Modeling the dynamics of COVID19 spread during and after social distancing: interpreting prolonged infection plateaus

Natalia L Komarova and Dominik Wodarz

Supplementary Information

Contents

1 The standard SIR ODE model

1.1 Model formulation and elements of analysis

Denote by x is the number of susceptibles, y the number of infecteds, and z is the number of removed. Consider the standard SIR system,

$$\dot{x} = -\frac{\beta xy}{N}, \tag{1}$$

$$\dot{y} = \frac{\beta xy}{N} - \gamma y, \tag{2}$$

$$\dot{z} = \gamma y \tag{3}$$

with initial conditions

$$x(0) = x_0 N, \quad y(0) = y_0 N, \quad z(0) = z_0 N.$$

Here β is the rate of infection and γ the rate of removal. Let us denote

$$r_0 = \frac{\beta}{\gamma}.$$

Let us define function u as follows:

$$\gamma t = \int_{u_0}^{u} \frac{d\xi}{\xi(-r_0 - \ln \xi + C\xi)}, \tag{4}$$

where we used the following combination of constants:

$$C = r_0 x_0 e^{r_0 z_0}. \tag{5}$$

This constant is important for defining the behavior of the solution. Note that the values x_0 and z_0 belong to the simples, $x_0 + z_0 \leq 1$, $x_0, z_0 \geq 0$. The maximum value of C corresponds to $x_0 = 1/r_0$ and $z_0 = 1 - x_0$, and igiven by $C_{max} = e^{r_0 - 1}$.

The solution of the SIR model is given in terms of u, see [**?**]:

$$x(t) = \frac{N}{r_0} C u, \tag{6}$$

$$y(t) = \frac{N}{r_0} \left(\ln u - Cu + r_0 \right), \tag{7}$$

$$z(t) = -\frac{N}{r_0} \ln u. \tag{8}$$

Note that only values of u that satisfy $u \leq 1$ give biologically relevant solutions. The constant u_0 can be found e.g. from the initial condition applied to equation (8),

$$u(0) = u_0 = e^{-r_0 z_0}. \tag{9}$$

In particular, if $z_0 = 0$ we have $u_0 = 1$, and if $z_0 > 0$ we have $u_0 < 1$.

Note that the number of infected individuals will increase initially if and only if

$$x_0 r_0 > 1, \tag{10}$$

that is, an epidemic will develop only if this condition is satisfied.

We are interested in three measures of the epidemic: the final epidemic size (defined as $z_\infty = \lim_{t\to\infty} z(t)$), the maximum value of $y(t)$, y_{max}, and the width of the peak of $y(t)$ (the latter two quantities are only defined if condition (10) holds)).

The peak of infection. To find y_{max}, we differentiate (7), to see that $y' = 0$ if $u' = 0$ or if

$$u = C^{-1}. \tag{11}$$

Under the former choice, we can calculate dt/du from (4) and invert it to find

$$\frac{u'}{\gamma} = u(-r_0 - \ln u + Cu) = 0,$$

which yields either $u = 0$ (which is not biologically relevant) or $y = 0$ from (7). Under choice (11) we have from (7)

$$y_{max} = N\left(-\frac{\ln C}{r_0} - \frac{1}{r_0} + 1\right).$$

The time at which this maximum is reached is found from equation (4) evaluated at (11).

The width of the peak. Using equation (7), we can find the 2nd derivative of $y(t)$ and substitute equation (11), such that it corresponds to $y' = 0$ (the peak). We obtain

$$y'' = \frac{-N\gamma^2}{r_0}(1 - r_0 + \ln C)^2.$$

To estimate the width of the peak, let us approximate it as a parabola with a maximum value of y_{max}. Finding the x-intercepts, we get an estimate of the width as

$$t_w = 2\sqrt{\frac{y''}{y_{max}}} = \frac{2}{\gamma\sqrt{r_0 - 1 - \ln C}}.$$

The final epidemic size. To find the final epidemic size, we consider equation (7), set $y(t) = 0$ and solve for u. Then we have

$$z_\infty = \lim_{t\to\infty} z(t) = -\frac{N}{r_0}\ln u_\infty, \quad \ln u_\infty - Cu_\infty + r_0 = 0. \tag{12}$$

Note that the above equation for u_∞ can be solved graphically by finding intersections of $\ln u$ and $-r_0 + Cu$ with $0 < u < 1$. Note that if $C < r_0$, there is a single intersection with $u \leq 1$. When $C = r_0$, a second point of intersection appears with $u = 1$. If $r_0 \leq C < e^{r_0 - 1} = C_{max}$, two intersections exist, and finally there is a single intersection at $u = 1/C$ when $C = C_{max}$. In the case when there are two intersection points, u_∞ corresponds to the smaller one (because this corresponds to the first time when y becomes zero). Also, we can see that the lower intersection value, u_∞, increases with C.

We can see that all three measures of the epidemic depend on the initial conditions through quantity C, equation (5). In particular, y_{max} decreases with C, t_w increases with C, and the final epidemic size, z_∞, decreases with C (because u_∞ increases). In other words, as C increases, the peak of infection becomes lower and flatter, and the final epidemic size decreases. The maximum value of C is reached at $x_0 = 1/r_0, z_0 = 1 - x_0$, and the minimum value corresponds to $x_0 = 1/r_0, z_0 = 0$.

1.2 Using the SIR ODEs to model intervention by social distancing

One can use this model to describe the effect of social distancing in the following way. Let us use $N = 1$ and $\gamma = 1$, which is equivalent to rescaling the population size by N and time by γ. Suppose the epidemic starts at $t_0 = 0$ and is characterized by a certain initial condition $(x(0) = x_0, z(0) = z_0)$ and a certain β, which we denote by β_1. Then, at time $t_1 > 0$, social distancing is implemented, which is manifested in a reduction of β below its original value, $\beta_2 < \beta_1$. Finally, at time $t_2 > t_1$, social distancing is relaxed. For simplicity, we will assume that it is relaxed completely, meaning that β returns to its original value, β_1.

This process can be described by using the characteristic lines,

$$r_0 x e^{r_0 z} = C.$$

In figure S1(a), such lines for $\beta_1 = 2$ are shown by black contour plots in the $x-z$ simplex. Let us suppose that the epidemic starts with a particular initial condition ($x_0 = 0.999, z_0 = 0$, see the lower blue line in figure S1(a)). Then, in the absence of intervention, the trajectory of the epidemic in terms of x, z coordinates will trace the corresponding characteristic upwards (increasing

the z-coordinate), starting from its initial point (in our case corresponding to its x-intercept), and ending at the point where the characteristic line crosses the diagonal, $x + z = 1$. The resulting z value is the final epidemic size, which can be found as the solution of $r_0(1-z)e^{r_0 z} = C$ (this is he same as solution (12)).

