\quad \left(y_{\tau 2+\tau 3} - y_{\tau 1+\tau 2+\tau 3}\right) - k_0 x_{\tau 2+\tau 3}(1-\mu_1)^2 \\ &\quad \left(y_{\tau 2+\tau 3} - y_{2\tau 2+\tau 3}\right) - \mu_2 k_3 x_{\tau 3} w_{\tau 3} \\ &\quad -\left(1-\mu_2\right)k_3 x_{\tau 3} w_{\tau 3} + k_4 x_{\tau 3}, \qquad (14)\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}w}{\mathrm{d}t} &= k_0 x_{\tau 2+\tau 3}\mu_1\left(1-\mu_1\right)\left(y_{\tau 2+\tau 3} - y_{\tau 1+\tau 2+\tau 3}\right) \\ &\quad +k_0 x_{\tau 2+\tau 3}(1-\mu_1)^2\left(y_{\tau 2+\tau 3} - y_{2\tau 2+\tau 3}\right) \\ &\quad +\mu_2 k_3 x_{\tau 3} w_{\tau 3}, \qquad (15)\end{aligned}$$

In the above equations, variables with subscripts of τ indictes the variable delayed by that amount, thus $y_{\tau 1}$ means $y(t-\tau_1)$ etc. The parameters are k_0: the basic transmission rate of the disease which includes factors such as R_0 of the pathogen and degree of social distancing; k_3: a proportionality constant expressing the number of anticipatory quarantines; k_4: a proportionality constant reflecting the number of flu and other diseases who report for COVID testing; μ_1: the fraction of asymptomatic carriers; μ_2: the fraction of anticipatory quar-

antines who test positive; and the three delays τ_1: the duration that an asymptomatic carrier is transmissible, τ_2: the pre-symptomatic period i.e. the interval between start of transmissibility and manifestation of symptoms, and τ_3: the interval between reporting for testing and the results being obtained. The most important parameters among these are k_0, τ_2 and τ_3.

In Fig. 5, we have fitted this data to the case histories of Japan, Italy, Spain and USA. For each fit, we have taken the parameter values $\tau_1 = 7$, $\tau_2 = 3$, $\tau_3 = 5$, $\mu_1 = 0.18$, and $\mu_2 = 0.02$ and we have started the free evolution of (12) at the time when the case count just crossed 100. Unlike an ordinary differential equation which needs only an initial value, a DDE needs an initial function, with the duration of this function being the longest delay involved in the equation. In (12) this maximum delay is 15 days. Hence we need to seed (12) with initial conditions over a 15-day period. During this seeding period, we assume constant value of x, zero value of z, and linear growths in z and w. We have obtained these growth rates from a fit to the subsequent data. Because of the nature of the initial conditions, a comparison between model and reality in the first 15 days is meaningless. During the next 15 days also, we find considerable error on account of the initial condition, and only after that do these errors die out and the model converges to the actual case trajectories.

We have performed an error analysis of the fits by calculating the percent error at each data point and averaging. We find large errors in the first 15-day period so we ignore the same for consideration, and perform the error analysis over the next few days. The errors are 6.17 percent for Japan, 5.09 percent for Italy, 10.26 percent for Spain and 2.79 percent for USA.

B. Self-burnout

In the controlled endgame phase of the epidemic, which we call self-burnout [13], we have a slightly different equation. In this phase, there is extensive enforcement of separation minima (a term we prefer to social distancing as it does not carry connotations of emotional isolation) so the rate of new cases does not depend on x. Rather, we assume that each sick person spreads the disease at a constant rate m_0. Moreover, contact tracing manages to capture a fraction $1 - \mu_3$ of all the sick patients and places them into quarantine. Under these conditions, the dynamic model for the spread of new cases is

$$\frac{\mathrm{d}y}{\mathrm{d}t} = m_0[y(t) - (1-\mu_3)\, y(t-\tau_2/2) - (1-\mu_1)\,\mu_3 y(t-\tau_2) - \mu_1\mu_3 y(t-\tau_1)] \quad (16)$$

One solution to the above is $y = const$. It has been shown in Shayak and Rand [13] that this solution is stable if and only if

$$m_0\left(\frac{1+\mu_3-2\mu_1\mu_3}{2}\tau_2 + \mu_1\mu_3\tau_1\right) < 1 \quad (17)$$

This identifies a maximum value of m_0 for which the solution is stable i.e. the epidemic gets over in time. For the plausible parameter values $\tau_1 = 7$, $\tau_2 = 3$, $\mu_1 = 1/5$ and $\mu_3 = 1/2$, the critical value of m_0 turns out to be $20/53$. Here we assume that the test results are instantaneous due to high testing capacity present in the region.

