ARTICLE

The impact of travel and timing in eliminating COVID-19

Alexander F. Siegenfeld [1,2 ✉] & Yaneer Bar-Yam[2]

While the spread of communicable diseases such as coronavirus disease 2019 (COVID-19) is often analyzed assuming a well-mixed population, more realistic models distinguish between transmission within and between geographic regions. A disease can be eliminated if the *region-to-region reproductive number*—i.e., the average number of other regions to which a single infected region will transmit the disease—is reduced to less than one. Here we show that this region-to-region reproductive number is proportional to the travel rate between regions and exponential in the length of the time-delay before region-level control measures are imposed. If, on average, infected regions (including those that become re-infected in the future) impose social distancing measures shortly after experiencing community transmission, the number of infected regions, and thus the number of regions in which such measures are required, will exponentially decrease over time. Elimination will in this case be a stable fixed point even after the social distancing measures have been lifted from most of the regions.

[1] Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. [2] New England Complex Systems Institute, 277 Broadway, Cambridge, MA 02139, USA. ✉email: asiegenf@mit.edu

The outbreak of coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, China, reportedly in December 2019[1] and has since become a severe pandemic[2]. Understanding the dynamics of disease transmission both within and between regions[3–10] can provide insight into how to eliminate the outbreak by imposing restrictive measures only where the virus is locally spreading, thus minimizing larger-scale economic impacts[11–14]. Regions could be cities, counties, states, or any other partition of a population such that the disease transmission occurs predominantly within (as opposed to between) regions. The choice of region size depends on the spatial granularity at which policy makers are willing and able to control disease transmission.

Since it is impossible to model all the details of real-world systems, identifying large-scale behaviors is necessary to determine which details matter and how[15–17]. For COVID-19, the parameters for these large-scale behaviors include the average size of an outbreak within a region and the transmissibility of the outbreak between regions. The values of these two parameters, both of which can be controlled with interventions, determine whether the behavior of the epidemic within a collection of regions (e.g., a country) is that of exponential spread until saturation (e.g., dynamic endemic equilibrium or herd immunity) or exponential decay until elimination. In the latter regime, elimination will be stable and most regions within the collection can fully open up their economies, with only local and sporadic social distancing measures needed to contain outbreaks arising from hidden or imported cases.

A central concept in the study of disease spread is the reproductive number R, i.e., the average number of people to whom a typical infected individual will transmit the disease[18]. To consider the transmission between regions, an analogous *region-to-region reproductive number* R_* can be defined as the number of other regions (including those that have been previously infected) to which a single infected region will transmit the infection on average[19,20]. Just as an outbreak cannot sustainably propagate (i.e., elimination is stable) among individuals within a region if $R < 1$, an outbreak cannot sustainably propagate *among regions* within a collection if $R_* < 1$.

Here, we analyze how local measures can support the elimination of COVID-19 while avoiding large-scale lockdowns where they are unnecessary. We find that reductions in travel and the speed with which regions act to contain future outbreaks play decisive roles in whether COVID-19 is eliminated from a collection of regions (i.e., whether $R_* < 1$). Such an elimination is stable: for each outbreak caused by imported or undetected cases, only a few regions—fewer than $1/(1 − R_*)$ on average—need to temporarily impose measures while the rest of the regions in the collection remain open. If infected regions (including those that become re-infected in the future) impose control measures shortly after experiencing community transmission (i.e., shortly after no longer being a "green zone"), the number of infected regions, and thus the number of regions in which such measures are required, will exponentially decrease over time.

Results

General model. The disease is modeled as being transmitted among individuals within a region, with travel allowing the disease to spread between regions. A collection of regions is defined as any partition of a population such that travel/social contact within each region far exceeds that between them (e.g. the U.S. could be partitioned by county or commuting zone boundaries), in order that an infected individual is far more likely to transmit the virus to someone within the same region rather than to someone in a different region (i.e. transmission between regions can be treated as a first-order perturbation). Policy makers are assumed to act homogeneously within each region and to have the ability to act independently between regions—thus, the choice of how to partition a population into regions must take into consideration the spatial granularity at which policy makers are willing and able to make decisions. Treating larger areas as single regions means that social distancing measures will be homogeneously applied to larger areas but also means that it is easier to achieve lower per capita travel rates between such areas.

We define a region as *infected* if someone with the infection enters the region. Community transmission, also known as community spread, is said to occur when individuals within a region are infected from an unknown source[21]. Conditioning on region c being infected, we let N_c be a random variable (that could be zero) denoting the total number of new infections that occur in region c from the time after the region is infected to the time at which there is no more community transmission, at which point we define the region as no longer infected. Let p_c be the probability that an infected individual in region c will travel to another region during the period in which that individual is contagious. Then, the region-to-region reproductive number for region c—which we define as the expectation of the number of regions that region c will infect if it becomes infected—is $R_*^c = \mathbb{E}[N_c]p_c$. The expected outbreak size $\mathbb{E}[N_c]$ can be greatly reduced if regions impose strong social distancing measures shortly after detecting community transmission (these social distancing measures can be lifted once community transmission has been contained), while p_c can be reduced by reducing travel out of infected regions.