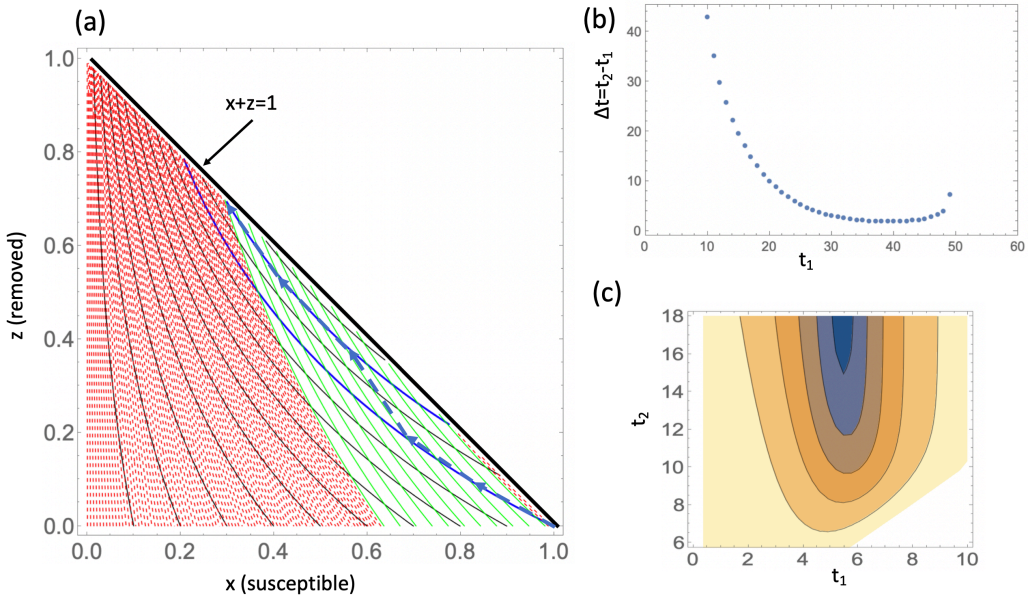

Figure S1: SIR ODE model used to study intervention strategies. (a) For particular parameters values, epidemic trajectories (corresponding to different initial conditions) are shown as black lines in the $x - z$ simplex. Upon intervention, the value of infectivity parameter decreases from β_1 to β_2, resulting in a new family of characteristic lines shows by red and green lines. The lower blue trajectory corresponds to the example of the epidemc illustrated in the main text and in panels (b,c). The upper blue trajectory corresponds to the target final epidemic size of panel (b). The green trajectories are intervention strategies that are capable of achieving the target epidemic size, while red ones are not. The blue arrows indicate an example of intervention that results in the target final epidemic size. (b) The length of the intervention as a function of the time t_1 of intervention start. (c) The final epidemic size as a function of intervention start and stop times, t_1 and t_2. Darker colors correspond to lower values. Parameters are $\beta_1 = 2, \beta_2 = 1.1, x_0 = 0.999, z_0 = 0$.

Intervention by social distancing corresponds to lowering β to its value β_2. The corresponding trajectories are denoted by green and red lines. They are steeper than the black lines. Intervention would correspond to hopping from the blue line to one of the green (or red) lines, and will result in a smaller finite epidemic size. In the scenario of interest here, after a finite

period of intervention that starts at time t_1, the intervention is stopped at time t_2, which corresponds to a return to one of the characteristic lines of the original family. One example is shown by blue arrows in figure S1(a).

In panel (b) we solve a specific optimization problem. Given an epidemic (the lower blue line of panel (a)), let us suppose that we have a target final epidemic size that is smaller than that of the original trajectory. In the example here we have $z_\infty^1 = 0.797$, and we chose $z_\infty^2 = 0.71$. The target epidemic size corresponds to the second blue trajectory of panel (a). Suppose further than β_2 is fixed. Is there a way to achieve this result with a minimum intervention time? In other words, the goal is to minimize the quantity $\Delta t = t_2 - t_1$. The target epidemic size can be reached by hopping on one of the green trajectories and then hopping onto the second blue line (relaxing the intervention) some time later. There is an infinite number of ways of doing this (one of them is shown). Note that there are some restrictions. For example, if we start intervention too early (one of the red dashed lines that start near the corner $x_0 = 1, z_0 = 0$), the target epidemic size cannot be achieved. If we never relax intervention (which is not a practical solution), then the final epidemic size will be relatively small (smaller than target), but a relaxation of the intervention at any finite time will lead to a finite epidemic size larger than target. Similarly, if we wait too long (the red trajectories in the left half of the graph), again, the finite epidemic size will be larger than target. Only one of the trajectories in the green region can result in the desired finite epidemic size. Panel (b) shows that there is an optimal such trajectory that minimizes the time of intervention. Again starting too early or too late will lead to a very long intervention time.

In panel (c) we study the final epidemic size as a function of the variables t_1 and t_2. Again, starting with the lower of the blue trajectories, we vary t_1 and t_2, but this time we do not constrain what final trajectory we choose, but simply relax intervention at time t_2, which puts us on one of the trajectories in the black family of panel (a). In panel (c), a contour plot of the final epidemic size is presented, where darker shading corresponds to lower values of z_∞. We can see that increasing t_2 will always result in a decrease of z_∞ (the longer the intervention, the lower the final epidemic size). The dependence on t_1 however is non-monotonic, and, as already hinted by the previous results, there is an optimal time to start the intervention.

1.3 Intervention in the absence of relaxation

To investigate the dynamics of a single stage intervention, we take $t_2 \to \infty$, and find the characteristics of the infection peak and the final epidemic size. In figure S2, we vary t_1 and look at the peak of infection (panel (a)) and the number of removed individuals (panel (b)). We can see that delaying intervention, while resulting in a faster overall infection decay, leads to a higher infection peak and a higher finite epidemic size.

