Acceleration of evolutionary spread by long-range dispersal

Oskar Hallatschek[a,1] and Daniel S. Fisher[b]

[a]Biophysics and Evolutionary Dynamics Group, Departments of Physics and Integrative Biology, University of California, Berkeley, CA 94720; and [b]Departments of Applied Physics, Bioengineering, and Biology, Stanford University, Stanford, CA 94305

Edited* by Herbert Levine, Rice University, Houston, TX, and approved September 2, 2014 (received for review March 12, 2014)

The spreading of evolutionary novelties across populations is the central element of adaptation. Unless populations are well mixed (like bacteria in a shaken test tube), the spreading dynamics depend not only on fitness differences but also on the dispersal behavior of the species. Spreading at a constant speed is generally predicted when dispersal is sufficiently short ranged, specifically when the dispersal kernel falls off exponentially or faster. However, the case of long-range dispersal is unresolved: Although it is clear that even rare long-range jumps can lead to a drastic speedup—as air-traffic–mediated epidemics show—it has been difficult to quantify the ensuing stochastic dynamical process. However, such knowledge is indispensable for a predictive understanding of many spreading processes in natural populations. We present a simple iterative scaling approximation supported by simulations and rigorous bounds that accurately predicts evolutionary spread, which is determined by a trade-off between frequency and potential effectiveness of long-distance jumps. In contrast to the exponential laws predicted by deterministic "mean-field" approximations, we show that the asymptotic spatial growth is according to either a power law or a stretched exponential, depending on the tails of the dispersal kernel. More importantly, we provide a full time-dependent description of the convergence to the asymptotic behavior, which can be anomalously slow and is relevant even for long times. Our results also apply to spreading dynamics on networks with a spectrum of long-range links under certain conditions on the probabilities of long-distance travel: These are relevant for the spread of epidemics.

long-range dispersal | selective sweeps | epidemics | range expansions | species invasions

Humans have developed convenient transport mechanisms for nearly any spatial scale relevant to the globe. We walk to the grocery store, bike to school, drive between cities, or take an airplane to cross continents. Such efficient transport across many scales has changed the way we and organisms traveling with us are distributed across the globe (1–5). This has severe consequences for the spread of epidemics: Nowadays, human infectious diseases rarely remain confined to small spatial regions, but instead spread rapidly across countries and continents by travel of infected individuals (6).

Besides hitchhiking with humans or other animals, small living things such as seeds, microbes, or algae are easily caught by wind or sea currents, resulting in passive transport over large spatial scales (7–16). Effective long-distance dispersal is also widespread in the animal kingdom, occurring when individuals primarily disperse locally but occasionally move over long distances. And such animals, too, can transport smaller organisms.

These active and passive mechanisms of long-range dispersal are generally expected to accelerate the growth of fitter mutants in spatially extended populations. However, how can one estimate the resulting speedup and the associated spatiotemporal patterns of growth? When dispersal is only short range, the competition between mutants and nonmutated ("wild-type") individuals is local, confined to small regions in which they are both present at the same time. As a consequence, a compact mutant population emerges that spreads at a constant speed, as first predicted by Fisher (17) and Kolmogorov et al. (18): Such selective sweeps are slow and dispersal is limited. In the extreme opposite limit in which the dispersal is so rapid that it does not limit the growth of the mutant population, the competition is global and the behavior the same as for a fully mixed (panmictic) population: Mutant numbers grow exponentially fast. It is relevant for our purposes to note that in both the short-range and extreme long-range cases, the dynamics after the establishment of the initial mutant population are essentially deterministic.

When there is a broad spectrum of distances over which dispersal occurs, the behavior is far more subtle than that of either of the well-studied limits. When a mutant individual undergoes a long-distance dispersal event—a jump—from the primary mutant population into a pristine population lacking the beneficial mutation, this mutant can found a new satellite subpopulation, which can then expand and be the source of further jumps, as shown in Fig. 1 *B* and *C*. Consequently, long-range jumps can dramatically increase the rate of growth of the mutant population. Potentially, even very rare jumps over exceptionally large distances could be important (14). If this is the case, then the stochastic nature of the jumps that drive the dynamics will be essential.