R_*^c may differ from region to region, with R_*^c for a particular region depending not only on internal factors but also on the network connectivity between that region and others, which in turn depends on the sizes of regions (i.e. the scale/level of granularity of the interventions). If the interventions are sufficiently fast and strong such that R_*—the average value of R_*^c over a collection of regions with each region weighted by its probability of being infected[18]—is less than one, then the outbreak will not be self-sustaining within that collection of regions. Put another way, a collection of regions can exist in one of two regimes: a regime for which elimination is a stable fixed point of the system and a regime for which it is unstable (see Fig. 1). A change in policy can shift a collection of regions from the unstable regime ($R_* > 1$) to the stable regime ($R_* < 1$) or vice versa. Although the values of N_c for a collection of regions could currently be high, R_* is determined by $\mathbb{E}[N_c]$ and p_c for the regions that will be infected or re-infected in the future after social distancing has eliminated the virus from currently infected regions.

We note that if region c were infected multiple times, $\mathbb{E}[N_c]$ would be higher than if it were infected once, but it is assumed that infecting an already infected region will on average contribute no more to disease spread than infecting a currently uninfected region. Thus, like the basic reproductive number, this region-to-region reproductive number overestimates the disease spread away from the limit of most regions being uninfected by counting a single region that has been infected multiple times during a single outbreak as multiple regions being infected. To the extent that a region being infected multiple times has a linear effect on its expected total number of cases, this approximation will not impact the values of R_*. However, if a region that is currently implementing control measures because of previous importations receives additional imported infections, these additional importations are not likely to have as large an effect
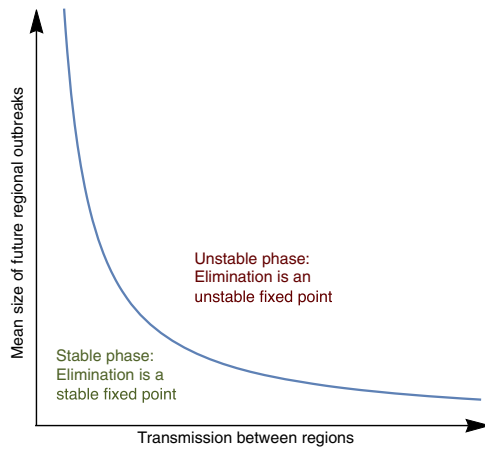
Fig. 1 Phase diagram of COVID-19 for a collection of regions. Elimination can either be stable or unstable; the stability of elimination is a function of (1) the average total number of cases that will result from the disease being transmitted to a region (which depends on, among other factors, how quickly regions locally impose social distancing measures if they experience community transmission), and (2) the probability that an infected individual in one region will infect an individual in another (which depends on the rate of travel between regions). Note that the stable regime does not require that every region implement social distancing measures but rather only those with active community transmission. Thus, once elimination is achieved, it can be maintained while most regions remain economically open, with outbreaks caused by hidden or imported cases contained by social distancing measures that are localized in both space and time. Image source: ref. [25].

as the previous ones that occurred before control measures were implemented. Thus, R_* may be overestimated when a high total number of infections within a collection of regions makes it common for regions to experience multiple importations during a single outbreak. We also note that, as in SIS compartmental models, an infected region in which the virus is contained can later be re-infected.

It is important to keep in mind that a collection of regions can be in the stable regime, in which the *region-to-region* reproductive number for the collection R_* is less than one, while, at any given time, the reproductive number R of most regions within the collection is greater than one. In other words, in order for $R_* < 1$ for the collection of regions, individual regions within the collections need only maintain $R < 1$ so long as community transmission persists within them; otherwise, they can lift social distancing measures and open up economically. For every infection that a collection of regions with $R_* < 1$ imports (not including cases that were quarantined at the border), the average number of regions within that collection that will need to impose social distancing measures is bounded by $\alpha \sum_{n=0}^{\infty} (R_*)^n = \alpha/(1 - R_*)$ where α, which can be reduced with testing and contact tracing, is the probability that an imported case will result in community transmission within a region. Thus, border control need not be perfect; if a collection of regions has sufficiently good border control and policies that ensure $R_* < 1$, elimination will remain stable while most of the collection's regions remain open most of the time.

Modeling the approximate size of regional outbreaks. We now describe a simple mathematical model to estimate N_c, not to provide a precise description of the epidemic trajectory but rather to clarify how various interventions may affect outbreak size. This specific model assumes exponential growth before control measures are implemented followed by exponential decline afterwards.

However, the validity of this assumption is not essential to the main results, which depend only on the number of infected individuals traveling out of each region. Modeling the number of active cases within a region using exponential growth and decay serves mainly to give a quantitative handle on the rates of increase and decrease in cases, but comparable results could be obtained using other epidemic trajectories. (Similarly, although SIS and SIR models assume an exponential distribution of generation intervals, this particular assumption does not affect the conditions under which such models are valid, so long as their recovery rate γ is treated as an effective parameter.) The one key assumption is that the reproductive number within individual regions can be reduced below one with sufficiently strong measures such as mask-wearing, social distancing, contact tracing, etc. This assumption has been validated empirically by the many regions around the world in which the number of cases has declined for a sustained period of time.