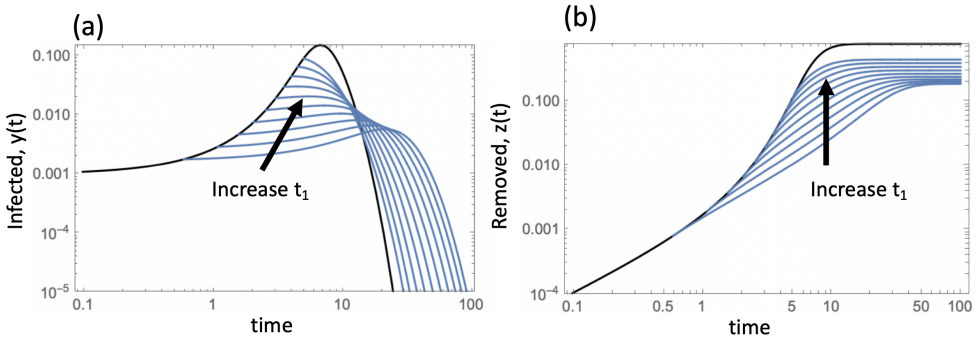

Figure S2: Intervention in the absence of relaxation. The number of infected (a) and removed (b) individuals are shown as functions of time. The black lines correspond to no intervention. The blue lines correspond to intervention starting at different times, t_1. The parameters are $\beta_1 = 2; \beta_2 = 1.1; x_0 = 0.999; z_0 = 0$.

1.4 Dependence on the transmission rate during the intervention

Here we explore the effect of the reduction of β on the infection peak, y_{max}, and the final epidemic size z_∞. In figure S3 we perform intervention simulations for two values of β_2, with the higher value corresponding to the solid lines and the lower value to dashed lines. For comparison, the trajectory in the absence of intervention is shown in blue. The number of infecteds (y) is plotted as a function of t, and the intervention time-period ($t_1 \leq t < t_2$) is shown in read, while the relaxation period ($t \geq t_2$) is shown in black. Trajectories for six different values of t_2 are shown. We can see that if the intervention time is very short, the infection peak after the relaxation is nearly

as high as that in the absence of intervention, and the two values of β_2 lead to nearly identical peak heights.

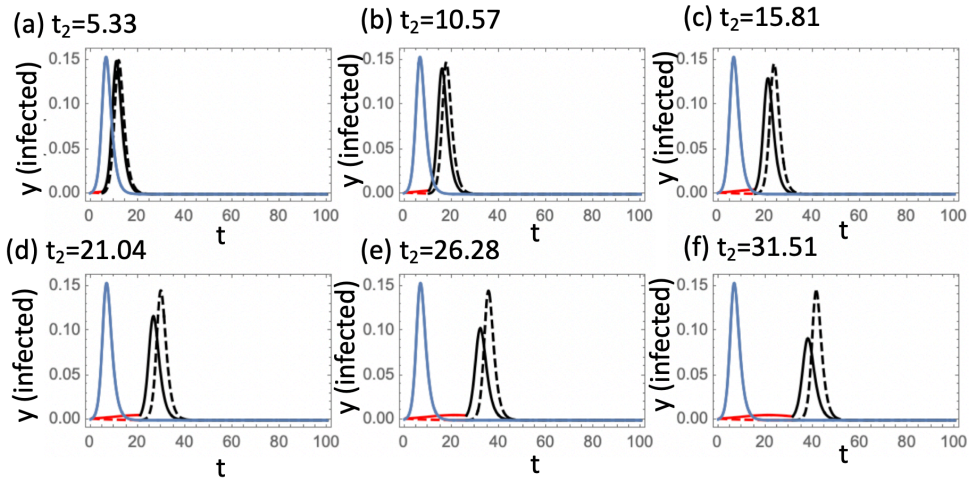

Figure S3: Dependence on the severity of intervention. The numbers of infected individuals are shown as function of time; different panels correspond to different times of intervention relaxation (t_2), while t_1 is the same for all panels. Blue lines correspond to no intervention (and also the curve before intervention). Red is the intervention period and black is after the intervention stops. Solid curves correspond to higher and dashed curves to lower values of β_2. The parameters are $\beta_1 = 2, \beta_2 = 1.1$ (solid), $\beta_2 = 0.95$ (dashed), $x_0 = 0.999, z_0 = 0$.

As t_2 increases, we can see that the peak becomes lower, especially the peak corresponding to the higher value of β_2. The peak corresponding to the lower β_2 does not decrease much with t_2 and is always higher than that for the higher β_2. In other words, the higher the "intervention" β value, the lower the peak.

Figure S4 varies the values of β_1 while keeping t_1 and t_2 fixed. The nine panels correspond to nine different values of β and present the time-series of infecteds, in the absence (blue) and in the presence (red and black) of intervention. We can see that as β_2 increases, the peak of infection decreases. Panel (b) shows the peak height (y_{max}) during the relaxation phase (the "relaxation peak") as well as the peak height during the intervention phase (the "intervention peak"), as functions of β_2. The height of the relaxation peak becomes negligible for high values of β_2, because as $\beta_2 \to \beta_1$, the intervention becomes nonexistent, but instead, the infection peaks during the intervention phase. The regime where the intervention peak is higher than

the relaxation peak corresponds to intervention that is not effective. The final epidemic size (z_∞) as a function of β_2 is shown in panel (c). The final epidemic size, after first dropping for intermediate values of β_2, eventually increases back to the value that characterizes the epidemic without intervention. So, to conclude, there is an optimal, intermediate, value of β_2 that corresponds to the lowest infection peak and the lowest finite epidemic size.

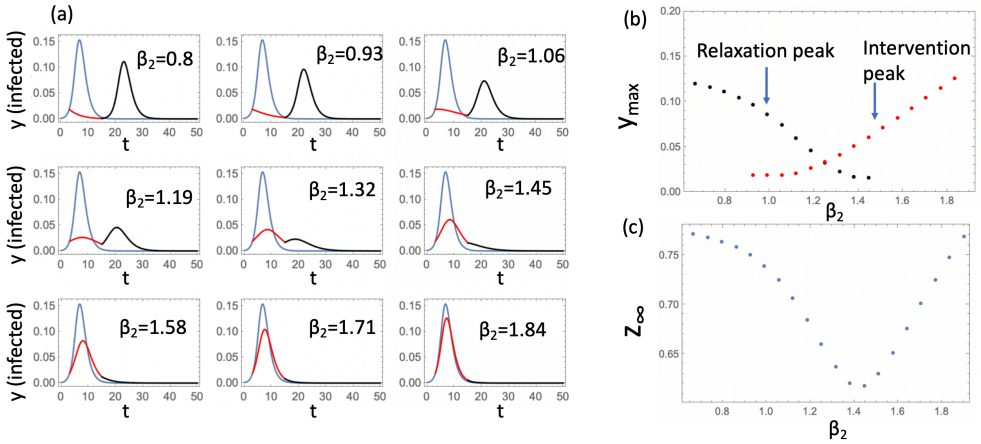

Figure S4: Varying the severity of intervention. (a) The number of infecteds as a function of time, for different values of intervention infectivity parameter. Blue lines correspond to before intervention (and in the absence of intervention), red curves during the intervention, and black after the intervention. (b) The hight of the peaks of the infecteds as functions of β_2 for the relaxation (black) and intervention (red) peaks. (c) The final epidemic size as a function of β_2. The parameters are $\beta_1 = 2; x_0 = 0.999; z_0 = 0; t_1 = 3; t_2 = 15$.

