



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

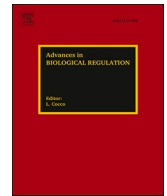
Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



ELSEVIER

Contents lists available at ScienceDirect

Advances in Biological Regulation

journal homepage: www.elsevier.com/locate/jbior

Virus spread on a scale-free network reproduces the Gompertz growth observed in isolated COVID-19 outbreaks

Francesco Zonta^b, Michael Levitt^{a,*}

^a Department of Structural Biology, Stanford University School of Medicine, Stanford, CA, 94305, USA

^b Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China

ABSTRACT

The counts of confirmed cases and deaths in isolated SARS-CoV-2 outbreaks follow the Gompertz growth function for locations of very different sizes. This lack of dependence on region size leads us to hypothesize that virus spread depends on the universal properties of the network of social interactions. We test this hypothesis by simulating the propagation of a virus on networks of different topologies or connectivities. Our main finding is that we can reproduce the Gompertz growth observed for many early outbreaks with a simple virus spread model on a scale-free network, in which nodes with many more neighbors than average are common. Nodes that have very many neighbors are infected early in the outbreak and then spread the infection very rapidly. When these nodes are no longer infectious, the remaining nodes that have most neighbors take over and continue to spread the infection. In this way, the rate of spread is fastest at the very start and slows down immediately. Geometrically we see that the "surface" of the epidemic, the number of susceptible nodes in contact with the infected nodes, starts to rapidly decrease very early in the epidemic and as soon as the larger nodes have been infected. In our simulation, the speed and impact of an outbreak depend on three parameters: the average number of contacts each node makes, the probability of being infected by a neighbor, and the probability of recovery. Intelligent interventions to reduce the impact of future outbreaks need to focus on these critical parameters in order to minimize economic and social collateral damage.

1. Introduction

Detailed analysis of SARS-CoV-2 reported number of cases show that in most of cases early isolated outbreaks follow a Gompertz function (Gompertz, 1823; Levitt et al., 2020; Castorina and Lanteri, 2020; Catala et al., 2020; Pelinovsky et al., 2022; Akira Ohnishi and Fukui, 2020)). This simple function is characterized by three parameters: N , U and T , so that total counts, $X(t)$ at time t are

$$X(t) = Ne^{e^{-\frac{(t-T)}{U}}} \quad (1)$$

where N is the total size of the outbreak, U is a time constant that determines how fast the growth and the slowdown occur, and T is the day the number of daily counts peaks.

Analogous behavior can be found in other respiratory disease epidemics, for example the 2003 SARS outbreak (Zhang and Wen, 2017), suggesting that time-course of respiratory virus spread follows simple and universal rules that are largely independent of genetic or social differences in the various countries.

In this paper, we analyze a simple model of virus diffusion on networks representing normal human social interactions (Moreno et al., 2002, 2004). We show that it is possible to reproduce the Gompertz behavior of actual data when the virus is spreading in a scale-free network. The model also allows us to derive a causal relationship between the total number of infections and some "microscopic" parameters that are properties of the virus and its interaction with society. In principle, Social intervention could affect

* Corresponding author.

E-mail address: Michael.levitt@stanford.edu (M. Levitt).

<https://doi.org/10.1016/j.jbior.2022.100915>

Received 6 June 2022; Received in revised form 31 July 2022; Accepted 6 September 2022

Available online 30 September 2022

2212-4926/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

these parameters, therefore, understanding their effect on the spread of the virus can help determine the most cost-effective strategy to contain virus spread in future outbreaks.

2. Methods

2.1. Definition

M Total number of nodes in the network

$X(t)$ Total number infected at time t

$N = \lim_{t \rightarrow \infty} X(t)$ Plateau value for the number of infected

P_I Probability of infection (per unit of time)

P_R Probability of recovery (per unit of time)

$R(t)$ Number of recovered people at time t

$I(t) = X(t) - R(t)$ Number of people that can infect other people at time t

$\Delta(t) = X(t) - X(t-1)$ Number of new daily infections at time t

$f(t) = \frac{X(t)}{X(t-1)}$ Fractional change of total number at time t

$r(t) = \ln[f(t)] = \ln \left[\frac{X(t)}{X(t-1)} \right]$ Exponential growth rate in that $X(t) = X(t-1) \exp[r(t)t]$

$J(t) = \ln[r(t)] = \ln \left[\ln \left[\frac{X(t)}{X(t-1)} \right] \right]$ Log of Exponential growth rate at time t

$Z(t) = \ln[\ln[N/X(t)]]$ log-log of the plateau value divided by the total number of case at time t

$\Sigma(t)$ Epidemic surface (total number of contacts between infected and susceptible nodes; multiple infected nodes may be in contact with the same susceptible node)

2.2. Simulations and computer code

The original code for generating the network and the stochastic model of virus diffusion is written in C. Results have also been double-checked and reproduced using a Python script and the Networkx package (Hagberg and Swart, 2008).

2.3. Choice of parameters for simulations

It is possible to roughly estimate realistic values for P_R , as it represents the probability that a person recovers from the virus in a given time interval.

The average time that a node remains infected can be calculated analytically from P_R as follows.

$P(n)$ the probability of recovering by day n , can be written:

$$\begin{aligned} P(1) &= P_R \cdot (1 - P_R)^0 \\ P(2) &= P_R \cdot (1 - P_R)^1 \\ &\dots \\ P(n) &= P_R \cdot (1 - P_R)^{n-1} \end{aligned}$$

Now the average time can be computed as:

$$\langle n \rangle = \sum_{d=1}^{\infty} n P_R \cdot (1 - P_R)^{n-1} = \frac{1}{P_R}$$

Recovery time for COVID-19 is about one week but can take up to 2 months. For this reason, we choose P_R values in the range 0.01–0.2 (recovery times from 5 to 100 days).