Let i_0^c be a stochastic factor that roughly corresponds to the initial foothold that the virus gains in region c conditioning on an infected individual entering the region, with $i_0^c = 0$ corresponding to the case in which no one was infected or a few people were infected but the outbreak was contained by contact tracing/ quarantine or otherwise spread no further (e.g. no community transmission). The probability distribution of i_0^c will depend on the distribution of infectiousness[22], a function of both biological and social factors. If the outbreak is contained ($i_0^c = 0$), then the number of active infections remains at zero for the purposes of this model because—by definition—the outbreak has no chance of spreading to other regions.

If the outbreak is not contained ($i_0^c \neq 0$), the number of active infections is modeled as growing with time t at an exponential rate $e^{r_c t}$. Then, after time T_c (the delay in response), the region implements social distancing measures that cause the number of active infections to decay as e^{-t/τ_c}, where τ_c is the time-constant (i.e. the inverse of the rate) of decay. Such exponential decrease will occur if the social distancing measures, together with mask-wearing and testing/contact tracing/quarantine, can reduce the reproductive number (R) of the virus below one. The greater the reduction in R, the smaller the value of τ_c and thus the faster the decrease in infections. (For $R < 1$, R is related to τ_c by $1 = \int_0^{\infty} Rg(t)e^{t/\tau_c}dt$ where $g(t)$ is the distribution of generation intervals[23].) We note that this assumption of exponential increase followed by exponential decay after intervention assumes that the proportion of susceptible individuals is roughly constant, i.e. that the region intervenes before a significant fraction of its population is infected, which is the regime with which we are concerned. To the extent that this assumption does not hold, its use will result in an overestimate of the number of infected individuals and thus does not affect our main conclusions.

The number of active infections in region c as a function of time (see Fig. 2) can therefore be written as:

$$i_c(t) = \begin{cases} i_0^c e^{r_c t} & t \leq T_c \\ i_0^c e^{r_c T_c} e^{-(t-T_c)/\tau_c} & t \geq T_c \end{cases} \quad (1)$$

The social distancing measures can be lifted once no active infections remain in the region or once there is no more community transmission and the remaining infections can be contained with contact tracing. Solving for $i_c(t) = 1$ (assuming $i_0^c \neq 0$) yields an approximate duration for the social distancing measures of

$$\tau_c(r_c T_c + \ln i_0^c) \quad (2)$$

As the number of cases becomes increasingly small, testing and contact tracing become increasingly effective and can hasten the
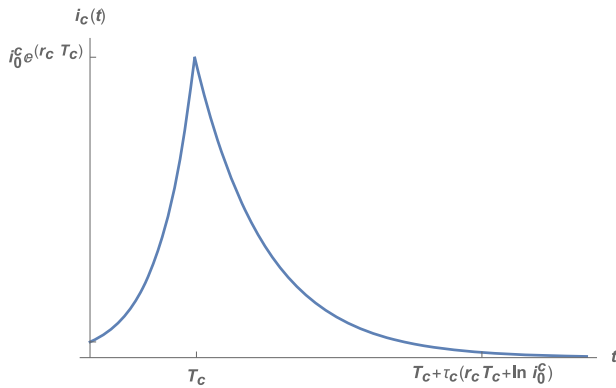
Fig. 2 Representative epidemic trajectory of an outbreak within a single region. The number of active infections i_c in a region c according to the model described by Eq. (1) are shown as a function of time. The time at which community transmission begins is defined as $t = 0$. At $t = T_c$ (a duration of T_c after the start of community transmission), control measures that reduce the reproductive number of the virus to less than one are implemented. At $t = T_c$, the size of the outbreak is $i_0^c e^{r_c T_c}$ where i_0^c is a stochastic factor and r_c is the rate of exponential growth without the control measures. In order to eliminate the community transmission, we estimate that the control measures must remain in place for a duration of approximately $\tau_c(r_c T_c + \ln i_0^c)$ (see Eq. (2)), where τ_c^{-1} is the rate of exponential decay in the number of infections under the control measures. Thus, the longer the region waits to enact the measures, the longer the total amount of time they must remain in place.

end of community transmission, thereby allowing the social distancing measures to be lifted. The probability that the number of infections will rebound after the social distancing measures are lifted—in which case an additional phase of such measures will be needed, as in the Imperial College report[24]—will depend on the probability of importation from other regions, which in turn will depend on the region-to-region reproductive number R_*. Even though the virus can be re-imported, as long as $R_* < 1$, the number of infected regions will on average decrease over time, since re-importation events are included in R_*.

Modeling transmission between regions. Each infected region c infects a currently uninfected region with a probability rate proportional to the number of active infections $i_c(t)$ times the probability rate p_c that an infected individual will travel to an uninfected region. The number of new infected regions spawned by region c can thus be modeled as a Poisson process with rate $i_c(t)p_c$. As described above, this modeling assumption overestimates the spread of the disease to new regions by counting a single new region that has been infected multiple times (possibly by multiple other regions) as multiple new infected regions. The smaller the number of regions infected and the smaller the probability of one region infecting another, the smaller the probability that the same region will be infected twice. Nonetheless, for certain regional connectivity networks, this model may overestimate R_*. (Our main conclusions are unaffected because R_* will be less than one if its overestimate is.)