We have performed simulation runs of (16) with the above parameter values and m_0 having the values 70, 80 and 90 percent of the critical respectively. We seed the equation with the linear function $y = 1000t$ for the first ten days. We find a considerable region thereafter where the case histories show a $\sqrt{t}$ profile before saturating. This explains why the countries which are achieving saturation are showing a pronounced $\sqrt{t}$ phase after the linear phase.

To further bolster our case, we consider two countries (South Korea and Austria) which have shown a very good linear region followed by saturation. South Korea showed linear regime from March 28 to April 04 (with 9478 and 10156 cases respectively), after which it enters the burnout phase. Using this as the seeding data and taking the parameter values mentioned above, we find the best fit for the next 20 days for m_0 equal to 77 percent of the critical. The fit error is 0.34 percent. Note that Shayak and Rand [13] has found an m_0 of 75 and not 77 percent of critical, because the fit was upto a smaller duration.

Austria showed linear regime from March 28 to April 01 (with 6250 and 10711 cases respectively) before entering self-burnout phase. We find the best fit for m_0 equal to 79 percent of the critical. The error is 2.6 percent. However, the actual data for the 8th to the 15th day appears to be too low—the curve has a convex profile which is probably unrealistic. If we consider the error from the 16th to the 30th day then we find a value of 1.1 percent only. We present these best-fit curves in Fig. 6. Other regions which are in the self-burnout phase are Vietnam, Australia, New Zealand, and Goa, Kerala and Odisha in India. We have chosen South Korea and Austria since their data shows the smoothest profile on account of high testing capacity.

Both SEIR and the DDE models describe the evolution of COVID-19 epidemic quite well for many countries. A detailed comparison between the two models will be performed in future. Also, we plan to employ the two models to understand the epidemic evolution for many nations.

VI. DISCUSSIONS AND CONCLUSIONS

COVID-19 pandemic involves many factors, for example, asymptomatic carriers, lockdown, social distancing, quarantine, etc. Considering these complex issues, we focus on data analysis. In particular, we analyze the real-time infection data of COVID-19 epidemic for 21 nations up to April 20, 2020. Our analysis shows that many nations are close to flattening the epidemic curve.

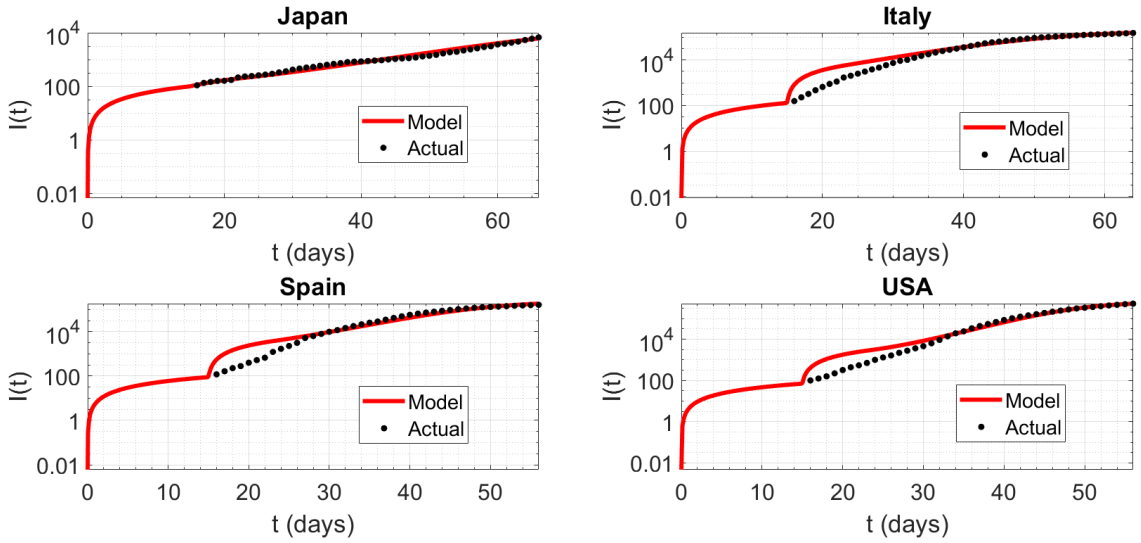

FIG. 5. (color online) For the COVID-19 pandemic, *semi-logy* plots of total infected individuals ($I(t)$) vs. time (t) (black dotted curves) for USA, Italy, Spain and Japan from the available data at *worldOmeter* [1]. The red curves are the best fit curves using the community transmission delayed-differential equation model described in Section. V.