It is less clear how to relate P_I to observables in the real world. In the model, P_I represents the probability that one node becomes infected if it has been in contact with another infected node for a chosen time interval. This number depends on many variables in the real world, such as how strong the contact is, which kind of interaction the nodes have, the viral load, etc. Taking all these possibilities into account would result in a more realistic model. At this stage, we opt for a simpler model with a single value for P_I and also for using the same range for both P_I and P_R .

A posteriori, we discovered that P_I is inversely proportional to the key Gompertz function time-constant U , and changing P_I can be also be interpreted as a rescale of the time.

2.4. Assessing whether a single trajectory follows the Gompertz growth function

We consider the function $J(t) = \ln \left[\ln \left[\frac{X(t)}{X(t-1)} \right] \right]$. Elementary algebra shows that $J(t)$ is a linear function if, and only if, $X(t)$ is the Gompertz growth function. Assessing the linearity of $J(t)$ is then sufficient to determine if the simulated infections are growing as the Gompertz function. By visual inspection of the $J(t)$ function, we observe that in the vast majority of cases, infection spread like Gompertz function in scale-free networks, but not with other networks. The results were further confirmed by computing the

correlation coefficient between $J(t)$ and t , which is always close to 1. The only cases in which the growth is not Gompertz are cases in which the virus does not propagate effectively in the network (less than 0.5% of the network is infected). The same is not true for other networks (Supplementary Fig. 1).

2.5. Proportionality between the daily newly infected and the "surface" of the epidemics

We define the function $E(t)$ as the nodes that are exposed to the infection, i.e. the nodes connected to at least one node in $I(t)$. Each node in $E(t)$ can be connected to multiple nodes in $I(t)$. We define d_i as the number of connections of the i -th node belonging to the set $E(t)$ that are received from all the nodes in the set $I(t)$.

From r_i , we can derive an approximate formula to derive the probability that the i -th node is infected. The probability that the i -th node will not be infected at a certain step $t+1$ is indeed:

$$P_0 = (1 - P_I)^{r_i}$$

So that the probability of being infected, for the same node, is:

$$P_{\text{INFECTION}} = 1 - P_0 = 1 - (1 - P_I)^{r_i} \approx r_i P_I$$

where the last approximate equality holds if $P_I \ll 1$.

The total nodes that are infected at time $t+1$ are then:

$$\Delta(t+1) = P_I \cdot \sum_{i \in C(t)} r_i = P_I \cdot \Sigma(t)$$

where we defined the quantity $\Sigma(t)$ as the total number of contacts that each infectious node makes with non-infected nodes (this is the "surface" of the epidemic).

3. Results

3.1. A simple model of virus diffusion on a scale-free network reproduces Gompertz dynamics found in observed data

We considered a simple infection model on a network. Individual people are nodes in the network and are connected by links representing the possibility of interaction (and thereby virus spreading). If node i interacts with node j , node j also interacts with node i

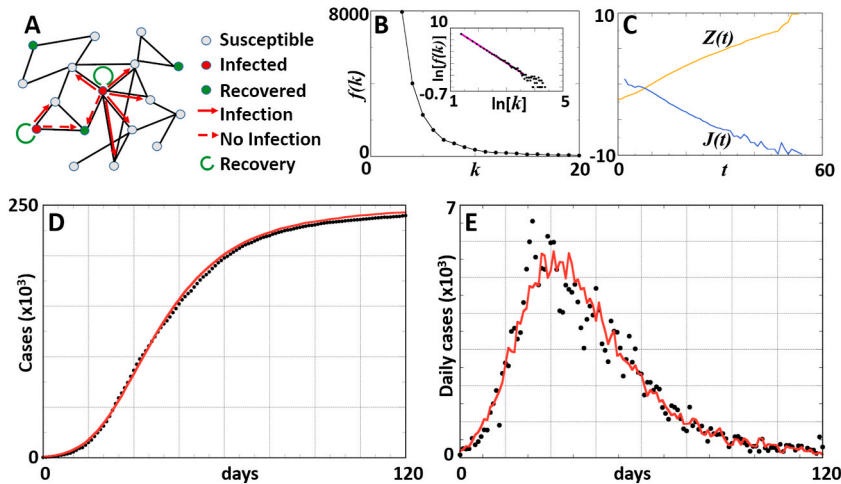


Fig. 1. Spread of the virus modeled on a network. (A) Our model of virus spread: at each time step, infected nodes (in red) can transmit the infection to a susceptible node (gray) with a probability P_I , only if there is a connection between the two nodes. After being infected, infected nodes can recover with a probability P_R . Recovered nodes cannot be re-infected. Panel (B) shows the distribution of the ranks of the nodes in the scale-free network used in the simulations. The inset shows the same plot on a ln-ln scale. As expected from theory (Prettejohn et al., 2011), there is a straight-line fit with exponent (or slope) $\gamma = -3.0$. Panel (C) Shows the functions $J(t) = \ln[\ln[X(t)/X(t-1)]]$ and $Z(t) = \ln[\ln[N/X(t)]]$ (where $X(t)$ is count at time t ; N is plateau or maximum count level) during a simulated infection with $P_I = 0.1$ and $P_R = 0.1$. The fact that functions $J(t)$ and $Z(t)$ are linear functions of time indicates that viral spread follows Gompertz growth. The slopes of the $J(t)$ and $Z(t)$ lines are equal to $-1/U$ and $1/U$, where $U = 4.7$ & 4.6 , days for $J(t)$ and $Z(t)$, respectively. Gompertz growth is observed only when the network is scale-free (Supplementary Fig. 1). Panels D (total cases) and E (daily cases) show a comparison between the results produced by a simulation (red lines, $P_I = 0.02$ and $P_R = 0.02$) and the actual confirmed cases observed in Italy in the first outbreak (black dots). Since the network used in the simulations is relatively small (20,000 nodes), results need to be rescaled by a factor of 13.4 for the comparison to real data.

