A random-walk based epidemiological model

Andrew Chu[1], Greg Huber[1], Aaron McGeever[1], Boris Veytsman[2], and David Yllanes[1,*]

[1]Chan Zuckerberg Biohub, 499 Illinois Street, San Francisco, CA 94158, USA
[2]Chan Zuckerberg Initiative, 601 Marshall Street, Redwood City, CA 94063, USA
[*]david.yllanes@czbiohub.org

ABSTRACT

Random walkers on a two-dimensional square lattice are used to explore the spatio-temporal growth of an epidemic. We have found that a simple random-walk system generates nontrivial dynamics compared with traditional well-mixed models. Phase diagrams characterizing the long-term behaviors of the epidemics are calculated numerically. The phase boundary separating those sets of parameters leading to outbreaks dying out and those leading to indefinite growth is mapped out in detail. The functional dependence of the basic reproductive number R_0 on the model's defining parameters reveals the role of spatial fluctuations and leads to a novel expression for R_0. Special attention is given to simulations of inter-regional transmission of the contagion. The attack rate and the (growing) radius of gyration of the affected zones are used as measures of the severity of the outbreaks, in cases where R_0 is not sufficiently prescriptive to chart the epidemic dynamics.

Introduction

Classic epidemiological models[1,2] often assume panmictic populations: every infected person has an equal chance to affect any other person in the population. While these models have been successful in describing an ideal dynamics of epidemic spread, they do not account for inhomogeneity: an infected person has higher probability to transmit the disease to a member of their household[3,4] or to a person in their locale. The overall reduction of long-distance travel due to the COVID-19 pandemic makes the latter cause of inhomogeneity especially relevant: many infected persons (but not all) spread the infection only in their immediate vicinity. Likewise, the local neighborhoods of both susceptible and infected persons are dynamic – they vary in space and time depending on the local state of the epidemic. These spatio-temporal inhomogeneities are studied in this paper.

There are two approaches to geographical inhomogeneity: detailed analyses based on contact-tracing and mobility data,[5,6] or analyses based on stylized models.[4,7,8] The first approach, while potentially highly accurate, requires a large number of parameters and is not robust with respect to unforeseen changes in mobility patterns. The second approach, in contrast, requires a small number of well-defined parameters and provides a sound intuition about their effects on the outcome. It is especially useful when planning interventions and policy changes, or investigating various "what if" scenarios. It is the latter approach that we have taken here.

In the present paper, we treat a very simple model: a random walker (modeled on an infected person) takes τ steps (a positive integer) on a 2D square lattice. After each step, the visited site can become infected with probability p. In this case, at the site of the new infection, it branches off another random walker, also with lifetime τ, that can also further spread the contagion to further sites, and so on. We study the resulting epidemic in space and time.

This model depends on two parameters: τ, representing the length of time of the infectious period, and p, representing the infectiousness. It also depends on the "walking pattern", which may include steps to the neighboring sites or jumps to more distant locations.

Panmictic models predict that the spread of epidemics depends only on the basic reproduction number R_0, the number of infected persons produced by one walker in a fully susceptible population, which, in turn, can be estimated in the panmictic approximation as

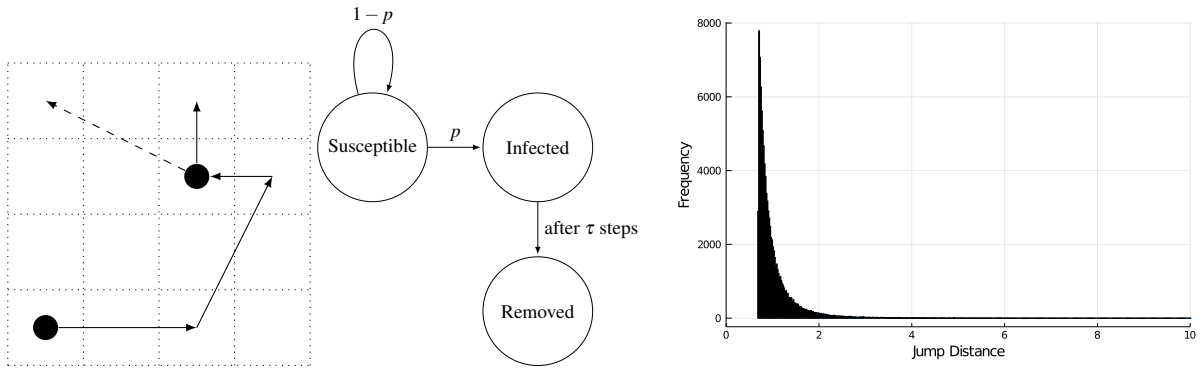
$$R_0 \approx p\tau. \tag{1}$$

We will refer to this estimate as the *naïve approximation*.

However, the random-walk model exhibits a surprisingly rich behavior, well beyond the "naïve approximation" above. We will show that the spread of the epidemics depends on the interplay between p and τ. Moreover, the detailed geometry of the epidemics depends on the stochasticity of individual walks and the nature of random walks on the plane. In subsequent sections, we describe our model in full, study its behavior using numerical simulations and theory, and draw conclusions for real-world epidemics.

Model

We model the geographic distribution of the population by an infinite 2D square lattice. At any given time, each site can be in one of three states. Although the model proposed here is fundamentally different than an SIR model, the same nomenclature is



(a) A random walk (b) State diagram (c) Histogram of jump distances

Figure 1. The model. 1(a) A walker (solid line) at some step infects a susceptible site, which generates another walker (dashed line). 1(b) The walker state diagram for each step, where p and $1-p$ transitions are triggered by infectious-susceptible interactions. 1(c) Histogram of jump distances of infective agents for 100000 realizations.

adopted; namely, each site in the lattice is either *susceptible*, *infected*, or *removed*. (Figure 1(b)). All sites are initialized as susceptible, with the exception of the origin which contains a single infectious agent. An infected site generates a brand new random walker, which starts to move on the lattice (Figure 1(a)). If the walker lands on a susceptible site (sites are represented as square cells in the figure), it becomes infected with probability p. After τ steps an infectious agent becomes removed. Thus, there are two fundamental parameters that control the outbreak: p, the probability of infection, and τ, the number of steps for which an infective agent (walker) is active before being removed.