Let p_0^c be the per capita probability rate before time T_c of individuals in region c traveling to other regions and p_1^c be the probability rate afterwards (p_1^c will be less than p_0^c if travel is discouraged and/or restricted at the time social distancing measures are implemented). The number of new regions that are infected by region c will then be a Poisson random variable

with a mean of

$$i_0^c p_0^c \int_0^{T_c} e^{r_c t}\, dt + i_0^c p_1^c e^{r_c T_c} \int_0^{\tau_c(r_c T_c + \ln i_0^c)} e^{-t/\tau_c}\, dt \qquad (3)$$

$$= \begin{cases} i_0^c \left(p_0^c \frac{e^{r_c T_c} - 1}{r_c} + p_1^c \tau_c \left(e^{r_c T_c} - \frac{1}{i_0^c} \right) \right) & i_0^c > 0 \\ 0 & i_0^c = 0 \end{cases} \qquad (4)$$

Taking the expected value over i_0^c (and allowing for a slight overestimate of R_*^c by treating the $\frac{1}{i_0^c}$ term as negligible when $i_0^c > 0$) yields

$$R_*^c = \mathbb{E}[N_c] p_c = \mathbb{E}[i_0^c] \left(p_0^c \frac{e^{r_c T_c} - 1}{r_c} + p_1^c \tau_c e^{r_c T_c} \right) \qquad (5)$$

(The value of R_* for a collection of regions is then an appropriately weighted average of R_*^c over that collection, as described above.)

Estimating R_* for COVID-19. In order to better understand the extent of the measures required to achieve $R_* < 1$, we estimate the values of the parameters in Eq. (5) (see Fig. 3 and Methods) in order to determine R_*^c as a function of the time-delay before social distancing measures are enacted, as shown in Fig. 4. Note that the time-delay is measured from the time at which exponential growth begins to occur—which could be as early as the beginning of community transmission within the region—not the time at which exponential growth is first measured. The latter may lag the former due to delays in testing and the possibility of pre-symptomatic/asymptomatic spread.

Discussion
Our analysis and parameter estimates are conservative and are therefore likely to overestimate R_*, meaning that there may be more room for error than Fig. 4 suggests. However, given the considerable uncertainty surrounding pandemics and the impossibility of precisely predicting their trajectories[25], R_* should be reduced as much as possible so as not to take any chances, as well as so that the virus will be eliminated (and economies more fully reopen) as quickly as possible. Since R_* is proportional to both $\mathbb{E}[N_c]$ and the travel rates between regions, any intervention that reduces the sizes of outbreaks within regions and/or travel between them will also reduce R_*. In the language of Eq. (5) (see Table 1):

- A reduction in travel from region c results in a linear reduction in R_*^c through p_0^c and p_1^c.
- Improvements in testing, contact tracing, and quarantine reduce R_*^c through $\mathbb{E}[i_0^c]$, r_c, and τ_c.
- Preemptive measures (i.e. measures before T_c, including when no spreading has been detected) such as mask-wearing and the reduction of large gatherings reduces the probability of a super-spreader event as well as general transmission, reducing R_*^c through both $\mathbb{E}[i_0^c]$ and r_c. Because—in the absence of interventions—a small fraction of COVID-19 cases cause most of the spread[26,27], the reduction of super-spreader events can have an outsized impact.
- Reductions in $r_c T_c$ not only exponentially reduce R_*^c (as well as the total number of infections within the region) but also linearly reduce the amount of time for which the social distancing measures must remain in place. Early-detection systems[28,29] and more comprehensive testing may greatly reduce T_c.
- Stronger social distancing measures (after time T_c) decrease τ_c, which results in a linear decrease in both R_*^c and the time for which the distancing measures must remain in place.
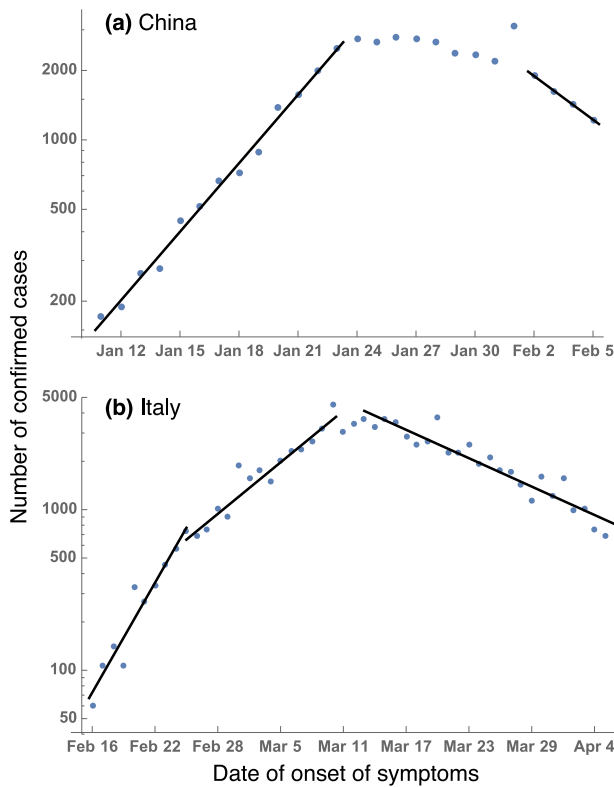
Fig. 3 Logarithmic plots of confirmed cases by date of symptom onset in China and Italy. a The daily number of confirmed cases in China—by date that these patients self-reported as the onset of their symptoms—are shown as dots on a logarithmic scale. The solid lines are the best ordinary least squares linear fits to the natural logarithm of the number of cases: For Jan. 11-23 (up until the lockdown), the slope (in units of day^{-1}) is 0.228 ($R^2 = 0.991$, 95% confidence interval (CI) [0.214, 0.242]), which corresponds to a doubling time of 3.04 days. For Feb. 2-5 the slope is -0.145 ($R^2 = 0.999$, 95% CI [-0.160, -0.131]), which corresponds to a halving time of 4.78 days. Data are from the Chinese Center for Disease Control and Prevention[38], which includes cases diagnosed through Feb. 11. Not pictured: There is a drop in cases with onsets of symptoms after Feb. 5[38], likely due to many of those cases being diagnosed after Feb. 11. **b** The daily number of confirmed cases in Italy by date of symptom onset are shown as dots on a logarithmic scale (data are from Italian authorities[47]). The best ordinary least squares linear fits are shown as solid lines and have slopes (in units of day^{-1}) of 0.262 ($R^2 = 0.927$, 95% CI [0.202, 0.322]) for Feb. 16-25, 0.123 ($R^2 = 0.923$, 95% CI [0.102, 0.144]) for Feb. 25 - Mar. 10, and -0.068 ($R^2 = 0.901$, 95% CI [-0.078, -0.058]) for Mar. 13 - Apr. 5. The change in the exponential growth rate from 0.262 to 0.123 likely occurred due to partial measures implemented by Italy, but it was not until a nationwide lockdown was implemented on March 9 that exponential growth changed to exponential decline. The rate of decline is much slower in Italy than in China, perhaps due to China's stronger lockdown enforcement and contact tracing/quarantine measures.