(undirected network). The number of links k that are connected to a particular node is called the *degree* of the node. The network is static, i.e., links do not change with time. One can interpret links as connections between people whose interaction is sufficiently strong to allow infection.

We follow the dynamics of the spreading of the virus using a discrete version of the SIR model (Moreno et al., 2002; Kermack and McKendrick, 1927) in which each node can be in one of three states: State S defines a *Susceptible* node, which can become *Infected* (State I) with a given probability P_I , if and only if it is connected to another infected node (in SEIRs models, these nodes are considered Exposed – we do not consider this status here for simplicity, but exposed nodes are well defined within each simulation). Infected nodes can become *Recovered* (State R) with a probability P_R , and they cannot then be re-infected. These ‘immune’ nodes represent individuals who can no longer infect or be infected, either because they have successfully recovered from the disease, have been isolated from the network (for example, placed in hospitals or quarantine), or are no longer alive. The infection propagates on the network according to the following algorithm. At time $t = 0$, a small number (m_0) of nodes are randomly infected. At each subsequent time step, an infected node infects a node it is connected to with a probability P_I and recovers with a probability P_R . The process is iterated until reaching an equilibrium state in which there are no more infected nodes (Fig. 1A).

We observe that the shape of the virus spread in the network depends on the network topology: if the network connectivity is *scale-free* (Albert and Barabási, 2002; Barabasi and Albert, 1999; Prettejohn et al., 2011), i.e., the distribution of the degrees of the nodes follows a power law $P(k) = k^{-\gamma}$ with $2 < \gamma < 3$ (Fig. 1B), the propagation of the virus follows a Gompertz law (Fig. 1C). With appropriate choice of parameters, the simulation produces results that can be compared with actual observed data (as an illustrative example, in Fig. 1 panels D and E we show a comparison with cases in Italy cases, after rescaling). Following Gompertz and fitting real-world data is a characteristic of the scale-free network: other topologies (small world or random networks) fail to produce the same behavior (Fig. 5B & Supplementary Fig. S1).

Due to the stochastic nature of our model, we still observe instances in which the spread of the virus does not follow a Gompertz growth function. This happens only when the virus infects a very small region on the network (less than 50 nodes). Such instances become more probable as the ratio between P_R and P_I increases, but occur in no more than 30% simulations even when $P_R/P_I = 5$. In