Another feature of the model is the way the infectious agent chooses its next move. The model is quasilocal: most of the time the walker moves to one of the eight adjacent sites, but sometimes it makes a longer jump. Namely, if the walker is at $\mathbf{r}_0 = (x_0, y_0)$, its next location, $\mathbf{r}$, is given by

$$\mathbf{r} = \left(\lfloor x_0 + r\sin\theta \rfloor + \tfrac{1}{2}, \lfloor y_0 + r\cos\theta \rfloor + \tfrac{1}{2}\right), \qquad r = \frac{1}{(3u)^{1/3}} \tag{2}$$

with u and θ drawn from uniform distributions:

$$u \in U(0,1), \qquad \theta \in U(0,2\pi). \tag{3}$$

Here $\lfloor X \rfloor$ is the integer part of X (the largest integer number not exceeding X), and $U(a,b)$ is the uniform distribution in the interval $[a,b]$.[1] This function produces a one-step walk to one of the eight adjacent sites approximately 66.8% of the time; 95.7% of jumps are within a distance of 2 sites. The random-variate generator for the jump distances r in eq. (2) is derived from a probability density function of the form $\phi(r) = A/r^4$, where the constant A is of order the lattice constant cubed a^3 (in this paper, $a = 1$ is chosen throughout). The exponent of -4 corresponds to interactions of the form $\sim 1/r^{d+\sigma}$ when $d = 2, \sigma = 2$, a case of special interest in a number of early studies,[9,10] because it sits at the margin between long-range and short-range interactions. However, in the context of epidemic models, as noted by Bunde *et al.* [11], Grassberger [7], and Hallatschek and Fisher [8], it presents no special difficulties.

Results

Initial stages of the epidemics

In the initial stage of the outbreak, the infected person (the index case) is completely surrounded by the susceptible population. The key parameter describing this state is R_0: the number of people infected by the index case. The value of R_0 depends on p and τ in some form, but the naïve approximation, $R_0 \approx p\tau$, must fail, as can be seen whenever the path of even a single walker self intersects; *i.e.*, the approximation does not account for the recurrence of a two-dimensional random walker.

We ran 100000 realizations of the simulation, collecting how many agents were infected by the index case. Table 1 shows the R_0 calculated for p and τ tabulated on log-log axes,[2] an idea that will be explored further in the discussion of the phase diagram. The breakdown of the naïve approximation is evident in the lack of constancy of R_0 along the diagonals of the table, which is particularly apparent when $p\tau$ is large.

[1] The nodes of the lattice have integer coordinates, but our walkers move on the dual lattice, consisting of the centers of the elementary cells.

[2] R_0 is slightly less than 1 for $p = \tau = 1$ because there is a very small probability that the walker will not leave its starting cell.

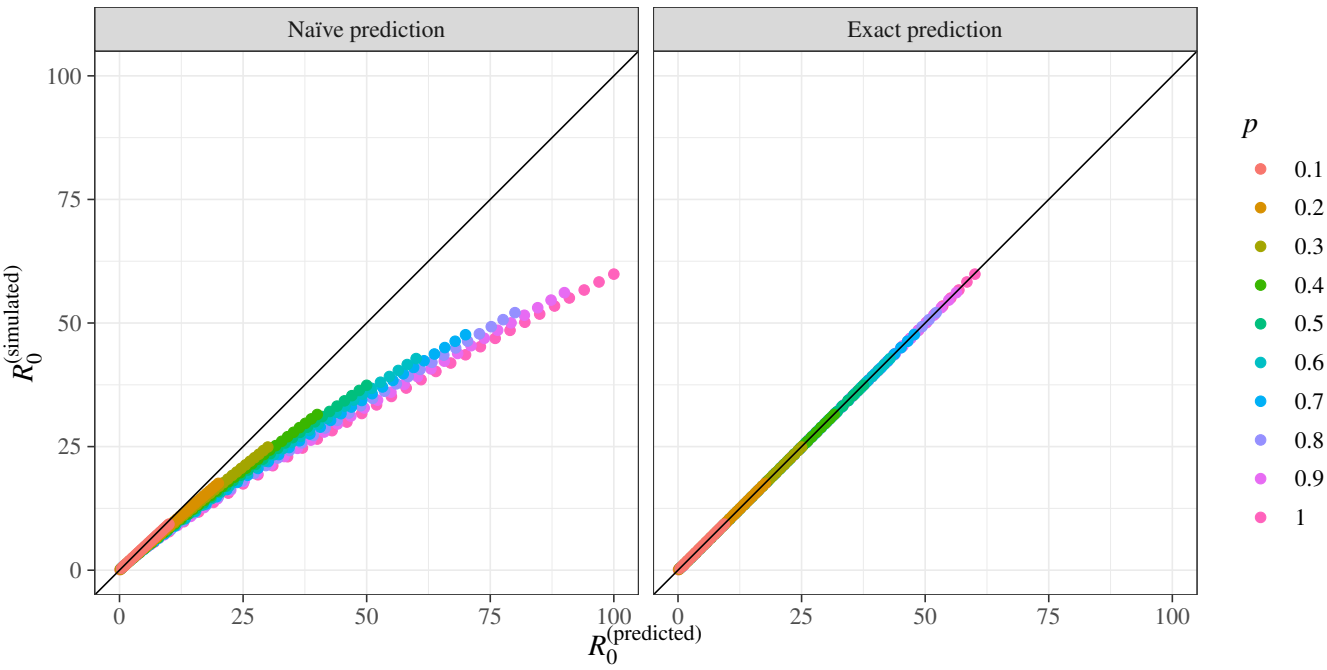


Figure 2. The prediction for R_0 and simulation data. Left panel: actual $R_0^{(\text{simulated})}$ against the naïve prediction $R_0^{(\text{predicted})} = p\tau$ (equation (1)). Right panel: $R_0^{(\text{simulated})}$ against the prediction of eq. (4). The straight line corresponds to $R_0^{(\text{simulated})} = R_0^{(\text{predicted})}$.