Although evolutionary spread with long-range jumps has been simulated stochastically in a number of biological contexts (6, 8, 19–23), few analytic results have been obtained on the ensuing stochastic dynamics (19, 24–26). Most analyses have resorted to deterministic approximations (27–35), which are successful for describing both the local and global dispersal limits. However, in between these extreme limits, stochasticity drastically changes the spreading dynamics of the mutant population. This is particularly striking when the probability of jumps decays as a power law of the distance. Just such a distance spectrum of dispersal is

Significance

Pathogens, invasive species, rumors, or innovations spread much more quickly around the world nowadays than in previous centuries. The speedup is caused by more frequent long-range dispersal, for example via air traffic. These jumps are crucial because they can generate satellite "outbreaks" at many distant locations, thus rapidly increasing the total rate of spread. We present a simple intuitive argument that captures the resulting spreading patterns. We show that even rare long-range jumps can transform the spread of simple epidemics from wave-like to a very fast type of "metastatic" growth. More generally, our approach can be used to describe how new evolutionary variants spread and thus improves our predictive understanding of the speed of Darwinian adaptation.

Author contributions: O.H. and D.S.F. designed research, performed research, contributed analytic tools, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: ohallats@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1404663111/-/DCSupplemental.

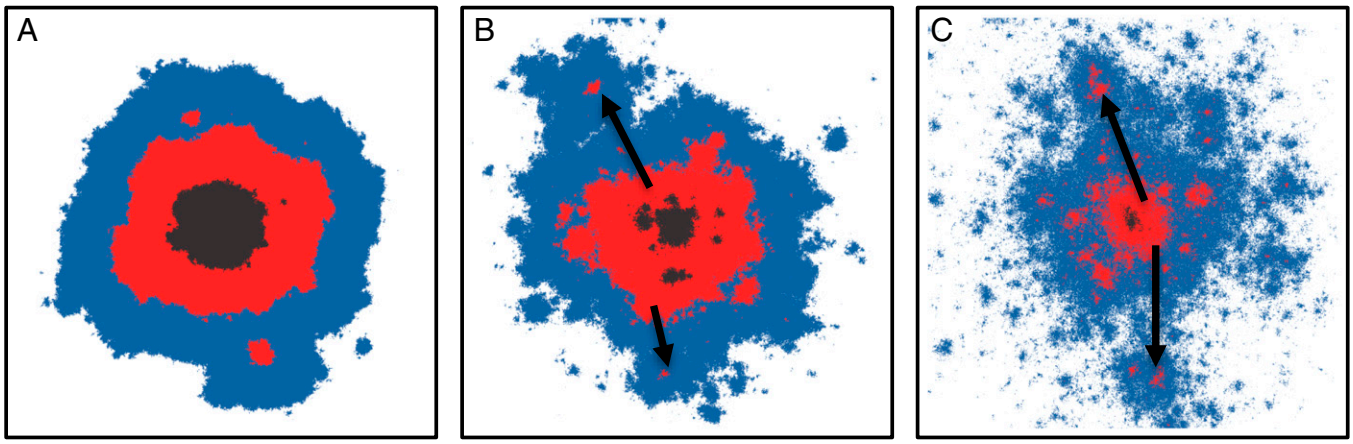

Fig. 1. Evolutionary spread sensitively depends on the dispersal behavior of individuals. For a broad class of models with "short-range migration," the mutant subpopulation expands at a constant speed that characterizes the advance of the mutant–nonmutant boundary. With long-range dispersal the spread is much faster. Shown are 2D simulations for the simple case of a jump distribution that has a broad tail characterized by a power-law exponent $-(\mu+2)$. *A–C* show the distribution of the mutant population at the time when half of the habitat is occupied by mutants. The color of a site indicates whether a site was filled in the first (black), second (red), or last (blue) third of the total run time. (*A*) When the jump distribution decays sufficiently rapidly, $\mu > 3$, the asymptotic growth resembles, in two dimensions, a disk growing at a μ–dependent constant speed. (*B*) For $2 < \mu < 3$, satellite seeds become clearly visible and these drive superlinear power-law growth. These seeds were generated by long-range jumps, as indicated by arrows. (*C*) The dynamics are changed drastically for $0 < \mu < 2$, becoming controlled by very long-distance jumps, which seed new expanding satellite clusters. As a result of this "metastatic" growth, the spreading is faster than any power law, although markedly slower than exponential. *A–C* were created by the simulations described in the main text, with parameters $\mu = 3.5$ in *A*, $\mu = 2.5$ in *B*, and $\mu = 1.5$ in *C*. (To avoid any boundary effects, the lattice size was chosen to be much larger than depicted regions.)

characteristic of various biological systems (36–41), including humans (3). We will show that the behavior is controlled by a tradeoff between frequency and potential effectiveness of long-distance jumps and the whole spectrum of jump distances can matter. The goal of this paper is to develop the theory of stochastic spreading dynamics when the dispersal is neither short range nor global.