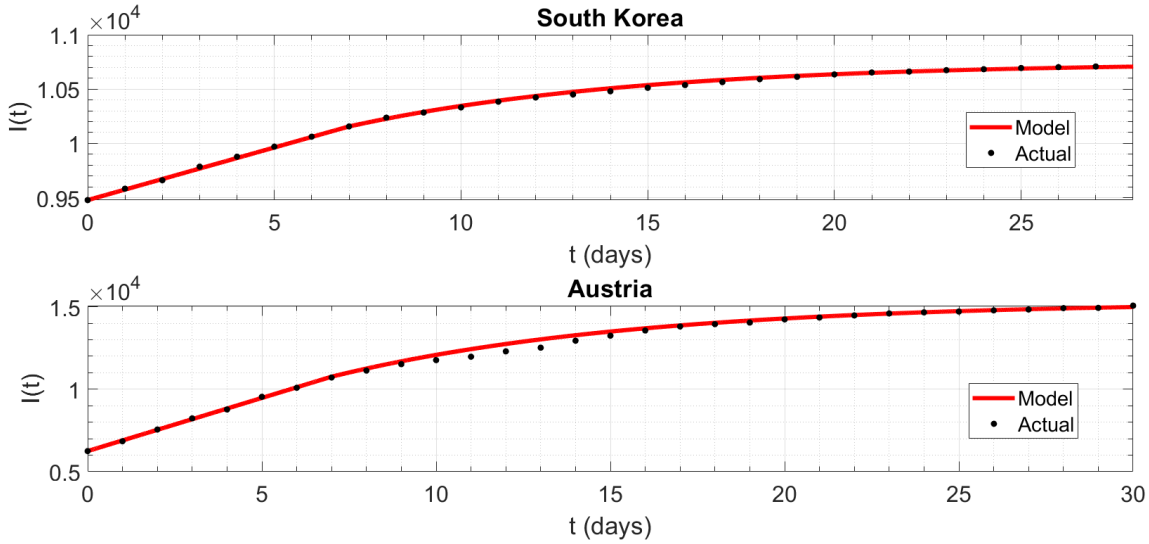

FIG. 6. (color online) (color online)For the COVID-19 pandemic, *semi-logy* plots of total infected individuals ($I(t)$) vs. time (t) (black dotted curves) for South Korea and Austria from the available data at *worldOmeter* [1]. The red curves are the best fit curves using the self-burnout delayed-differential equation model described in Section. V.

A key feature of our analysis is the emergence of power law behavior after an exponential growth, which has also been observed by other researchers [14–21]. The exponential growth is easily explained using $\dot{I} \propto \beta I$ relation, which arises due to the spread by contact. For power-law growth, $I(t) \sim t^n$, the above relation is modified to $\dot{I} \sim I^{1-1/n}$. The suppresion of $I^{-1/n}$ in $\dot{I}$ could be attributed to lockdowns and social distancing etc. A careful analysis of the epidemic models should yield this feature. Interestingly, Ebola and MERS also exhibit similar behavior. This generic feature is very useful for the forecast of the epidemic evolution.

Note that the $I(t)$ curve needs to turn from convex (during the exponential growth) to concave for flattening. Hence, a transition from an exponential growth to a power law growth is expected. The lockdowns and social distancing is likely to make the transition earlier, thus suppressing the exponential growth to some degree. Earlier, Verma et al. [29] had conjectured that the power-law growth may occur due to asymptomatic carriers and/or

community spread. It appears that this conjecture may be incorrect.

In this paper we only studied the infection counts. However, it is obvious that during the growth phase, the active cases, as well as death counts, would follow similar pattern as $I(t)$. The total death count too flattens along with the infection count, but the active cases decreases with time during the saturation.

Prakash et al. [24] studied the phase space portraits, that is, $\dot{I}$ vs. I plots. They observed the phase-space curves to be linear. This is natural for the exponential growth ($\dot{I} \propto \beta I$), as well as for the power-law growth with large exponent n because $\dot{I} \sim I^{1-1/n}$. In another interesting of analysis of COVID-19 epidemic, Schüttler et al. [25] and Marsland and Mehta [18] argued that $I(t)$ or total death count could be modelled using error function. Using this result, we may be able to predict the asymptotic behavior of $I(t)$ that may yield valuable clues regarding the extent and duration of the epidemic.

Epidemic spread has similarities with rumor spread, as well as with the growth of a network [4, 40]. A comparison of the power-law growth in these systems will yield fruitful results for the epidemic forecast.

In summary, COVID-19 epidemic data reveals interesting properties that can be used for its forecast.

ACKNOWLEDGMENTS

The authors thank Santosh Ansumali for useful discussions. We also thank the *worldOmeter* for the data that made this work possible. Ali Asad is supported by Indo-French (CEFIPRA) project 6104-1, and Soumyadeep Chatterjee is supported by INSPIRE fellowship (IF180094) of Department of Science & Technology, India.