2 Growth laws in the context of edge removal for different types of networks

2.1 Edge removal and the plateauing of the infections

Depending on the type of random network, infection spread can happen differently. Moreover, as edges are removed, the kinetics of spread will change, and this change also depends on the underlying type of network. To this end, we have investigated the spread of infection in 4 types of random graphs, listed below. They were constructed by using the built-in procedure RandomGraph in *Mathematica*, using the appropriate distributions:

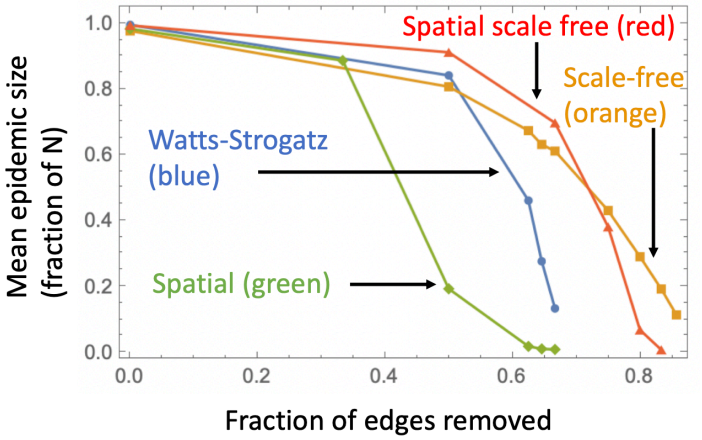

Figure S5: The mean final epidemic size (measured as the fraction of all nodes that became infected) as a function of the proportion of edges removed, for 4 different network types. The epidemic size is obtained by averaging among the runs that resulted in 200 or more infected individuals, from the total of 100 runs per condition. $P_{inf} = 10^{-4}$, $P_{remove} = 1.5 \times 10^{-4}$.

- Small world (Watts-Strogarz) networks, constructed from a circular graph with 6 nearest neighbors on each side, and the probability of rewiring $p = 0.1$.
- Scale-free (Barabasi-Albert) networks with the minimum of 6 connections per node.
- Spatial networks built according to the spatial distribution for graphs with N vertices uniformly distributed over the unit square and edges between vertices that are at distance at most $r = 0.02$.
- Spatial scale free ("hybrid") networks which were constructed as a cross between spatial and scale-free networks. To build such a network we created an adjacency matrix W_1 for the spatial network exactly as above, and an adjacency matrix for a scale-free network (but unlike the network above, we assumed the minimum of one edge per node). Then the hybrid graph used the adjacency matrix $W_1 + W_2$.

In all the numerical examples, we used networks with $N = 10^5$ nodes. For each network, we performed 100 simulations, where we start with one randomly infected agent and stop when there are no more infected individuals.

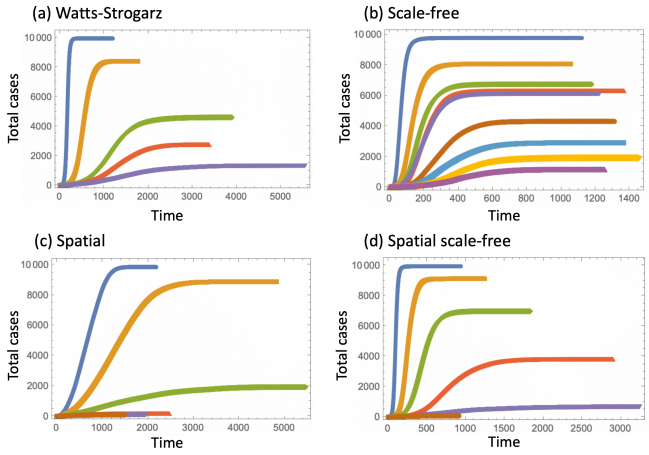

Figure S6: Same as figure S5, with more detail. The mean total number of cases is plotted as a function of time, for the 4 networks. In each panel, different lines correspond to the different proportion of edges removed; the lower the curve the lower the fraction of retained edges.

Next, we removed a certain fraction of randomly chosen edges and performed similar simulations with such "reduced" networks, to study the effect of edge removal (or social distancing) on the epidemic spread. Figure S5 shows the mean epidemic size (as the fraction of the $N = 10000$ nodes) developing on the four networks, when we remove different fractions of edges. We observe that there is a monotonic dependence of the final epidemic size on the density of the remaining network: the more edges are removed, the smaller the resulting size. Also, different networks demonstrate different degrees of robustness against edge removal. In particular, for the chosen network parameters, the spatial network is most "vulnerable" among the four and results in extinction for the smallest proportion of edges removed. More details can be seen in the time-series presented in figure S6, where for each network, we plot the mean total number of cases as a function of time, for different fractions of removed connections. The more corrections are removed, the slower the epidemic growth, the lower the final epidemic size, and the fewer runs result in an epidemic (the latter trend is not shown here, as the averaging was performed conditioned on the epidemic rising above 200 cases).