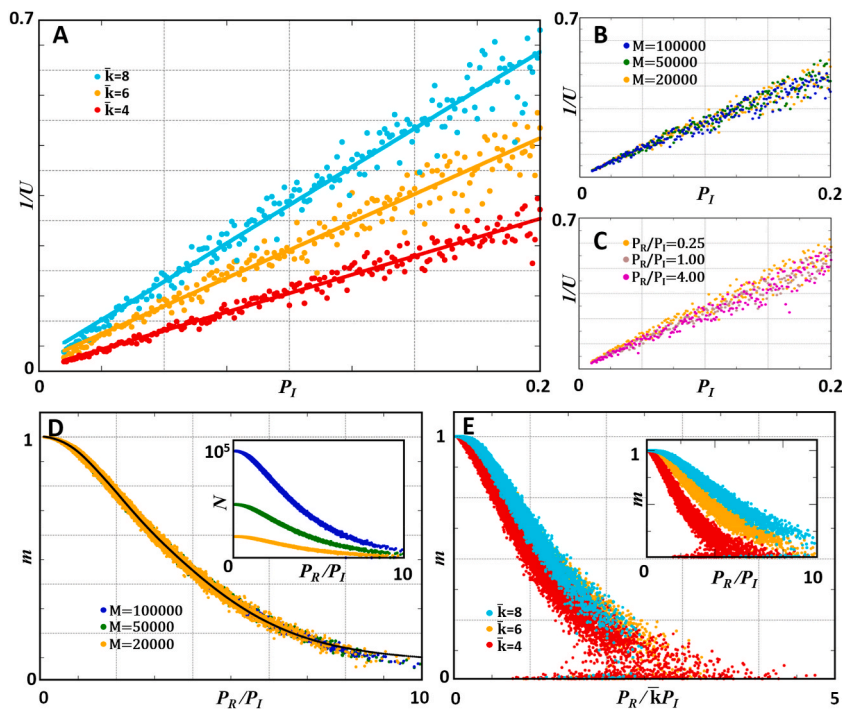


Fig. 2. The speed of spread and the fraction of the network that is infected at the plateau depend on the microscopic properties of the virus infectivity and on the average network connectivity. Each dot in the graphs corresponds to a different simulation. Panel (A): the parameter $1/U$ is a linear function of P_I with slope proportional to the average number of contacts of the network, \bar{k} . The three sets of data correspond to three networks with $\bar{k} = 4$ (red dots), $\bar{k} = 6$ (orange dots), and $\bar{k} = 8$ (cyan dots) and $M = 20,000$ nodes. Straight lines are best linear fits and have slopes of 1.49, 2.23, and 3.05, respectively, making the value for the constant α in Eq. (2) equal to 0.375. Panels (B) and (C) show that $1/U$ does not depend on the size of the network, M , or the recovery probability, P_R/P_I . Panel (D) shows the fraction of the network that is infected $m = N/M$ (where M is the total number of nodes in the network), depends on P_R/P_I for a fixed value of $\bar{k} = 6$. Note that m does not depend on the size of the network. The black line is the fit of m as a function of P_R/P_I given in Eq. (3). The first four terms achieve a good fit, with the following values for the constants: $a_1 = 0.20$, $a_2 = 0.68$, $a_3 = 0.10$, $a_4 = 0.02$, $c_1 = 0.07$, $c_2 = 0.05$, $c_3 = 0.11$, $c_4 = 0.42$. The inset shows the plateau values, N , as a function of P_R/P_I for different values of M . Panel (E) shows how m changes as a function of P_R/P_I . The main graph shows that if P_R/P_I is rescaled by \bar{k} , we obtain the same dependence on m , whereas the inset shows the functions before rescaling.

other words, when the virus actually spreads through the network, it always does so by following the Gompertz growth function.

3.2. In a scale-free network, outbreaks are entirely determined by microscopic parameters and network connectivity

Having shown that when outbreaks are significant, they always follow the Gompertz growth function, we wanted to understand how the parameters N and U in Eq. (1) are linked to the microscopic rate of infection and recovery and possibly to other network properties. The remaining parameter (T) can be ignored as it simply describes a translation of the Gompertz curve along the time axis.

To study the effect of microscopic parameters, we run many simulations with different values for P_I and P_R on the same network, but with different random initial seeding of the infection. Similarly, we tested how the results depend on the parameters that define the network (namely the total number of nodes M and the average connectivity \bar{k}). In order to have a rough comparison with actual data, simulations were run with P_I and P_R varying in the range 0.01–0.2, with a step of 0.001 for either parameter. The total amount of different runs analyzed in this way was 36100 for each network considered (see Methods).

Despite our model being stochastic, we find that the percentage of the total nodes ($m = N/M$ where M is the size of the network) and the time-constant (U) are well-defined functions of the two parameters P_I and P_R as well as of the average connectivity of the network, and they do not depend on the initial seeding of the virus in the network. This allows these quantities to be predicted quite well in our simulation. In particular, U is inversely proportional to the product of \bar{k} and P_I , ($U = 1/\alpha \bar{k} P_I$) while m is a decreasing function of $x = U \cdot P_R = P_R/\alpha \bar{k} P_I$ (Fig. 2):

$$U = 1/\alpha \bar{k} P_I \quad (2)$$

$$m = m(x) = \sum_{i=1}^{\infty} a_i e^{-c_i x^i} \quad (3)$$

$m(x)$ is a monotonically decreasing function of x ; it does not have a simple functional form but can be well approximated by the first four terms of the series in Equation (3) (see fit in Fig. 2). Constants α , a_k , and c_k in Equations (2) and (3) are fitting parameters and do not depend on the size of the network. As shown in Fig. 2 panel D and E, when $P_R/P_I < 0.5$, m is close to 1 and the network is almost entirely infected. Taking into account Eqs. (2) and (3), we can rewrite Eq. (1) as:

$$X(t) = Mn \left(P_R / \alpha \bar{k} P_I \right) \cdot e^{-\alpha \bar{k} P_I (t-T)} \quad (4)$$

Equation (4) shows that it is possible to describe viral spread on a scale-free network from microscopic parameters, which could be available *a priori* from intrinsic, known characteristics of both the virus and the network.

3.3. Infection of major hubs

We further analyzed the role of major hubs, the nodes with many neighbors, (the super-spreaders) in spreading the virus. Fig. 3 panel A gives the time, t_0 , at which each node of the network is infected as a function of $\ln[1/k]$, averaged over 100 simulations run with $P_I = 0.01$ and $P_R = 0.01$ using different random seeding of the infection. Major hubs with high k values are consistently infected at shorter t_0 than smaller hubs. Smaller hubs show much larger variances in the value of t_0 than larger hubs. Similar trends can be observed for other choices of parameters (Fig. 3 Panel B and Supplementary Fig. 2). The time to become infected is largely independent of P_R , unless the ratio P_R/P_I is close to 1.

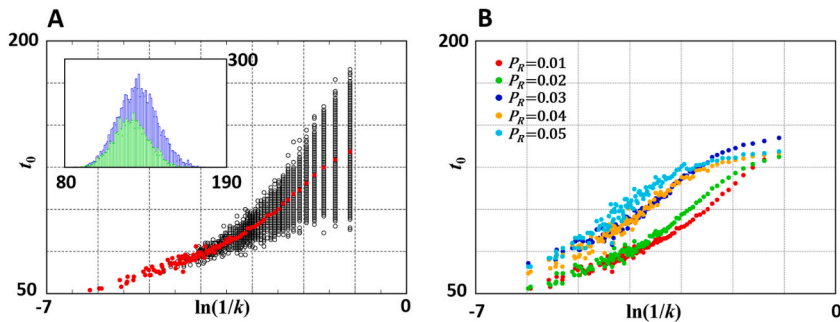


Fig. 3. Larger nodes are infected earlier in the epidemic. Panel (A) shows t_0 , the average time at which a particular node is infected as a function of the degree of the node (more precisely $-\ln[k]$). Each black dot represents a different specific node. The time of infection is averaged over 100 replica simulations. The t_0 value averaged over all the nodes of same degree k is shown by the red line. The inset shows histograms of the average time of infection for nodes with $k = 3$ (blue) and $k = 4$ (green); both distributions are normal distributions. Panel B plots the average t_0 , value for various P_R values when P_I is set to 0.01 (we average over all 100 replicas and over all the nodes with the same degree, k). The red dots correspond to the red line in panel A. Differences in t_0 are less evident with higher values of P_I (see Supplementary Fig. 3).