The theoretical result (see Methods) is

$$R_0 = \frac{p\tau}{1 + Kp\ln(\tau/\tau_0)}, \tag{4}$$

where K and τ_0 are constants that depend on how the walker's next jump is selected. For the jump generator in equation (2) we have

$$K \approx 0.174, \qquad \tau_0 \approx 2.19. \tag{5}$$

As shown in Figure 2, this formula is in excellent agreement with the numerical experiments without any adjustable parameters.

Outbreak progression and phase diagrams

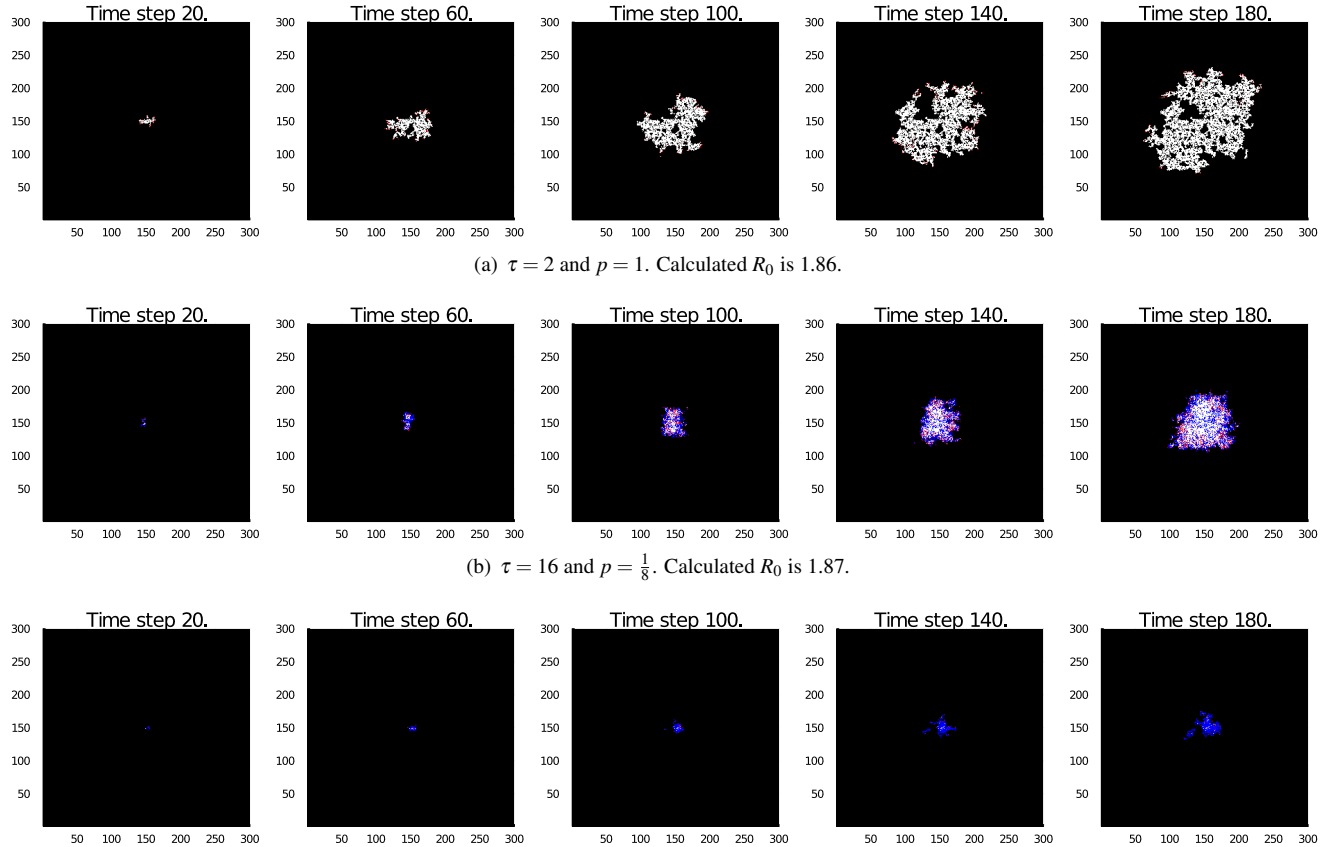
The progression of the outbreak beyond the initial stage is illustrated by Figure 3. Depending on the parameters, the outbreak may die out or spread indefinitely, and it is a natural starting point to describe this dependence by a phase diagram.

Phase diagrams are used in physics and related fields as a visualization of how pressure and temperature (or other thermodynamic variables) affect the bulk thermodynamic state of a substance. Here, phase diagrams are used in an analogous manner to provide insight into how p and τ affect the long-term behavior of an epidemic (*i.e.*, how likely it is to die out after a certain amount of time has passed). In contrast to traditional phase diagrams, it is challenging to define discrete phases for the behavior of an epidemic, so the measures used in this section of our study are continuous variables with finite gradients between phases.

Each point in the phase diagram describes the expected behavior of an epidemic with the given p and τ inputs. The data from running 100 realizations of each of these outbreaks were used to produce a histogram of the growth rate (defined as the number of new infections) at the final time step, as shown in Figure 7 in the Appendix. The final time step is chosen by the following function:

$$t(\tau) = \begin{cases} 200 & \text{if } \tau < 20 \\ 10\tau & \text{if } \tau \geq 20 \end{cases} \tag{6}$$

This sliding criterion adjusts for outbreaks with longer τ, *i.e.*, when infective agents have longer lifespans. Notably, the histogram exhibits a large spike at zero new infections, which represents the fact that a significant fraction of outbreaks will die out quickly, since, due to the intrinsic stochasticity of our model, not every infection reaches epidemic-level proportions. From this histogram, a cutoff of 10 new infections at the final time step is used as a proxy for determining whether an outbreak has



(a) $\tau = 2$ and $p = 1$. Calculated R_0 is 1.86.