Figure S7 displays typical stochastic runs for both original (uncut) networks and their counterparts with a fraction of edges randomly removed. We can see that all four original networks are characterized by the same pat-

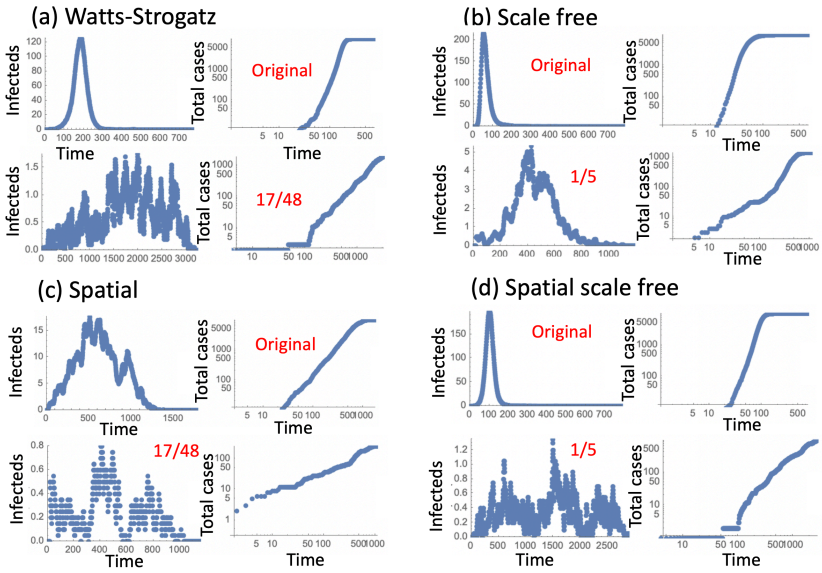

Figure S7: The effect of edge removal in 4 types of network. For each of the networks, Watts-Strogatz (a), Scale-free (b), Spatial (c) and Spatial scale free (d), typical individual stochastic runs are presented. For each panel (a-d), the first row of subpanels corresponds to the original (uncut) network, and the second row to the network that was obtained from the original one by randomly removing the fraction of edges that is indicated in red. Both the number of infected individuals and the total number of cases are plotted as functions of time. $P_{inf} = 10^{-4}, P_{remove} = 1.5 \times 10^{-4}$.

tern of infection time-series ($y(t)$): the number of infecteds rises and then falls relatively sharply. A different pattern is observed in three out of four cut networks: the number of infecteds does not experience a sharp hump but instead remains relatively flat (nearly constant), entering a period of plateau-like behavior, before eventually decreasing. This can be seen in panels (a), (c), and (d), but not in panel (b) which corresponds to scale-free networks.

To study the formation of plateaus systematically, we performed large numbers of computer simulations with different types of networks and different degrees of edge removal. Figure 3 of the main text shows results for spatial networks. In figure S8 we also provide results for scale free (Barabasi-Albert) and hybrid networks. In these simulations, the networks with a fraction of connections randomly removed were used after the epidemic rose to size 100 individuals. While the spatial network exhibits a well defined plateau dynamics, the scale-free network shows a steep rise and drop in the number of infections, resembling more the ODE dynamics. As expected, the hybrid

network shows intermediate behavior.

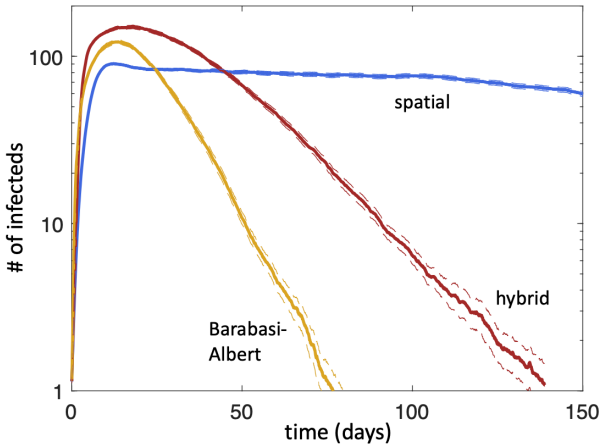

Figure S8: The mean number of infected individuals as a function of time for an epidemic on a spatial, Barabasi-Albert, and hybrid network ($N = 10^5$, constructed as described in this section), with edges randomly removed. For the spatial network, 1/2 connections are kept; for Barabasi-Albert, 1/9 of the connections are kept; for the hybrid network, 1/5 of the connections are kept. The other parameters are $P_{inf} = 10^{-4}, P_{remove} = 1.5 \times 10^{-4}$. Averages over 200 runs are plotted together with standard errors (dashed lines). $P_{inf} = 10^{-4}, P_{remove} = 1.5 \times 10^{-4}$.

2.2 Parameter dependence of plateau dynamics

The length of the plateau phase following intervention depends on several parameters, including the degree to which the network's connections are removed, the timing of the intervention and the relative removal/infection rates. In figure S9 we explore the plateau region in spatial networks under varying parameters. Panel (a) shows the existence of plateau under milder distancing measures compared to figure 3F of the main text. The colored lines correspond to different epidemic sizes when the distancing measures are first implemented; these take values $50, 100, 200, \ldots, 700$. As in figure 3F of the main text we observe that an earlier intervention start leads to a flatter plateau, but weaker distancing measures lead to a shorter plateau compared to the stronger measures of figure 3F.

Panel (b) on figure S9 shows that the length of the plateau depends on the removal rate of the epidemic (relative to the infection rate). Higher removal rates correspond to smaller epidemic sizes and shorter plateaus

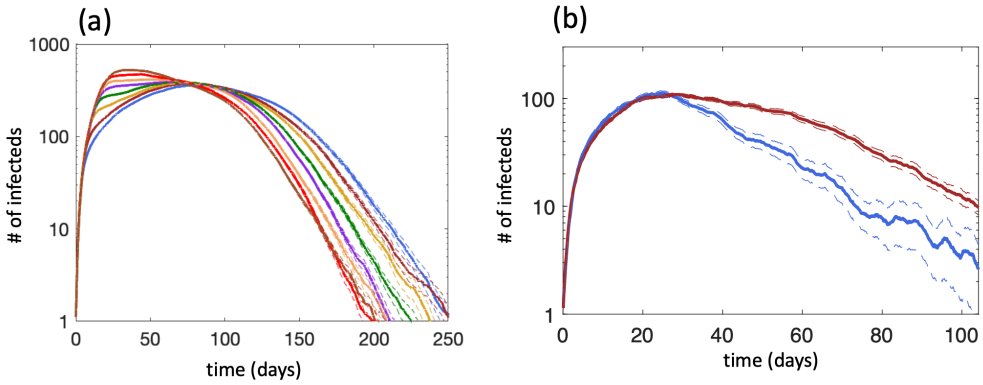

Figure S9: Examples of plateau dynamics on spatial networks with different parameters. (a) Similar as figure 3F of the main text, except the network upon social distancing retains 2/3 of its edges. Social distancing is initiated when the number of infected individuals reached size 50, 100, 200, 300, 400, 500, 600, 700. Parameters are as in figure 3F of the main text. Averages over 900 runs are plotted. (b) The number of infecteds as a function time for social distancing where 1/2 (blue) and 2/3 (red) of edges are retained, and the removal rate is higher $P_{removal} = 3 \times 10^{-4}$. Social distancing was initiated when the number of infected individuals reached 200. The rest of the parameters are as in figure 3F. Averages over 200 runs are plotted together with standard errors (dashed lines).