We conclude with a few comments. First, without the timely implementation of control measures that reduce the local reproductive number to less than one, restricting travel from infected regions serves only to delay the spread of the outbreak, as found in other studies[30–36]. However, when a reduction in travel is coupled with such measures, the travel reduction will not only delay the spread of the outbreak but in some cases will also be the determining factor in whether or not an outbreak is eliminated. (If $R_* < 1$ can be achieved without reducing travel, travel reductions can greatly decrease the duration and total case count of the epidemic by further reducing R_*.) Empirically, travel
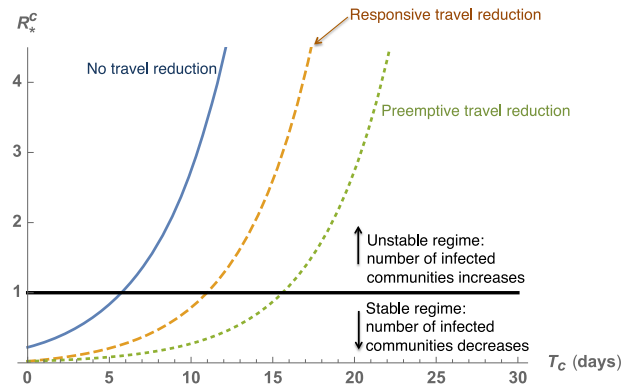

Fig. 4 Dependence of the region-to-region reproductive number on travel policies and the time delay before regions that experience community transmission impose social distancing measures. Each curve depicts, under different travel policies, the average number of regions R_*^c to which region c will transmit the disease as a function of the time delay T_c between when region c experiences community transmission and when it imposes social distancing measures (Eq. (5)). If the region-to-region reproductive number R_* (a weighted average of R_*^c over a collection of regions) is less than one, the number of infected regions will exponentially decrease and the elimination of the disease within that collection of regions will be stable; otherwise, the number of infected regions will increase until saturation. Parameter values: All curves use average initial foothold $\mathbb{E}[r_0^c] = 3.7$, inverse exponential decline rate $\tau_c = 15$ days, and exponential growth rate $r_c = 0.228$ day^{-1}. Solid curve: no travel reduction; the per capita travel rates out of region c before and after T_c are, respectively, $p_0^c = p_1^c = 0.004$ day^{-1}. Dashed curve: 10-fold (responsive) travel reduction after time T_c; $p_0^c = 0.004$ day^{-1} and $p_1^c = 0.0004$ day^{-1}. Dotted curve: general (preemptive) 10-fold travel reduction; $p_0^c = p_1^c = 0.0004$ day^{-1}. Note that some of these parameter estimates are conservative and are likely to overestimate R_*^c as a function of T_c—see Methods for discussion.

restrictions, when combined with other sufficiently strong interventions, have been found to substantially curb the COVID-19 epidemic[37].

Second, because R_*^c depends exponentially on T_c, the longer a region waits to implement social distancing measures, the more important it becomes to act without delay. It is important to note, however, that there is no advantage to delaying for even a short time. Immediately implementing social distancing measures as soon as there is evidence of the disease spreading within the region will not only reduce the total amount of time for which such measures must remain in place but will also exponentially reduce the probability of infecting or re-infecting another region. Thus, there is no tradeoff here between health and economics—acting quickly will reduce the duration of both the disease and the economic harm.