(b) $\tau = 16$ and $p = \frac{1}{8}$. Calculated R_0 is 1.87.

(c) $\tau = 122$ and $p = \frac{1}{64}$. Calculated R_0 is 1.87.

Figure 3. Progression of the outbreak. Red sites are infective agents, white are removed and black are susceptible. Blue sites are not technically part of the SIR categories but represent susceptible sites which have been previously visited (but not infected). This fourth color brings attention to the pockets of susceptibles left in the wake of the infection. Note that the model exhibits very different behavior for combinations of τ and p with almost identical R_0, demonstrating that R_0 is not always an accurate characterization of the overall behavior of an outbreak. For example, the outbreak in Figure 3(a) grew much more rapidly in the earlier time steps, but by the final time step, the outbreak in Figure 3(b) had a much larger population of active infective agents. The outbreak shown in Figure 3(c) is clearly of much smaller magnitude despite having the same R_0.

τ \ p	1	1/2	1/4	1/8	1/16	1/32	1/64
1	0.99927(2)	0.4997(3)	0.2498(3)	0.1252(2)	0.0627(2)	0.03130(11)	0.01577(8)
2	1.8614(2)	0.9307(4)	0.4658(4)	0.2339(3)	0.1171(2)	0.0583(2)	0.02896(11)
4	3.4590(4)	1.8042(6)	0.9211(5)	0.4658(4)	0.2339(3)	0.1171(2)	0.0586(2)
8	6.3862(7)	3.4885(8)	1.8251(7)	0.9334(6)	0.4723(4)	0.2375(3)	0.1189(2)
16	11.7767(13)	6.7137(11)	3.6020(10)	1.8686(8)	0.9512(6)	0.4802(4)	0.2414(3)
32	21.725(2)	12.8762(16)	7.0810(14)	3.7223(11)	1.910(8)	0.9677(6)	0.4870(4)
64	40.214(3)	24.643(2)	13.8614(19)	7.3885(15)	3.8184(12)	1.9432(9)	0.9793(6)
128	74.711(6)	47.143(3)	27.092(2)	14.629(2)	7.6183(16)	3.8883(12)	1.9663(9)

Table 1. R_0 values calculated over 2 500 000 realizations. The numbers in parenthesis are the standard error in the last digits.

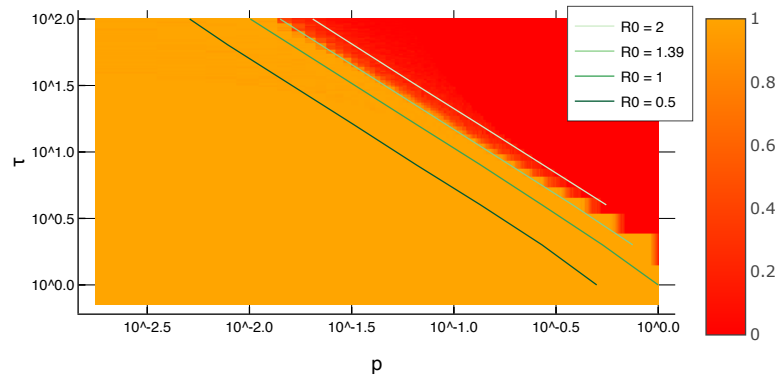


Figure 4. Phase diagram in log-log scale, including iso-R_0 lines. The heatmap scale shows the phase variable of Eq. (7), namely, the fraction of realizations where the outbreak dies out.

died out. The cutoff includes the spike at zero new infections (realizations which have definitively died out) as well as runs where the number of new infections is very small and the end of the epidemic is imminent. Using this criterion, we define the *phase* of an infection by this single measure:

$$\text{phase variable} = \frac{\text{number of realizations with growth rate below cutoff}}{\text{total number of realizations}}. \tag{7}$$

Conveniently, the phase variable is already normalized on the scale 0 to 1, and represents what fraction of epidemics, with a given p and τ, are expected to die out. The plot of this phase variable is shown in Figure 4. The determination of the phase boundary allows us to quickly predict the behavior of the system in the regions above or below the boundary, even in the absence of extensive simulations.

For each horizontal cross-section of the phase diagram (corresponding to a 1D phase diagram for a fixed value of τ), we have fitted the phase variable as a function of p to a smooth interpolating function in order to find the location of the phase boundary, defined as the point with $p = 0.5$. The resulting phase boundary is well approximated by a hyperbola in the p, τ plane. In particular, a least-squares fit gives the curve $p_{\text{boundary}} \approx 2.0\tau_{\text{boundary}}^{-1.1}$.

It is customary in epidemiology[2] to predict the progression of epidemics based on the value of R_0. Accordingly, on Figure 4 "iso-R_0" lines were overlaid on the phase diagram: each point on an iso-R_0 line has p and τ values which produce outbreaks with a common R_0. As expected, the phase boundary indeed coincides with an iso-R_0 line, having $R_0 \approx 1.39$ (and not $R_0 = 1$, because of the role of fluctuations inherent in a particle-based, spatial model). As shown in the Methods section, this finding corresponds to a robust theoretical prediction.

Outbreaks across regional boundaries

In order to complement the mapping out of the phase diagram, the random-walk outbreak model was also used to simulate real-world infectious dynamics. In this section, the effects of having multiple regions with separate p values (representing adjacent states or provinces with different distancing or shelter-in-place policies) is explored. The spatial aspect of this scenario is simplified to two infinite half-planes with distinct p values.

Most notably, the random-walk model corroborates the hypothesis that it is possible for an infection to grow very slowly (constantly on the brink of completely dying out) until it eventually reaches a region with a higher value of p, upon which it exhibits rapid growth. This is demonstrated in Figure 5. Moreover, there is also a significant backflow of infectors returning from the high-p region to the original low-p region.