[1] WorldOMeter, URL https://www.worldometers.info/coronavirus/.
[2] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society A **115**, 700 (1927).
[3] O. N. Bjørnstad, *Epidemics: Models and Data using R* (Springer, 2018).
[4] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction* (Cambridge University Press, 2001).
[5] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, Science **6**, DOI: 10.1126/science.abb3221 (2020).
[6] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, medarxiv.org (doi: 10.1101/2020.03.14.20036202) (2020), 2002.06563v1.
[7] L. R. López and X. Rodo, medarxiv.org (doi: 10.1101/2020.03.27.20045005) (2020).
[8] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, and J. Edmunds, The Lancet Global Health **8**, DOI: 10.1016/S2214 (2020).
[9] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, et al., Science **(preprint)** (2020).
[10] S. Mandal, T. Bhatnagar, N. Arinaminpathy, A. Agarwal, A. Chowdhury, M. Murhekar, R. Gangakhedkar, and S. Sarkar, Indian Journal of Medical Research **(preprint)** (2020).
[11] R. Singh and R. Adhikari, arXiv.org p. arXiv:2003.02223 (2020).
[12] B. Shayak, M. M. Sharma, R. H. Rand, A. K. Singh, and A. Misra, medrxiv.org (doi:10.1101/2020.03.29.20047035) (2020).
[13] B. Shayak and R. H. Rand, medrxiv.org (doi:10.1101/2020.04.17.20069443) (2020).
[14] A. L. Ziff and R. M. Ziff, medrxiv.org (DOI: 10.1101/2020.02.16.20023820) (2020).
[15] N. L. Komarova and D. Wodarz, medrxiv.org (DOI: 10.1101/2020.03.30.20047274) (2020).
[16] C. Manchein, E. L. Brugnago, R. M. da Silva, C. F. O. Mendes, and M. W. Beims, arXiv p. arXiv:2004.00044 (2020), 2004.00044.
[17] B. Blasius, arXiv p. arXiv:2004.00940 (2020), 2004.00940.
[18] R. Marsland and P. Mehta, arXiv p. arXiv:2004.10666 (2020), 2004.10666.
[19] M. Li, J. Chen, and Y. Deng, arXiv p. arXiv:2002.09199 (2020), 2002.09199.
[20] H. M. Singer, arXiv p. arXiv:2004.03859 (2020), 2004.03859.
[21] B. K. Beare and A. A. Toda, arXiv p. arXiv:2004.12772 (2020), 2004.12772.
[22] I. Cherednik and U. C. Hill, arXiv.org p. arXiv:2004.06021 (2020).
[23] M. K. Verma, A. Asad, and S. Chatterjee, medrxiv.org (DOI:10.1101/2020.04.02.20051680) (2020).
[24] M. K. Prakash, S. Kaushal, S. Bhattacharya, A. Chandran, A. Kumar, and S. Ansumali, medrxiv.org (DOI:10.1101/2020.04.08.20057414) (2020).
[25] J. Schüttler, R. Schlickeiser, F. Schlickeiser, and M. Kröger, preprints.org (2020), doi:10.20944/preprints202004.0175.v1.
[26] K. Wu, D. Darcet, Q. Wang, and D. Sornette, arXiv.org (2020), arXiv2003.05681.
[27] Y. Min, X. Jin, Y. Ge, and J. Chang, PLoS ONE **8**, DOI: 10.1371/journal.pone.0057100 (2013).
[28] Johns Hopkins University, Corona Resource Center (2020), URL https://coronavirus.jhu.edu/map.html.
[29] M. K. Verma, A. Asad, and S. Chatterjee, medRxiv (2020), URL https://www.medrxiv.org/content/early/2020/04/16/2020.04.02.20051680.
[30] Phoenix7777 on Wikipedia, URL https://upload.wikimedia.org/wikipedia/commons/7/7d/World_MERS_outbreak.svg.
[31] J. Ma, Infectious Disease Modelling **5**, 129 (2020).
[32] J. Labadin and B. H. Hong, medrxiv.org (doi: 0.1101/2020.02.07.20021188) (2020).

[33] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, arXiv.org (2020), 2002.06563v1.
[34] B. Pirouz, S. Shaffiee Haghshenas, S. Shaffiee Haghshenas, and P. Piro, Sustainability **12**, 2427 (2020).
[35] R. Gupta, G. Pandey, P. Chaudhary, and S. K. Pal, medrxiv.org (doi:10.1101/2020.04.01.20049825) (2020).
[36] R. Anne, medrxiv.org (doi:10.1101/2020.04.18.20070631) (2020).
[37] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, Data in Brief **29**, 105340 (2020).
[38] S. Deb and M. Majumdar, arXiv.org p. arXiv:2003.10655 (2020).
[39] R. Ranjan, medrxiv.org (doi:10.1101/2020.04.12.20061002) (2020).
[40] M. Newman, Phys. Today **61**, 33 (2008).