To understand whether the infections of major hubs actually causes faster spreading of the virus, we repeat the simulation with the additional condition that the 20 largest hubs (corresponding to nodes with $k > 100$) cannot be infected. Fig. 4 shows histograms for the fractional plateau value m for those simulations where the virus spreads widely in the network. We see that while the 'immunity' of the largest nodes makes negligible difference when $P_R/P_I = 1$, it becomes important when $P_R/P_I = 5$. In addition to the average value of m being smaller when big hubs are immune, the number of cases in which spread does not occur increases from 29% to 48% for $P_R/P_I = 5$, while it is almost unaltered (1% to 2%) when $P_R/P_I = 1$ (Fig. 4 inset). These numbers are for simulations in which 4 different random nodes are infected at $t = 0$; we expect more significant differences if the initial seeding number is smaller. Overall, we see that while infection of the nodes larger than 100 is not essential for the propagation of the epidemic, they are important drivers of virus spread by influencing the probability that an outbreak actually spreads widely and to what fraction of the network becomes infected.

3.4. In a scale-free network, the "surface" of the epidemic rapidly decreases after the largest nodes have been infected

The Gompertz function was initially envisioned to describe growth in an environment with limited resources (Gompertz, 1823). Its most important feature is that the Gompertz growth function starts with maximum velocity and then slows down immediately. Why does this happen in a scale-free network and not in other networks?

To answer this question and understand why this is a special property of the scale-free network, we track the time course of several observables and determine how they are correlated. Results are reported in Fig. 5 panels A for three different replicas of infection of a scale-free network and panels B for one example of infection in a random network (see also Supplementary Movies 1 and 2).

Supplementary video related to this article can be found at <https://doi.org/10.1016/j.jbior.2022.100915>

The top row panels show functions $X(t)$ (red line, total number of nodes that have been infected), $I(t)$ (gray line, number of infected nodes that have not recovered yet) and $\Delta(t)$ (orange line, daily new infected nodes, shown on a different scale for clarity). We can immediately observe that $I(t)$ and $\Delta(t)$ lose correlation after some time which happens much faster for a scale-free topology.

$\Delta(t)$ is proportional to $\Sigma(t)$, the number of contacts that infectious nodes make with non-infected nodes (the "surface" of the epidemic, see Methods section). The fact that $\Delta(t)$ decays faster than $I(t)$ means that nodes are infected later on, have less access to the pool of susceptible nodes. For a random network, this decay happens halfway through the infection, when the number of susceptible nodes are comparable or less than the number of infected nodes, as one would expect from a logistic growth; but happens much earlier in a scale-free network.

This can be understood by observing that in a scale-free network, bigger nodes are exhausted very early in the epidemic (Figs. 4 and 5 central rows), and thus the number of contacts that can be established with uninfected nodes immediately decays.

By contrast, in a random network, the biggest nodes are not very different from the others. While on average, they are also infected faster and more consistently, they do not appear to have any particular role in spreading of the epidemic. The shape of total infections and daily infection for the random network, looks like a perfect logistic function. Infection of the largest node in a scale free-network coincides with the beginning of the Gompertz growth of the epidemic, as is evident from the linear behavior of $J(t)$ (bottom rows in Fig. 5).

3.5. Multiple outbreaks cannot be observed in a single scale free-network, but emerge in multiple networks connected by a few links

In contrast with real epidemics, we never observe multiple outbreaks in single scale-free or random networks: in all the simulated trajectories, the virus exhausts itself following the Gompertz growth function (or other growth functions for the random network) even

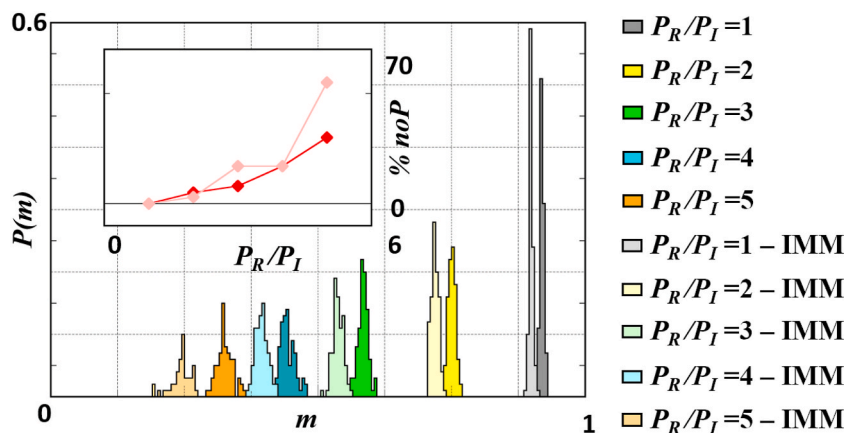


Fig. 4. Granting immunity to larger nodes reduces the fraction of the network that is infected. Each histogram shows the plateau fraction $m = N/M$ for a set of 100 simulations; for different values of the ratio P_R/P_I and for two different conditions: (i) every node can be infected, (ii) nodes with most neighbors (top 0.1% of all nodes) are immune from the beginning of the simulation. The inset shows the percentage of simulations in which the virus does not spread into the network, i.e., the final fraction of m is below 0.005. Immunization of highly connected nodes decreases the plateau value of m and increases the probability that the virus does not spread in the network.