Attack rate and radius of gyration of the affected zone

In this section, we explore two alternative metrics to characterize the incidence and growth of an epidemic. First, the *attack rate* is generally defined as the proportion of exposed susceptibles that have been infected. Given the set of infected sites, I, the set of removed sites, R, and the set of visited sites (which includes infected and removed ones), V, we may define the attack rate in the context of our lattice-based model. Using $|X|$ to represent the cardinality of set X, attack rate is calculated as

$$\text{attack rate} = \frac{|I| + |R|}{|V|}. \tag{8}$$

In our spatial model, we can also study the growth of the affected area. To this end, a simple and intuitive metric is the

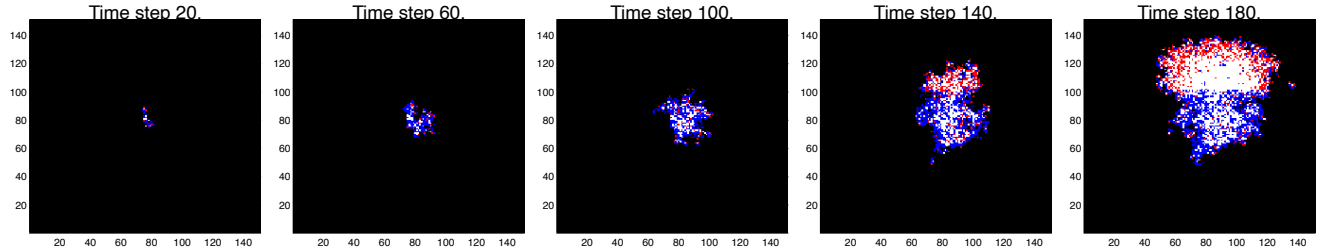


Figure 5. Simulation of an outbreak spanning two regions with distinct p values. The border between the two regions is located at the $y = 100$ line and the infection originates at the center of the grid. Once crossing the boundary to the upper region, the infection explodes. For $y \leq 100$, $p = 0.1$ and for $y > 100$, $p = 0.3$. This illustrates a stark example of the effects of an infection re-igniting upon crossing a regional border.

radius of gyration $\mathscr{R}_{\mathrm{g}}$ of the set of visited sites. If $\mathbf{r}_i$ is the position of the i-th visited site, then

$$\mathscr{R}_g^2 = \frac{1}{|V|} \sum_{i \in V} (\mathbf{r}_i - \mathbf{r}_{\mathrm{CM}})^2, \qquad \mathbf{r}_{\mathrm{CM}} = \frac{1}{|V|} \sum_{i \in V} \mathbf{r}_i. \tag{9}$$

The attack rate and radius of gyration for selected p and τ values along two cross-sections of the phase diagram near the phase boundary were plotted and are shown in Figure 6. These two transects of the phase diagram were chosen to capture the effects of crossing the boundary from different positions. For each transect, 10 different p, τ pairs are sampled and then the attack rate and radius of gyration are calculated for each pair. These computations are averaged from 1000 realizations of the model for 800 time steps. For the fixed-τ interval, $\tau = 50$ and p ranges from 0.01 to 0.10 (in increments of 0.01). For the fixed-p interval, $p = 0.4$ and τ ranges from 1 to 10 (in increments of 1).

Figures 6(b) and 6(d) show how the attack rate increases with p and τ. Notice that for very small values of τ, below the phase boundary, the attack rate can decrease with time. This is because we normalize by the number of visited sites, which is always one at start but increases with time. The attack rate of eq. (8), hence, is initially equal to p but can decrease for dying epidemics as more sites are visited but not infected.

The primary result from figure 6(c), which is shown in log-log scale, is the linear growth of the radius of gyration with time for developing epidemics. The curves for $\tau = 1$, $\tau = 2$, and $\tau = 3$ are below the phase boundary, and, hence, they plateau, indicating that the infection has died out and the radius is no longer increasing. Figure 6(e) shows largely the same results as figure 6(c). However, this plot has a more punctuated change after $t = \tau = 50$, when several infections begin to plateau and die out. This behavior is explained by the low p values, which cause most outbreaks to stop growing after the lifetime of the initial walker expires at $t = 50$.

This linear growth of $\mathscr{R}_{\mathrm{g}}$ translates into a quadratic, rather than exponential, growth for the cumulative incidence of the epidemic, as observed in recent works.[12–14]

Discussion

We see that a simple lattice model provides a number of interesting insights about epidemic spread. One of the most useful features of idealized models like this one is that they provide an understanding of which factors are salient for the outcome, and which are not. This understanding can be used in the construction of more detailed and realistic models, in the same way that a pencil sketch is useful when starting a full-size oil painting.

One of the most important questions when studying epidemics is whether the value of R_0 is sufficient to describe the epidemics behavior, or whether a detailed analysis of the contact network is necessary. Our model shows that the answer depends on what aspect of the epidemic is interrogated. The spatial features of the outbreak are not determined by R_0 alone, but depend on the the individual infectivity and the number of individual contacts separately. In contrast, the phase boundary between infinite spread and localization of the epidemic is mainly determined by R_0, though, as a consequence of having a single index case, not by the canonical value $R_0 = 1$.

A significant idealization in our model is the motion of the infective agents: the random walkers can go anywhere on the lattice and do not have a memory of their origins. In real life, people tend to have permanent residences and return there daily. These periodic movements would work to make the outbreaks even more localized than those in our model. Introducing more realistic walking patterns is, therefore, a compelling direction for future work. Another interesting possibility would be to use the end state of one simulation, with a lattice including both removed and still susceptible sites, as the initial condition to simulate the effect of subsequent epidemic waves. Such a numerical experiment has a direct bearing on the question of thresholds to herd immunity in a fully spatial model, a topic that has received scant attention.