Long-distance dispersal can occur either on a fixed network or more homogeneously in space. For simplicity, we focus on the completely homogeneous case and then show that many of the results also apply for an inhomogeneous transportation network with hubs between which the long-distance jumps occur. For definiteness, we consider for most of the paper the evolutionary scenario of the spread of a single beneficial mutation, but, by analogy, the results can be applied to other contexts, such as the spread of infectious disease or of invasive species.

Basic Model

The underlying model of spatial spread of a beneficial mutant is a population in a d-dimensional space with local competition that keeps the population density constant at $\hat{\rho}$ and equipped with a probability that any individual jumps to any particular point a distance r away of $J(r)$ per time per area, per length, or per volume. At a very low rate, mutants can appear that have a selective advantage, s, over the original population. A lattice version of this model is more convenient for simulations (and for aspects of the analysis): Each lattice site represents a "deme" with fixed population size, $\hat{n} \gg 1/s$ with the competition only within a deme and the jump migration between demes. Initially, a single mutant occurs and if, as occurs with probability proportional to s, it survives stochastic drift to establish, it will take over the local population. When the total rate of migration between demes is much slower than this local sweep time, the spatial spread is essentially from demes that are all mutants to demes that are all of the original type.

Short jumps result in a mutant population that spreads spatially at a roughly constant rate. However, with long-range jumps, new mutant populations are occasionally seeded far away from the place from which they came, and these also grow. The consequences of such long jumps are the key issues that we need to understand. As we shall see, the interesting behaviors can be conveniently classified when the jump rate has a power-law tail at long distances, specifically, with $J(r) \sim 1/r^{d+\mu}$ (with positive μ needed for the total jump rate to be finite). Crudely, the behavior can be divided into two types: linear growth of the radius of the region that the mutants have taken over and faster than linear growth. In Fig. 1, these two behaviors are illustrated via simulations on 2D lattices. In addition to the mutant-occupied region at several times, shown are some of the longest jumps that occur and the clusters of occupied regions that grow from these. In Fig. 1*A*, there are no jumps that are of comparable length to the size of the mutant region at the time at which they occur, and the rate of growth of the characteristic linear size $\ell(t)$ of the mutant region—loosely its radius—is roughly constant in time; i.e., $\ell(t) \sim t$. In Fig. 1 *B* and *C*, $J(r)$ is longer range and very long jumps are observed. These result in faster-than-linear growth of the radius of the mutant region, as shown. Before developing analytic predictions for the patterns of evolutionary spread, we report our simulation results in detail.

Results

Simulated Spreading Dynamics. We have carried out extensive simulations of a simple lattice model, in which the sites form either a one-dimensional, of length L, or two-dimensional $L \times L$ square array. Boundary conditions are chosen to be periodic, but typically do not matter unless the filling fraction becomes of order one. As it is the spreading dynamics at long times that we are interested in, we assume that the local sweeps in a deme are fast compared with migration. We can then ignore the logistic growth process within demes, so that when jumps occur and establish a new mutant population, it is saturated in the new deme by the next time step. Therefore, it is convenient to lump together the probability of an individual to jump, the density of the population from which the jumps occur, and the probability (proportional to s) that the mutant establishes a new population: We define $G(r) \equiv s\hat{\rho}J(r)$ so that $d^d r d^d r' G(|\mathbf{x}-\mathbf{y}|)$ is the rate at which a saturated mutant population near $\mathbf{x}$ nucleates a mutant population near $\mathbf{y}$. In each computational time step, we pick a source and target site randomly such that their distance r is sampled from the