Third, practically speaking, all regions within which the virus is spreading must implement and maintain control measures that reduce the local reproductive number to less than one until the virus has been eliminated or contained. If a locale experiences another outbreak of community transmission (e.g. because the virus was re-imported), sufficiently strong interventions should be implemented in a large enough region around that locale so that the per capita rate of travel out of the region is sufficiently low. Per capita travel rates between regions do not have to be zero, but they must be small—precisely how small depends on the parameters in Eq. (5). In the event that some regions "defect" by not containing their outbreaks (thereby putting other regions at risk), other regions must either quarantine travelers from those regions or maintain precautions that keep their local reproductive numbers below one to avoid further outbreaks. Due to the higher

Table 1 Summary of parameters affecting the region-to-region reproductive number.

Param.	Description
r_c	The rate of exponential growth in active infections in region c before control measures are implemented.
τ_c	The time-constant (inverse rate) of exponential decline in active infections after the control measures are implemented.
T_c	The time delay between the start of exponential growth and the implemention of the control measures.
i_0^c	A stochastic measure of the initial foothold that the virus gains prior to exponential growth, with $i_0^c = 0$ corresponding to containment.
p_0^c	The probability rate that an infected individual in region c will travel outside the region before the control measures are imposed.
p_1^c	The probability rate that an infected individual in region c will travel outside the region after the control measures are imposed.

rates of travel between neighboring regions, extra precautions are necessary in a region that has a neighbor with widespread disease transmission.

Finally, like transmission within a region, transmission between regions is an exponential process. At first the number of infected regions is deceptively small, but if $R_* > 1$, this number exponentially grows. If, however, a collection of regions adopts a set of protocols that achieves $R_* < 1$, sustained propagation of the disease between the regions will not be possible, and future outbreaks caused by importations will die out while leaving most regions within the collection unaffected.

Methods

Estimating COVID-19 parameters. The doubling time of the epidemic can vary from location to location and depends on pre-lockdown interventions (see e.g. the change in the growth rate for Italy in Fig. 3). Using the number of confirmed cases in China by date of symptom onset (rather than by date of diagnosis)[38] yields a doubling time of 3.04 days in the period leading up to the Jan. 23 lockdown, which corresponds to $r_c = 0.228$ day^{-1} (Fig. 3). Some studies estimated the doubling time for COVID-19 at approximately 7 days[39,40], but even a 5-day doubling period is implausibly long, given that in various countries, even with some preventative measures, the number of infections increased by far more than a factor of 64 over a 30 day period[41]. Part of the difficulty in estimating the doubling time from the initial period of transmission is that 'super-spreader' events may play an important role in the transmission process[22,26,27]. The presence of super-spreader events indicates that the transmission process may be dominated by relatively uncommon events and that therefore standard statistical approaches may underestimate the rate of spread—and thus overestimate the doubling time—when the total number of cases is still small[42]. (Overestimates of the doubling time lead to underestimates of R_0.) Furthermore, due to heterogeneity from both social and biological factors in transmission rates among various subgroups in a population, the expected growth rate in the very earliest phase of an epidemic depends on the subgroups in which the disease is initially introduced and may not be reflective of the true reproductive number[18].

While $i_c(t)$ is equal to the number of active infections rather than the daily number of new cases, the doubling time for each will be equal during a period of exponential growth, and therefore the exponential rate of growth of the daily number of new cases can be used to estimate r_c. We note that a doubling time of 3 days will likely be a substantial overestimate for regions that, even when cleared of the virus, still maintain some precautionary measures such as mask-wearing, working from home when possible, and avoiding large gatherings.

From this growth rate $r = 0.228$ day^{-1} in China before the Jan. 23 lockdown, the basic reproductive number R_0 (which also varies by location and time) can be calculated for China before Jan. 23 using

$$\frac{1}{R_0} = \int_{-\infty}^{\infty} g(t) e^{-rt} dt \qquad (6)$$

where $g(t)$ is the distribution of generation intervals[23]. Empirically, we generally observe the distribution of serial intervals (the times between the onsets of symptoms in two successive cases in a transmission chain) rather than the distribution of generation intervals (the times between two successive infections in a transmission chain)[43]. The means of the two distributions will, however, be the same. The distribution of generation intervals is affected by non-pharmaceutical interventions; for example, if infected individuals are quarantined, the transmission that is not prevented will more likely occur at shorter generation intervals. In the period before the Jan. 23 lockdown, the mean serial interval (which equals the mean generation interval) was estimated to be 7.8 days, significantly longer than estimates of 5.1 days and 2.6 days for the time periods Jan. 23-29 and Jan. 30-Feb. 13, respectively[44]. As we are estimating R_0 in China before Jan. 23, we focus on the distribution of generation intervals from that period (with the mean of 7.8 days). If all transmission occurred at the estimated mean generation interval of 7.8 days, $R_0 = e^{7.8r} = 5.9$. However, due to the spread of generation intervals, $R_0 = 5.9$ will be a substantial overestimate. R_0 can be underestimated by approximating the distribution of generation intervals $g(t)$ by the distribution of serial intervals

(estimated as a normal distribution with mean 7.8 days and standard deviation 5.2 days[44]), which yields $R_0 = 2.9$. ($R_0 = 2.9$ is an underestimate because the distribution of serial intervals generally has a larger variance than the distribution of generation intervals, and serial intervals, unlike generation intervals, can be negative[43].) Thus, for China before Jan. 23, $2.9 < R_0 < 5.9$, but without a distribution of generation intervals it is hard to be significantly more accurate. We estimate $R_0 = 3.7$, which can be obtained by taking the distribution of generation intervals $g(t)$ to be uniformly distributed between 0 and 15.6 days (15.6 being chosen so that the mean generation interval is 7.8 days).