2.3 Social distancing and relaxation dynamics on networks

Next we turn our attention to implementing social distancing measures followed by subsequent relaxation. In the models presented here, social distancing is modeled as removing random edges in the social network. Relaxation of social distancing amounts to restoring the original network completely. Figure 4 of the main text shows results for spatial networks. In these simulations, the starting conditions for social distancing are always the same: for each simulation, a given fraction of the edges is removed when the population of infecteds reaches a fixed number (100 individuals). What varies among simulations in panels of main figure 4(c) is the time of relaxation. We observed two interesting trends for spatial networks: (1) if the length of the social distancing stage is extended, the second wave of the epidemic becomes less pronounced, but this effect is only significant if relaxation occurs after the plateau starts lowering down; (2) stronger social distancing measures (more edges removed) lead to a lower second peak and a lower final epidemic size after relaxation. The latter can be observed in figure S10, where we plot

the number of removed individuals for two degrees of social distancing, with red curves correspond to a large fraction of edges randomly removed.

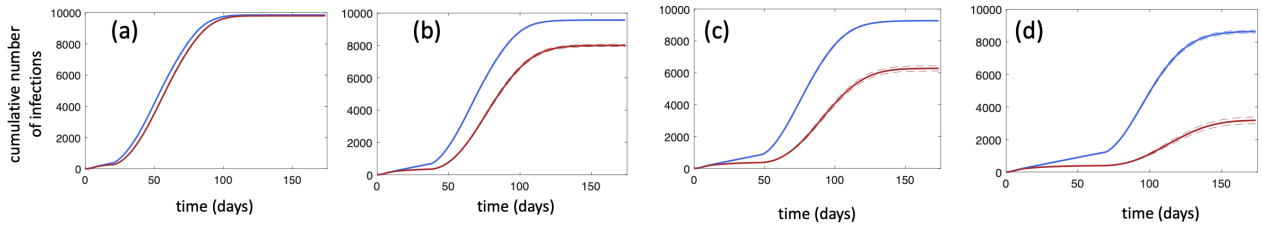

Figure S10: Additional panel for main figure 4: varying the intensity of distancing and the timing of relaxation for a simulated epidemic on a spatial network. The cumulative number of infections is plotted against time, for relaxation times (a) 21, (b) 38, (c) 49, and (d) 69 units. The degree of distancing: 1/2 edges retained (blue) and 17/48 edges retained (red). Other parameters are as in figure 4 of main text.

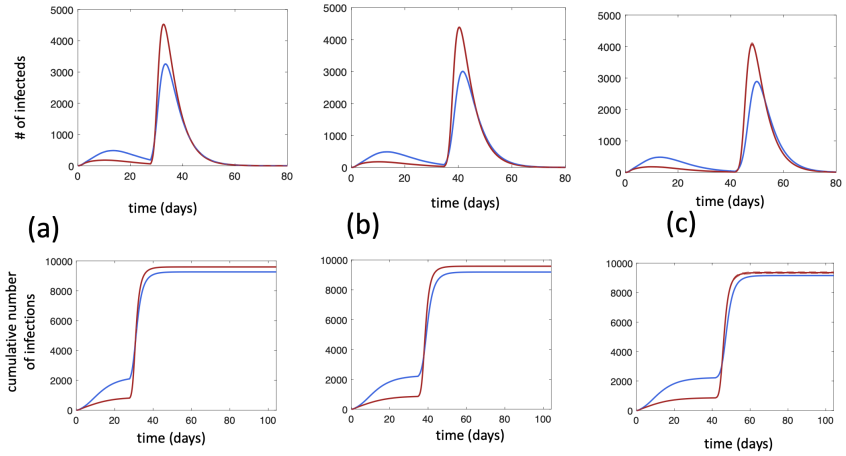

Figure S11: Varying the intensity of distancing and the timing of relaxation for a simulated epidemic on a Barabasi-Albert network ($N = 10000$, minimum 6 connections per node), with edges randomly removed: only 1/6 of all edges are left (blue) and only 1/9 of all edges are left (red). The mean number of infected individuals (top row) and cumulative number of infected individuals (bottom row) are plotted against time, for relaxation times (a) 28, (b) 35, and (c) 42 units. Averages over 400 runs are plotted together with standard errors (dashed lines). Other parameters are as in figure S8.

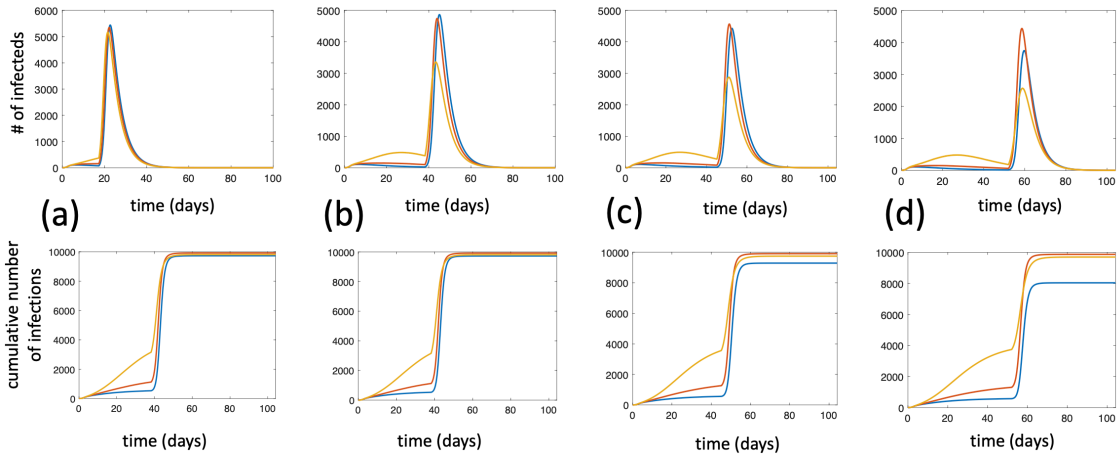

Figure S12: Same as in figure S11, but for a hybrid network. Blue: 1/6 edges retained; red: 1/5 edges retained, yellow: 1/4 edges retained. The relaxation times are (a) 17.4, (b) 38, (c) 45, and (d) 52 units. Averages over 400 runs are plotted together with standard errors (dashed lines). Other parameters are as in figure S11.