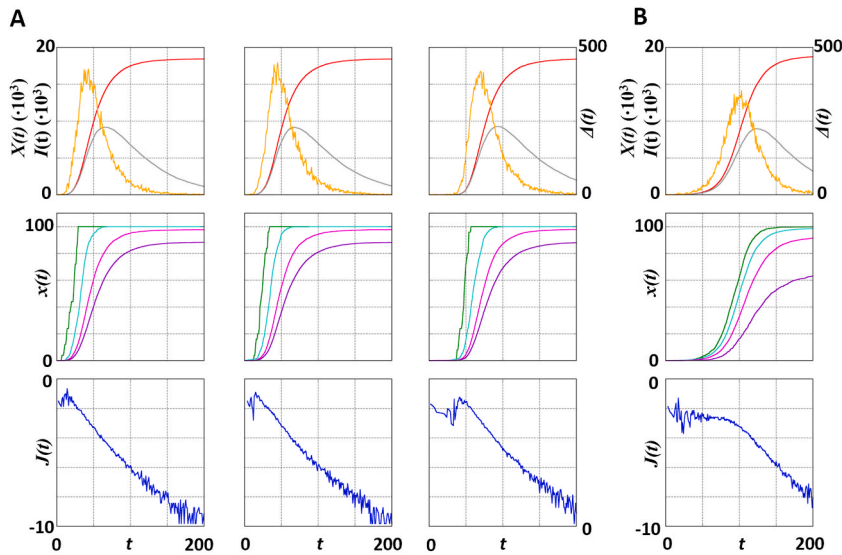


Fig. 5. In a scale-free network, rapid spread of the virus is caused by the infection of the larger nodes. Panel (A) shows three replicas of simulations for the scale-free network topology, whereas panel (B) shows one typical simulation for the random network topology. In all the cases, we ran simulations with $P_I = 0.02$ and $P_R = 0.02$. The top row plots functions $X(t)$ (total cases, red line), $I(t)$ (number still infected, gray line) and $\Delta(t)$ (new daily infections, orange line) against time, t . The middle row plots the percentage of infected nodes as a function of t for four ranges of k value (we called this function $x(t)$). For the scale-free network, the clusters are $k \geq 100$ ($\sim 0.1\%$ of the total, dark green line), $20 \leq k < 100$ (2.7% of the total, cyan line), $5 \leq k < 20$ (37.2% of the total, magenta line) and $k < 5$ (60.0% of the total, purple line). For the random network, the clusters, which involve much smaller k values, are $k \geq 10$ (8.5% of the total, dark green line), $6 \leq k < 10$ (47.1% of the total, cyan line), $3 \leq k < 6$ (38.2% of the total, magenta line) and $k < 3$ (6.26% of the total, purple line). Most remarkable is that in the scale-free network, the largest nodes are fully infected very early and before the next largest nodes. In the random network, the largest nodes are infected a little before smaller nodes but they take much longer to be fully infected. The bottom row plots the functions $J(t)$ (\ln (exponential growth rate) as a blue line).

if the final fraction of the network that has been infected is small. Indeed, as shown in Figs. 4 and 5, larger nodes, which fuel the Gompertz outbreak, are always infected early, and after they recover there is no possibility of creating a new outbreak.

However, we can observe a second Gompertz outbreak if we connect two different scale-free networks with a small number of links

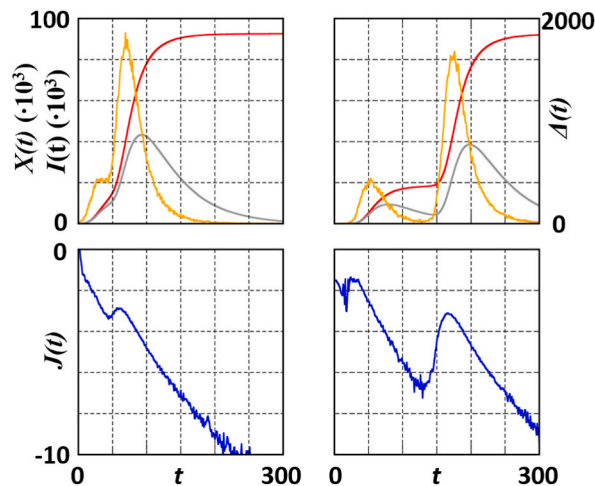


Fig. 6. Multiple outbreaks emerge in networks built from connected scale-free networks.

We simulated the infection in a network that consists of two scale-free networks ($M_1 = 20000$, $M_2 = 100000$) connected by a small number (Hagberg and Swart, 2008) of edges. In both cases, we infected 4 random nodes of the smaller network at the beginning of the simulations. The infection and recovery rate are fixed at $P_I = 0.02$ and $P_R = 0.02$. The top row plots functions $X(t)$ (total cases, red line), $I(t)$ (number still infected, gray line) and $\Delta(t)$ (new daily infections, orange line) against time, t ; the bottom row plots the functions $J(t)$ (\ln (exponential growth factor), blue line).