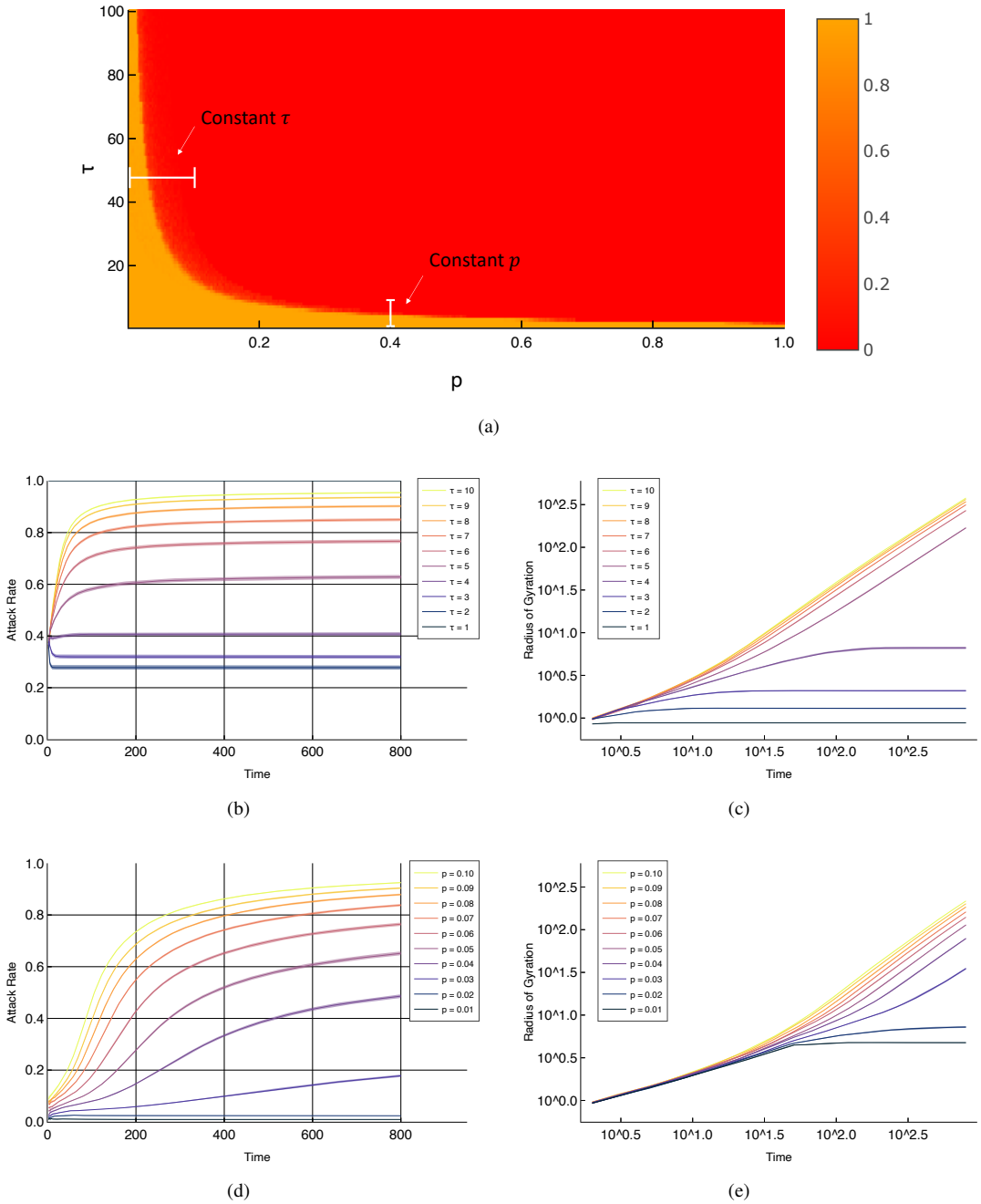


Figure 6. Study of the radius of gyration of the set of infected sites and of the attack rate performed across two transects of the phase diagram. 6(a) Location of the two transects in the phase diagram that are examined in detail. 1000 outbreaks of 800 time steps are simulated for each p and τ combination and the attack rate and radius of gyration are calculated. Note that both of the studied transects cross the boundary between phases. The boundary is located at $\tau = 50, p = 0.0337$ and $\tau = 5, p = 0.4$. 6(b) The attack rate for the fixed-p interval. 6(c) The radius of gyration for the fixed-p interval. 6(d) The attack rate for the fixed-τ interval. 6(e) The radius of gyration for the fixed-τ interval.

Methods

Effect of fluctuations on phase diagram

The phase boundary introduced above separates the cases where an outbreak dies out from those where it spreads indefinitely. Let P_0 be the probability that the outbreak dies out given one index case. We assume that the number of people infected by the index case has a Poisson distribution with mean R_0. The meaning of R_0 is thus the same as in classical epidemiology: the number of people infected by the index case in a naïve population.[2] The probability that the number of secondary infections is zero, and that the outbreak dies out immediately, is $\exp(-R_0)$. The probability that the index case infects exactly n persons is $\exp(-R_0)R_0^n/n!$ If the index case infected n persons, the epidemic dies out only if each secondary outbreak caused by each of these n persons dies out. The probability of such an event P_1 is, strictly speaking, different from P_0 since the secondary infections operate in a different environment than the index case. However, we neglect this and stipulate that $P_1 \approx P_0$. The probability that n secondary outbreaks die out is therefore P_0^n. Then we can write down

$$P_0 = \exp(-R_0) + \sum_{n=1}^{\infty} \exp(-R_0)\frac{R_0^n P_0^n}{n!} = \exp\big(-R_0(1-P_0)\big). \tag{10}$$

At $0 \le R_0 \le 1$ the only solution for this equation is $P = 1$. This means that at $R_0 \le 1$ the epidemics is always localized. At $R_0 > 1$ equation (10) gets a solution between 0 and 1. If we draw the phase boundary at the point where exactly 50% of outbreaks survive, then the critical value of R_0 is $R_0 = 1.39$. This agrees extremely well with Figure 4.