The values of τ_c that can be achieved depend on the effectiveness of the social distancing measures. The data from China (see Fig. 3) indicate a halving time of as few as 4.78 days is achievable, which corresponds to $\tau_c = 6.9$ days. However, as a more conservative estimate, we use $\tau_c = 15$ days from Italy's data (Fig. 3), which exhibited a particularly slow decline in cases compared with most other European countries[41]. (The halving time of the daily number of new cases will equal the having time of the number of active infections $i_c(t)$, which is why τ_c can be estimated from this data.)

$\mathbb{E}[i_0^c]$ is the expected "effective" number of people an infected traveler will infect while visiting region c, taking into account containment efforts. For instance, if the outbreak is contained such that exponential growth never occurs, the effective number of people infected by the traveler is zero, even if the traveler did infect some individuals in region c. We estimate that in the absence of any mitigating policies (mask-wearing, testing, etc.), $\mathbb{E}[i_0^c] \sim R_0 \sim 3.7$; the degree to which $\mathbb{E}[i_0^0]$ differs from R_0 depends on how likely a typical traveler is to transmit the virus relative to a typical resident, as well as on the effectiveness of contact tracing and other containment efforts. For regions in which preemptive measures (e.g. mask-wearing, avoiding large gatherings, etc.) are taken, we expect that $\mathbb{E}[i_0^c] \sim R_0 \sim 3.7$ is a substantial overestimate since there is a far greater chance that an infected traveler will not spark an uncontained outbreak (and thus i_0^c will be 0). Good testing and contact tracing policies can also substantially increase the chance that $i_0^c = 0$ (thereby reducing $\mathbb{E}[i_0^c]$) by reducing the probability of community transmission. And a rigorous enough quarantine policy at the border of a region may reduce $\mathbb{E}[i_0^c]$ to nearly zero by preventing infected travelers from even having a chance to infect individuals in the region.

The value of p_c depends on the frequency of travel out of region c. As previously noted, there is some choice in how to partition a population into a collection of regions. In general, the larger the regions, the lower the frequency of per capita travel out of them but the more homogeneous the application of the social distancing measures. Since p_c is smaller for larger regions and N_c is not strongly affected by the size of regions, R_* will be lower if larger regions are chosen, but at the cost of the social distancing measures being applied over larger areas for each new outbreak.

Regions do have to be large enough so that transmission between regions can be treated as a first-order (linear) perturbation to a system in which most spread occurs within regions. Thus, how small the regions can be depends on the extent to which travel between them can be reduced. A region could be as small as, for example, a neighborhood if the neighborhood was willing to take measures so as to greatly reduce contact with people outside of it. (We note that the regions within any given collection can differ in both geographic size and in population from one another.)

Although p_c will depend on the size of region c as well as the travel behavior of individuals within that region, we still wish to get an estimate of a plausible value for p_c. Considering a collection of regions within the U.S. that are large enough such that travel between the regions is predominantly by flight yields a per capita travel rate of 0.004 flights out of a region per person per day. This estimate is obtained by dividing the 1.01 billion total passengers traveling by plane to, from, or within the U.S. in 2018[45] by the 2018 U.S. population and the number of days in 2018, and then also dividing by 2 so that only flights out of and not into a region are counted. Using this estimate for p_0^c assumes that the probability that an infected individual will travel equals that of the general public. Travelers may on average have more social contacts than the general public or may become infected because of their travels, which could mean that infected individuals are more likely to be traveling than the general public. On the other hand, an individual who has COVID-19 symptoms or has tested positive for COVID-19 may be less likely to travel than the general public. It should also be noted that, compared to 2019 numbers, U.S. air travel in April and May of 2020 was down by approximately a

factor of 10 (the effects of a 10-fold reduction in travel are shown in Fig. 4) and U.S. air travel in June was down by approximately a factor of 5[46].

Data availability
The data analyzed in Fig. 3 can be found in refs. [38,47].