The two replicas show that the infection of the second network can happen at different times: in the first case we almost have superposition of the two peaks, while in the second case they are well separated in time. The function $J(t)$ resembles the one we observed in the real world data for multiple outbreaks.

(Fig. 6). The outburst of the second outbreak can happen almost immediately or well after the first outbreak has passed its peak and the overall dynamics are reminiscent of real world data (Paper I).

4. Discussion

4.1. Phenomenology of viruses spreading in scale free networks

The virus spread model we introduce and analyze in this work correctly reproduces the main characteristics of COVID-19 in the real world. At the beginning of the infection, the virus propagates in the network infecting a small number of other nodes. This phase is likely to go undetected in the real world, as the number of infected individuals is small and COVID-19 can be asymptomatic. Depending on the characteristics of the virus, the infection can either disappear during this first phase, or last enough to reach one of the major nodes. In a scale-free network, such nodes are much better connected than the average node, and thus they are, on average, infected earlier than other nodes. When one of such nodes is infected we observe a rapid increase in the number of infected nodes, resulting in a major outbreak which will hide any smaller or secondary outbreak, i.e. the network is synchronized once major nodes are involved in the infection. The largest nodes propagate the infection towards other nodes that are smaller but still better connected than the rest of the network, and the process iterates through smaller and smaller nodes. This spread of a large-scale infection is self-limited because the number of nodes connected to each infected node becomes smaller and smaller.

This spreading pattern gives indirect evidence that the virus can spread through asymptomatic cases (Peirlinck et al., 2020) and is airborne (Zhang et al., 2020), since no other transmission routes can explain why some individuals could infect hundreds or thousands of people in a short amount of time (Kang, 2020).

Gompertz growth appears to be the fastest possible for a virus on a scale-free network, as any initial exponential phase can last only until a central node has been reached, and this happens in a very short time period. It is reasonable to assume that this limit speed can be achieved only by highly virulent viruses such as influenza, measles and COVID-19.

4.2. “Gompertz growth” and “scale-free networks”

The main result presented in this paper is that the virus spreads on a scale free network according to the Gompertz growth function, as it appears to happen for isolated Covid outbreak in the real world. It is worth to notice that both terms have not to be taken in the mathematical sense but more as a physical realization of these concept. Gompertz growth is not rigorously proved, but inferred by observing the linearity of the $J(t)$ function for most of the outbreaks. It is possible that more complex growth functions can better fit the data of the simulation, but they would be in any case virtually indistinguishable from a Gompertz function.

Furthermore, the total number of cases does not immediately follow the Gompertz function, as there is always an initial transient time in which the virus lurks in the network. However, this transient is stochastic in nature, affects only a small portion of the network, and does not carry any information of the extent of the infection at the end. Moreover, it is reminiscent of real world data, as in the first days of the outbreak a small number of cases are registered, before the number of cases start following the Gompertz growth function.

Scale-free networks, are very well defined mathematical objects which are likely not to be accurate representations of the human interaction in the society and are expected to be rare (Broido and Clauset, 2019; Holme, 2019). Nevertheless, several studies indicate that the distribution of social contacts is long tailed (Glass and Glass, 2008; Leung et al., 2017), and this has been observed also during the COVID-19 pandemic (Feehan and Mahmud, 2021). It is possible that other networks following a long tail distribution could produce similar results, or that an analogous pattern is generated by the reaction of individuals as they tend to reduce the number of contact proportionally to the perceived danger created by of virus. Still our model is conceptually simple and explains the apparent fractal nature of the virus spread.

4.3. Effect of intervention in the virus propagation

Equation (3) shows that the time constant U , which characterizes the speed at which the infection propagates in the network, depends only on two factors. One is the infection probability (P_I), which is a characteristic of the virus. The other is the average degree of connectivity of the social network of human interactions in a city, region, or country (\bar{k}), which summarizes the complex interaction patterns between people.

P_I can be different for different virus strains and can perhaps be modulated by the ambient temperature (the season) or by using masks, visors, gloves, and other protective measures. Similarly, the average connectivity, \bar{k} , can be modified by spontaneous human behavior or by government intervention.

The fraction of the network infected by the virus (m) depends on U and the recovery rate probability (P_R). P_R represents the rate at which infected nodes recover (or die) or are isolated from the network. Its value does not depend only on the virus characteristics but can be influenced by social interventions such as isolating infected people.

Increasing P_R by a certain factor is equivalent to decreasing P_I or \bar{k} by the same factor. However, the methods to achieve either case are drastically different in the real-world: increasing P_R can be done *a posteriori* by tracking and isolating infected people while decreasing \bar{k} would require *a priori* intervention such as lockdowns or school closure. The feasibility and the economic cost of the two approaches are different and depend on the country's internal policies.

Another way to affect the final number of infected people is by capping the extent of the network that the virus has compromised. In

our model, this is represented by the size of the network M , and is a fixed number. In the real world, lockdowns and other form of intervention could help to isolate the virus before it infects a new "region" of the network.

Finally, as mentioned above, it is possible that the apparent organization of human contacts in a scale-free network is due to the response and adaptation of the society to a spreading virus, as infected people tend to self-isolate, while susceptible people tend to reduce risky behaviours (for example by reducing contacts or using masks), and both effects become stronger as the number of infected increases, even without external policies.