We next compare these results with those obtained for scale free (Barabasi-Albert, figure S11) and hybrid (figure S11) networks. In the case of Barabasi-Alber networks (figure S11), we observe that (1) as the length of social distancing is extended, the second infection wave becomes somewhat lower (and the final epidemic size smaller), but without the threshold-like effect of the spatial networks, because there is no plateau during the distancing phase. Observation (2) above however reverses compared to the spatial networks: stronger social distancing measures (more edges removed) lead to a higher second peak and a larger final epidemic size after relaxation. This behavior is similar to that obtained for the ODE model of infection.

Finally, hybrid networks exhibit an interesting cross between the two types of behavior characteristic of spatial and scale-free networks, see figure S12. We have run simulations for the mild (yellow), intermediate (red), and strict (blue) social distancing measures. The resulting second wave size depends on the time of relaxation, and the changes are non-monotonic in terms of the fraction of edges removed. Namely, for relatively short durations of the social distancing stage (panels (a) and (b)), we observe that the behavior is similar to that of scale-free networks (and the ODE models): stricter social distancing measures (a smaller fraction of edges retained) lead to a larger

second wave of infection. This pattern, however, changes as the duration of social distancing increases. In the last two panels of figure S12 we can see that the strongest social distancing measure result in second highest second wave (out of the 3 scenarios plotted), so the system starts molding its behavior to acquire more properties of the spatial system. The final epidemic size generally exhibits more complex behavior, because it is a phenomenon with "memory": it reflects both the social distancing stage of the epidemic where the smallest number of infected individuals corresponds to the most severe distancing, and the relaxation dynamics, where depending on the timing, different networks may correspond to the highest spike. For the parameters chosen for this figure, the strictest distancing corresponds to the lowest final epidemic size, and intermediate distancing leads to the highest final epidemic size. In general, depending on the relative size of spatial and scale-free network contributions used when constricting the network, one may expect different outcomes as the balance shifts and spatial and no-spatial effects trade-off differently.

2.4 The SIS dynamics

In order to study the role of immunity in plateau dynamics, we have simulated an epidemic in the absence of immunity, see figure S13. In this case, no plateau is observed, see discussion in the main text.

3 Growth patterns in the different states of the US

The goal of this part of the study is to find patterns in the epidemic development in different locations. For this purpose, we chose to work with different states of the US. Data were available from `https://covidtracking.com/`, where the total number of confirmed cases is available every day starting on 01/22/20 in Washington, and on 03/01/20 for the other states. The latest date included here is 5/22/20. We obtained the number of new confirmed cases (per million of the population) every day by taking the difference between consecutive days, and use the moving average for the time-series of this quantity (averaged over 7 day windows). In our calculations and figures below, day zero corresponds to 01/22/20. Then we separated the growth curves into 4 groups in the following way.

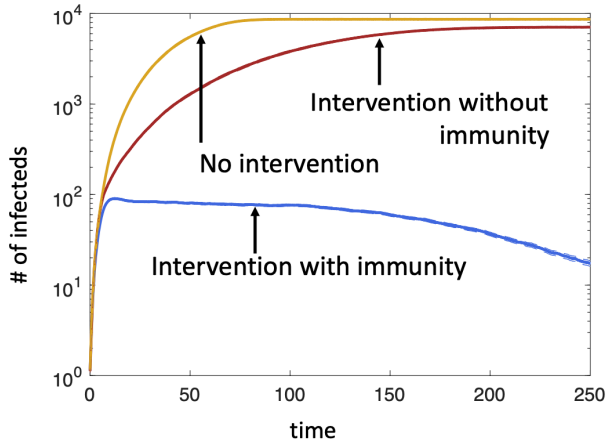

Figure S13: SIS dynamics on a spatial network. Compared are the number of infected individuals in the absence of intervention, with intervention in the presence of immunity and with intervention in the absence of immunity. Parameters are $P_{inf} = 10^{-4}, P_{recover} = 10^{-4}, P_{death} = 5 \times 10^{-5}$. In the absence of immunity, recovered individuals become susceptible. Intervention is modeled by removing 1/2 of connections. The average over 100 runs together with standard errors (dashed lines) is plotted for the simulations without immunity. The graph with immunity is the same as in Figure 3A.

For each dataset obtained as above, we considered the number of new infections starting on day 60 and until the end of the dataset (day 120). For each state i, these daily numbers of new cases were split into 3 bins: $[0, M_i/3], (M_i/3, 2M_i/3], (2M_i/3, M_i]$, where M_i is the maximal daily number of new cases for state i. These bins are shown as histograms in figure S14. We postulate that the number of cases experiences a plateau if the number of cases on day 120 falls within the largest bin. There are three cases of plateau in figure S14:

- Group 1 (6 states) corresponds to a relatively high peak following a fall in the number of daily cases that enters a plateau. For this group, the first bar is the highest in the histogram.

- Group 2 (7 states) is characterized by a relatively shorter peak followed by a plateau. For this group, the second bar is the highest in the histogram. Compared with group 1, the relative difference between the peak and the plateau is smaller.

- Group 3 (15 cases) shows a rise in the daily number of cases followed

by a plateau, without experiencing a prominent peak. For this group, the third bar is the highest in the histogram, and this is the largest of the three plateauing groups.

In addition to these three groups, two more non-plateauing groups were identified. For these groups, the number of cases on the last day did not belong to the largest bin.

- Group 0 (6 states) has the number of infections on day 120 that is smaller than a half of the maximum, $M_i/2$. This group is characterized by a rise and fall of the infection, but no subsequent plateau.

- Group 4 (10 states) has the number of infections on day 120 that is larger than $2M_i/3$, that is, it belongs to the third bin. This group is characterized by a rise of the infection, which has not experienced a significant decrease or a plateau.

Note that states in group 0 could be reclassified and enter group 1 later, as more data become available. Similarly, states in group 4 could enter any of groups 1,2,3, if they enter a plateau in the future. Further, we note that the above classification only operates with relative numbers and time-series shapes, and does not attempt to compare case numbers per million across states. For example, group 0 contains NY where daily cases reach to almost 500 (per million), and WA where the maximum number of cases was smaller than 60 (per million).