References

1. World Health Organization. Novel coronavirus – China. www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/ (2020).
2. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report, 72. https://apps.who.int/iris/bitstream/handle/10665/331685/nCoVsitrep01Apr2020-eng.pdf (2020).
3. Watts, D. J., Muhamad, R., Medina, D. C. & Dodds, P. S. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proc. Natl Acad. Sci. USA* **102**, 11157–11162 (2005).
4. Rauch, E. M. & Bar-Yam, Y. Long-range interactions and evolutionary stability in a predator-prey system. *Phys. Rev. E* **73**, 020903 (2006).
5. Colizza, V., Pastor-Satorras, R. & Vespignani, A. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282 (2007).
6. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* **251**, 450–467 (2008).
7. Ajelli, M. et al. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* **10**, 190 (2010).
8. Balcan, D. & Vespignani, A. Invasion threshold in structured populations with recurrent mobility patterns. *J. Theor. Biol.* **293**, 87–100 (2012).
9. Tanaka, G., Urabe, C. & Aihara, K. Random and targeted interventions for epidemic control in metapopulation models. *Sci. Rep.* **4**, 1–8 (2014).
10. Ball, F. et al. Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**, 63–67 (2015).
11. Hasell, J. Which countries have protected both health and the economy in the pandemic? https://ourworldindata.org/covid-health-economy (2020).
12. Goolsbee, A. & Syverson, C. Fear, lockdown, and diversion: comparing drivers of pandemic economic decline. Working Paper 27432, National Bureau of Economic Research (2020).
13. Aum, S., Lee, S. Y. T. & Shin, Y. COVID-19 doesn't need lockdowns to destroy jobs: the effect of local outbreaks in Korea. Working Paper 27264, National Bureau of Economic Research (2020).
14. Brzezinski, A., Kecht, V. & Van Dijcke, D. The cost of staying open: voluntary social distancing and lockdowns in the US. Working Paper 910, University of Oxford, Department of Economics (2020).
15. Kardar, M. Statistical Physics of Fields (Cambridge University Press, 2007).
16. Bar-Yam, Y. From big data to important information. *Complexity* **21**, 73–98 (2016).
17. Siegenfeld, A. F. & Bar-Yam, Y. An introduction to complex systems science and its applications. *Complexity* **2020**, e6105872 (2020).
18. Diekmann, O., Heesterbeek, J. A. P. & Metz, J. A. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* **28**, 365–382 (1990).
19. Ball, F., Mollison, D. & Scalia-Tomba, G. Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89 (1997).
20. Colizza, V. & Vespignani, A. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.* **99**, 148701 (2007).
21. Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19) frequently asked questions, accessed 28 June 2020. https://www.cdc.gov/coronavirus/2019-ncov/faq.html (2020).
22. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
23. Wallinga, J. & Lipsitch, M. How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B: Biol. Sci.* **274**, 599–604 (2007).
24. Ferguson, N. M. et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team (2020).
25. Siegenfeld, A. F., Taleb, N. N. & Bar-Yam, Y. What models can and cannot tell us about COVID-19. *Proc. Natl Acad. Sci. USA* **117**, 16092–16095 (2020).
26. Riou, J. & Althaus, C. L. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. *Eurosurveillance* **25**, 2000058 (2020).
27. Endo, A. et al. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020).
28. Ahmed, W. et al. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: a proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci. Total Environ.* **728**, 138764 (2020).
29. Kogan, N. E. et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real-time. arXiv:2007.00756 (2020).
30. Bajardi, P. et al. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PLoS ONE* **6**, e16591 (2011).
31. Poletto, C. et al. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Eur. Surveill.* **19**, 20936 (2014).
32. Brownstein, J. S., Wolfe, C. J. & Mandl, K. D. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the United States. *PLoS Med.* **3**, e401 (2006).
33. Germann, T. C., Kadau, K., Longini, I. M. & Macken, C. A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl Acad. Sci. USA* **103**, 5935–5940 (2006).
34. Epstein, J. M. et al. Controlling pandemic flu: the value of international air travel restrictions. *PLoS ONE* **2**, e401 (2007).
35. Wells, C. R. et al. Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proc. Natl Acad. Sci. USA* **117**, 7504–7509 (2020).
36. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. eaba9757 (2020).
37. Kraemer, M. U. G. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. eabb4218 (2020).
38. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* (2020).
39. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
40. Li, Q. et al. Early transmission dynamics in Wuhan, China of novel coronavirus–infected pneumonia. *N. Engl. J. Med.* (2020).
41. Worldometer. COVID-19 coronavirus outbreak, accessed 28 August 2020. https://www.worldometers.info/coronavirus/.
42. Taleb, N. N. How much data do you need? An operational, pre-asymptotic metric for fat-tailedness. *Int. J. Forecast.* **35**, 677 – 686 (2019).
43. Park, S. W. et al. Cohort-based approach to understanding the roles of generation and serial intervals in shaping epidemiological dynamics. medRxiv: 2020.06.04.20122713 (2020).
44. Ali, S. T. et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science*. eabc9004 (2020).
45. U.S. Department of Transportation. 2018 traffic data for U.S. airlines and foreign airlines U.S. flights, accessed 12 March 2020. https://www.bts.dot.gov/newsroom/2018-traffic-data-us-airlines-and-foreign-airlines-us-flights.
46. U.S. Department of Homeland Security Transportation Security Administration. TSA checkpoint travel numbers for 2020 and 2019, accessed 28 June 2020. https://www.tsa.gov/coronavirus/passenger-throughput.
47. Istituto Superiore di Sanita. COVID-19 integrated surveillance: key national data, accessed 21 April 2020. https://www.epicentro.iss.it/en/coronavirus/sars-cov-2-integrated-surveillance-data.

Author contributions
A.F.S. and Y.B.-Y. developed the concepts, performed the analyses, and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.F.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.