It will be surely important in future to test how dynamical changes in the parameters alter the spreading of the virus in the scale-free model and whether similar dynamics can be observed in real world data. This will help us identify the efficacy of intervention and optimize intervention to reduce the stress they cause to both single persons and to the society as a whole.

5. Conclusions

Despite being minimal, the model proposed in this paper reproduces most of the characteristics of isolated COVID-19 outbreaks. It successfully reproduces the observed Gompertz growth behavior, explains why outbreaks present universal characteristics, and most importantly, provides a theoretical framework on which to interpret differences between various countries as differences in the social structure or how society responds to the virus.

Virus spread in a single scale free network model cannot produce multiple outbreaks, but multiple outbreaks appears to be possible when different scale-free networks are linked together by few contacts. Assuming that this is a realistic representation of reality, isolating individual bridges between different networks (Salathe and Jones, 2010) (for example controlling long distance travels) could be a much more effective way to isolate the virus before it spread to new locations.

Despite its stochastic nature, the model suggests that the outcome of each individual outbreak of a high virulent virus would be largely predictable, if we could get detailed information about the contact pattern within a society. Indeed, the extent and the temporal span of the outbreak have low variance, and are well defined by the initial characteristics of the virus.

It is undoubtedly possible to improve the realism of the model in various ways, for example, allowing for a dynamic contact structure. We believe that such changes may improve the description of what happens between two outbreaks but will not affect the modeling of single outbreaks. Indeed, the initial phase, in which the virus lingers within the network before finding one of the larger nodes, appears to be a stochastic process that is likely to be largely independent of the network properties; while the rapid growth of the Gompertz phase happens so fast that the underlying network would not have time to change.

CRedit authorship contribution statement

Francesco Zonta: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Michael Levitt:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgements

Support was provided by US National Institutes of Health award R35GM122543 (M.L.) and National Natural Science Foundation of China Grant No. 31770776 (F.Z.). Michael Levitt is the Robert W. and Vivian K. Cahill Professor of Cancer Research.

We thank Dr. Steven Strong for contacting us and sharing his idea of simulating virus spread on a network. His preliminary results convinced us to do the full-scale study that led to this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jbior.2022.100915>.

References

- Akira Ohnishi, Y.N., Fukui, Tokuro, 2020. Universality in COVID-19 spread in view of the Gompertz function. *Progr. Theor. Exp. Phys.* 2020 (12), 123J101.
 Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.
 Barabasi, A.L., Albert, R., 1999. Emergence of scaling in random networks. *Science* 286 (5439), 509–512.

- Broido, A.D., Clauset, A., 2019. Scale-free networks are rare. *Nat. Commun.* 10 (1), 1017.
- Castorina, P.I.A., Lanteri, D., 2020. Data analysis on Coronavirus spreading by macroscopic growth laws. *Int. J. Mod. Phys. C* 31, 2050103, 07.
- Catala, M., et al., 2020. Empirical model for short-time prediction of COVID-19 spreading. *PLoS Comput. Biol.* 16 (12), e1008431.
- Feehan, D.M., Mahmud, A.S., 2021. Quantifying population contact patterns in the United States during the COVID-19 pandemic. *Nat. Commun.* 12 (1), 893.
- Glass, L.M., Glass, R.J., 2008. Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Publ. Health* 8, 61.
- Gompertz, B., 1823. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Phil. Trans. Roy. Soc. Lond.* 115, 513–585.
- Hagberg, A.A.S.D.A., Swart, P.J., 2008. Exploring network structure, dynamics, and function using NetworkX. *Proc. 7th Python Sci. Conf. SciPy2008*, 11–15.
- Holme, P., 2019. Rare and everywhere: perspectives on scale-free networks. *Nat. Commun.* 10 (1), 1016.
- Kang, Y.J., 2020. Lessons learned from cases of COVID-19 infection in South Korea. *Disaster Med. Public Health Prep.* 14 (6), 818–825.
- Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond.* 115 (772), 700–721.
- Leung, K., Jit, M., Lau, E.H.Y., Wu, J.T., 2017. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci. Rep.* 7 (1), 7974.
- Levitt, M., Scaiewicz, A., Zonta, F., 2020. Predicting the Trajectory of Any COVID19 Epidemic from the Best Straight Line. *medRxiv*.
- Moreno, Y., Pastor-Satorras, R., Vespignani, A., 2002. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* 26, 521–529.
- Moreno, Y., Nekovee, M., Vespignani, A., 2004. Efficiency and reliability of epidemic data dissemination in complex networks. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* 69 (5 Pt 2), 055101.
- Peirlinck, M., et al., 2020. Visualizing the invisible: the effect of asymptomatic transmission on the outbreak dynamics of COVID-19. *Comput. Methods Appl. Mech. Eng.* 372, 113410.
- Pelinovsky, E., et al., 2022. Gompertz model in COVID-19 spreading simulation. *Chaos, Solit. Fractals* 154, 111699.
- Prettejohn, B.J., Berryman, M.J., McDonnell, M.D., 2011. Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Front. Comput. Neurosci.* 5, 11.
- Salathe, M., Jones, J.H., 2010. Dynamics and control of diseases in networks with community structure. *PLoS Comput. Biol.* 6 (4), e1000736.
- Zhang, M., Wen, J., 2017. SARS Time Series Modeling and Spatial Data Analysis. *DEStech Transactions on Computer Science and Engineering*.
- Zhang, R., Li, Y., Zhang, A.L., Wang, Y., Molina, M.J., 2020. Identifying airborne transmission as the dominant route for the spread of COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* 117 (26), 14857–14863.