Analysis of the different infection patterns was performed and is shown in the main text. In particular, we determined the timing of infection rise. To do so, we determined the first day when the daily new infections (per million) reached 30. In figure S15(a), the colorful dots show the timing of the epidemic rise (as defined above) for each of the 5 groups. Solid blue lines connect the mean values for each of the groups, and dashed lines are linear fits. We can see that the relative timing of the epidemic rise is correlated positively with the group number ($p = 0.00008$). There is however no statistically significant difference among the 3 groups that experience a plateau. Therefore, in the main text we combine the 3 plateauing groups into one, and show that, with high significance, the epidemic started the earliest for group 0, later for the plateauing groups, and group 4 had the latest rise time.

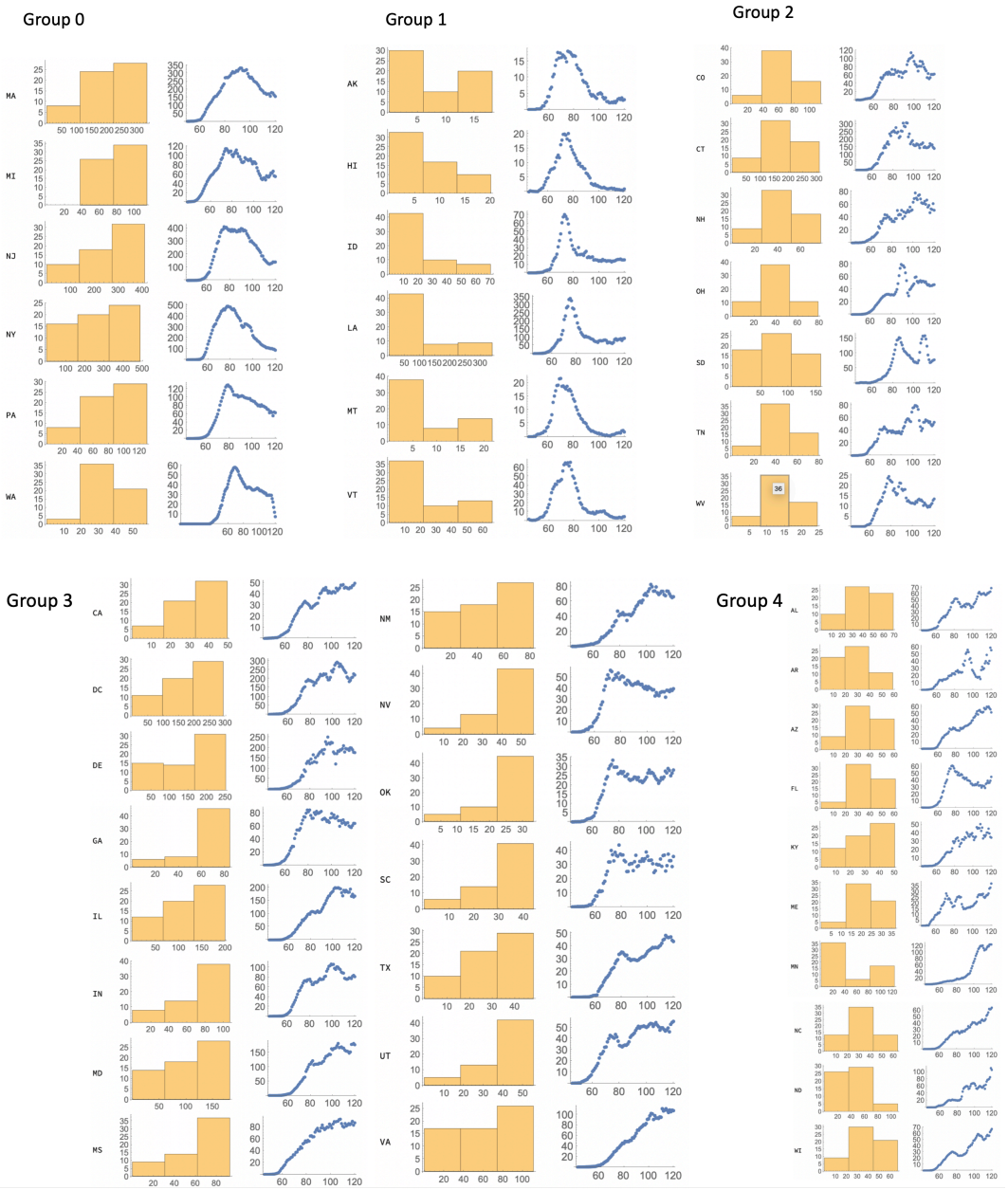

Figure S14: The groups of states classified by their infection patterns. For each state (marked on the left by the state abbreviation), the histogram for the daily numbers of cases is presented (yellow bar graphs, the horizontal axes is daily cases, the vertical axes is the number of days), followed by the time-series of the new cases (blue, he horizontal axes is time in days, and the vertical axes is the number of new cases, averaged over 7 days).

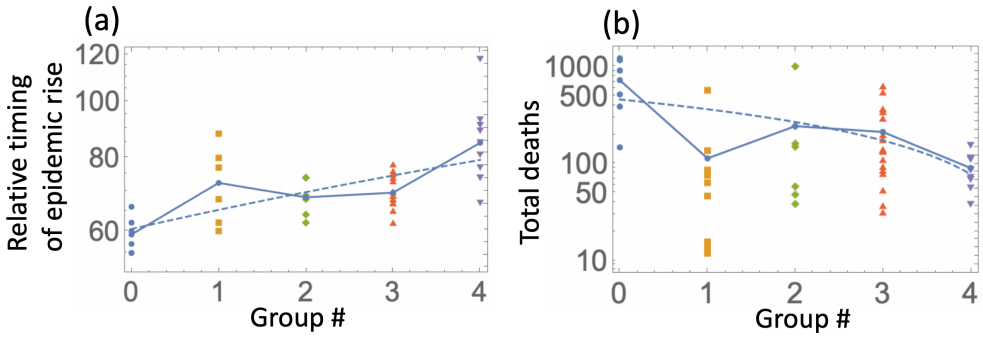

Figure S15: Correlations of groups of states with (a) relative time of epidemic rise (the day when the number of new infections first reached 30) and (b) the total number of deaths by day 120. Colored symbols represent the data for groups 0-4, the solid blue lines connect the means, and dashed lines correspond to the best linear fits.

Similar analysis was performed for the the total number of COVID-related deaths recorded in each state by day 120., see figure S15(b). Linear fit (the dashed line, please note the log scale) shows a strong negative correlation between groups and the level of deaths ($p = 0.002$). Again, since no statistically significant difference is observed among the 3 groups that experience a plateau, in the main text we combine the 3 plateauing groups and show that, with high significance, the total number of deaths (per million) is the highest for group 0, intermediate for the plateauing groups, and lowest for group 4.