If we have N index cases, the probability that the epidemic is localized is, to first approximation, P_0^N. At large N this probability quickly approaches zero, for $R_0 > 1$. It is important to note that the phase boundary, unlike the other features of the epidemic, depends only on R_0.

Asymptotics for R_0

In this section, we derive a theoretical estimate for R_0: the number of infections produced by one random walker in a susceptible population. We use ideas similar to those in[15–17].

Let us consider a walker that starts at the point $\mathbf{x}$. Let $C(\tau, m, \mathbf{x}, \mathbf{w}, \mathbf{y})$ be the probability for this walker to end up at the point $\mathbf{y}$ after making τ steps, while visiting the point $\mathbf{w}$ exactly m times. The probability that the walker does *not* infect the site $\mathbf{w}$ during this trip is $(1-p)^m$, since the probability to infect at each visit is p. Summing over $\mathbf{w}$, $\mathbf{y}$ and m, we get the total number of infected sites as

$$R_0 = \sum_{\mathbf{w}} \left[1 - \sum_{\mathbf{y}} C'(\tau, p, \mathbf{x}, \mathbf{w}, \mathbf{y})\right] = \sum_{\mathbf{y}} \left[1 - \sum_{\mathbf{w}} C'(\tau, p, \mathbf{x}, \mathbf{w}, \mathbf{y})\right], \tag{11}$$

where

$$C'(\tau, p, \mathbf{x}, \mathbf{w}, \mathbf{y}) = \sum_{m=0}^{\infty} (1-p)^m C(\tau, m, \mathbf{x}, \mathbf{w}, \mathbf{y}). \tag{12}$$

This function is (up to a normalization) the generating function introduced in Rudnick and Gaspari [17, Section 3.2]. To calculate it, we introduce a transformation similar to the transition from the canonical ensemble to the grand canonical ensemble in statistical physics. Namely, let us take a complex variable z and write down the series:

$$H(z, p, \mathbf{x}, \mathbf{w}, \mathbf{y}) = \sum_{\tau=0}^{\infty} z^\tau C'(\tau, p, \mathbf{x}, \mathbf{w}, \mathbf{y}). \tag{13}$$

The inverse transformation is

$$C'(\tau, p, \mathbf{x}, \mathbf{w}, \mathbf{y}) = \frac{1}{2\pi i} \oint \frac{H(z, p, \mathbf{x}, \mathbf{w}, \mathbf{y})}{z^{\tau+1}} dz, \tag{14}$$

where the complex integral is taken around the point $z = 0$. As we shall see below, the value of this integral is defined by the singularity of H at $z \to 1$.

Let $P(\tau, \mathbf{x}, \mathbf{y})$ be the probability for a random walker starting at the point $\mathbf{x}$ to end at the point $\mathbf{y}$. We introduce two functions

$$G(z, \mathbf{x}, \mathbf{y}) = \sum_{\tau=0}^{\infty} z^\tau P(\tau, \mathbf{x}, \mathbf{y}), \qquad G_1(z, \mathbf{x}, \mathbf{y}) = \sum_{\tau=1}^{\infty} z^\tau P(\tau, \mathbf{x}, \mathbf{y}). \tag{15}$$

A remarkable theorem [17, Section 3.3] states that function H can be expressed through the functions G and G_1:

$$H(z, p, \mathbf{x}, \mathbf{w}, \mathbf{y}) = G(z, \mathbf{x}, \mathbf{y}) - \frac{G(z, \mathbf{x}, \mathbf{w}) G(z, \mathbf{w}, \mathbf{y})}{1 + pG_1(z, \mathbf{w}, \mathbf{w})}. \tag{16}$$

Integration of G over y and w is easy due to the normalization. The only nontrivial contribution is from the function G_1 in the denominator of this expression. In two dimensions, this function can be expressed as [17, Section 3.5]

$$G_1(z,\mathbf{w},\mathbf{w}) = \frac{1}{(2\pi)^2}\int \frac{d\mathbf{q}}{1-z\chi(\mathbf{q})} - 1, \tag{17}$$

where $\chi(\mathbf{q})$ is the structure factor, *i.e.* the Fourier transform of the jump probability function. More precisely, let $p(\mathbf{x})$ be the probability for the walker to go from the point $\mathbf{0}$ to the point $\mathbf{y}$ in one step. We then define the structure factor $\chi(\mathbf{q})$ as

$$\chi(\mathbf{q}) = \sum_{\mathbf{x}} p(\mathbf{x}) e^{i\mathbf{x}\mathbf{q}}. \tag{18}$$

At small q, we can expand the structure factor as

$$\chi(\mathbf{q}) = 1 - cq^2 + \mathscr{O}(q^4), \tag{19}$$

which gives near the singularity $z \to 1$

$$G_1(z,\mathbf{w},\mathbf{w}) \approx \frac{1}{4\pi c}\ln\frac{cq_0^2}{1-z}. \tag{20}$$

Here $q_0 \approx 1$ is the upper limit in the integral (17) (assuming the lattice constant to be 1). Using these asymptotics, and integrating equations (11), (14), and (16), we get equation (4) with

$$K = \frac{1}{4\pi c}, \qquad \tau_0 \approx \frac{1}{c}. \tag{21}$$

Note that we use a different normalization from the one used in the Rudnick and Gaspari[17]: they count the number of random walks, whereas we count the probability. This means that the singularities of the functions H, G and G_1 are at $z = 1$ instead of $z = z_c$, and that $\chi(0) = 1$.

The coefficients K and τ_0 depend only on the structure factor χ. To calculate these two, let us generate a large number N of jumps according to equations (2) and (3). Let us take $\mathbf{q}$ along the x axis. If jump number i lands at the point $\mathbf{r}_i = (x_i, y_i)$, then it contributes the term

$$\frac{1}{N}e^{ix_i q} = \frac{1}{N}\left(1 - ix_i q - \frac{x_i^2}{2} + \dots\right) \tag{22}$$

to χ. Due to symmetry, the contribution of the second term is zero, and we arrive at

$$c = \frac{1}{2N}\sum_i x_i^2. \tag{23}$$

In our experiments, we generate $N = 10^6$ points, which gives $c \approx 0.458$, and the resulting values of K and τ_0 in equation (5).

Data availability

Code and data for figures and tables can be found at https://github.com/czbiohub/random-walk-epidemic-model.

References

1. Allen, L. J. S. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infect. Dis. Model.* **2**, 128–142, DOI: https://doi.org/10.1016/j.idm.2017.03.001 (2017).
2. Hethcote, H. W. The mathematics of infectious diseases. *SIAM Rev.* **42**, 599–653, DOI: 10.1137/S0036144500371907 (2000).
3. Grijalva, C. G. *et al.* Transmission of SARS-COV-2 infections in households — Tennessee and Wisconsin, April–September 2020. *Morb. Mortal. Wkly. Rep.* **30 October**, DOI: 10.15585/mmwr.mm6944e1 (2020).
4. Huber, G. *et al.* A minimal model for household effects in epidemics. *Phys. Biol.* **17**, 065010, DOI: 10.1088/1478-3975/abb209 (2020).
5. Chang, S. Y. *et al.* Mobility network modeling explains higher SARS-CoV-2 infection rates among disadvantaged groups and informs reopening strategies. *medRxiv* DOI: 10.1101/2020.06.15.20131979 (2020). https://www.medrxiv.org/content/early/2020/08/14/2020.06.15.20131979.full.pdf.

6. Aleta, A. *et al.* Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971, DOI: 10.1038/s41562-020-0931-9 (2020).

7. Grassberger, P. Two-dimensional SIR epidemics with long range infection. *J. Stat. Phys.* **153**, 289–311, DOI: 10.1007/s10955-013-0824-7 (2013).

8. Hallatschek, O. & Fisher, D. S. Acceleration of evolutionary spread by long-range dispersal. *Proc. Natl. Acad. Sci.* **111**, E4911–E4919, DOI: 10.1073/pnas.1404663111 (2014). https://www.pnas.org/content/111/46/E4911.full.pdf.

9. Fisher, M. E., Ma, S.-k. & Nickel, B. G. Critical exponents for long-range interactions. *Phys. Rev. Lett.* **29**, 917–920 (1972).

10. Stephen, M. J. & Aharony, A. Percolation with long-range interactions. *J. Phys. C: Solid State Phys.* **14**, 1665–1670 (1981).

11. Bunde, A., Herrmann, H. J., Margolina, A. & Stanley, H. E. Universality classes for spreading phenomena: A new model with fixed static but continuously tunable kinetic exponents. *Phys. Rev. Lett.* **55**, 653–656 (1985).

12. Ziff, A. L. & Ziff, R. M. Fractal kinetics of COVID-19 pandemic. *medRxiv* 8, DOI: 10.1101/2020.02.16.20023820 (2020).

13. Singer, H. M. The COVID-19 pandemic: growth patterns, power law scaling, and saturation. *Phys. Biol.* **17**, 055001, DOI: 10.1088/1478-3975/ab9bf5 (2020).

14. Merrin, J. Differences in power law growth over time and indicators of COVID-19 pandemic progression worldwide. *Phys. Biol.* **17**, 065005, DOI: 10.1088/1478-3975/abb2db (2020).

15. Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E. & Weiss, G. H. Number of distinct sites visited by *N* random walkers. *Phys. Rev. A* **45**, 7128–7138, DOI: 10.1103/PhysRevA.45.7128 (1992).

16. Larralde, H., Trunfio, P., Havlin, S., Stanley, H. E. & Weiss, G. H. Territory covered by n diffusing particles. *Nature* **355**, 423–426 (1992).

17. Rudnick, J. & Gaspari, G. *Elements of the Random Walk: An Introduction for Advanced Students and Researchers* (Cambridge University Press, 2004).

Acknowledgments

A.C., G.H., A.M., and D.Y. are supported by the Chan Zuckerberg Biohub; B.V. is supported by the Chan Zuckerberg Initiative.

Author contributions statement

G.H. and A.M. designed research; A.C. performed research with guidance from G.H., A.M. and D.Y.; A.M. wrote the simulation code; B.V. carried out analytical computations; all authors analyzed data and contributed to the writing of the paper.

Additional information

Competing interests: The authors declare no competing interests.

Appendix

Figure 7 shows an example of the histograms used to draw our phase diagram.

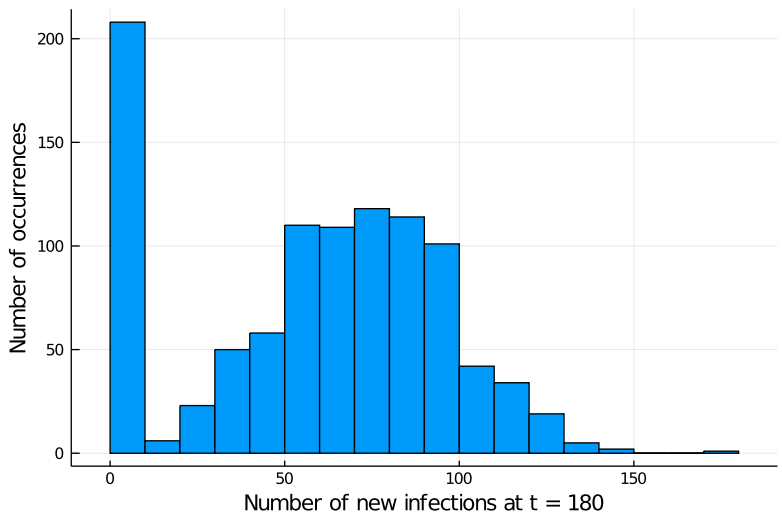


Figure 7. Histogram of epidemic growth rate across several realizations of the outbreak for $p = 0.5$ and $\tau = 4$. Note that 1000 independent runs are used here to provide additional detail for the reader, but the model uses 100 runs for computational efficiency.