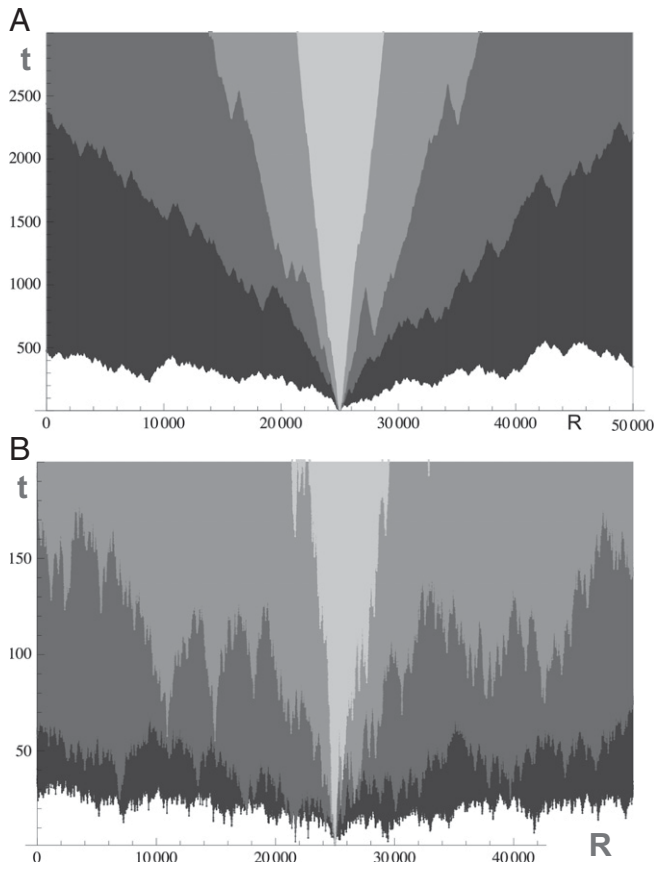

Fig. 2. Stochastic growth of a mutant population over time in one dimension. Each level of shading represents a single simulation run. (*A*) Regime of power-law growth. The values of μ are 2.5, 2.0, 1.75, 1.5 in order of increasing darkness. Note that, in two dimensions, $\mu = 2.0$ corresponds to the marginal case separating linear growth ($\mu > 2$) from superlinear growth ($\mu < 2$). (*B*) Regime of very fast growth with μ near $\mu = 1.0$, which is the marginal case separating power-law growth ($\mu > 1$) from stretched exponential growth ($\mu < 1$). The values of μ are 1.5, 1.25, 1.0, 0.75 in order of increasing darkness.

(discretized) jump distribution $G(r)$—with the $d^d r$ a lattice site. If the source site is a mutant and the target site a wild type, the identity of the target site is updated to mutant. We measure time in units of L^d time steps. The lattice sizes are chosen large enough (up to $L^d \sim 10^9$) so that we can observe the growth dynamics over several orders of magnitude undisturbed by boundary effects. See *SI Text*, section *SI1* for more details on the simulation algorithm.

The growth of mutant populations generated by our simulations is best visualized in a space–time portrait. Fig. 2 *A* and *B* shows the overlaid space–time plots of multiple runs in the regimes $1.5 < \mu < 2$ and $0.8 < \mu < 1.1$, respectively. Fig. 3 shows the growth dynamics of the mutant population over large timescales and length scales for various values of μ. For $\mu \gtrsim 1.4$, the dynamics clearly approach a power law. For $\mu \lesssim 0.7$, the simulations are consistent with stretched exponentials. The intermediate regime $0.7 < \mu < 1.3$ is elusive, as we cannot extract a clear asymptotic behavior on the timescales feasible in simulations. The behavior in two dimensions is qualitatively similar, as shown in Fig. S1. Lowering the rate of long-range jumps compared to short-range jumps between neighboring sites leads to a delayed cross-over to the super-linear regimes, as shown in Fig. S2.

To explain these dynamics in detail, we develop an analytical theory that is able to predict not only the asymptotic growth dynamics but also the crucial transients.

Breakdown of Deterministic Approximation. Traditionally, analyses of spreading dynamics start with a deterministic approximation of the selective and dispersal dynamics—ignoring both stochasticity and the discreteness of individuals. To set up consideration of the actual stochastic dynamics, we first give results in this deterministic approximation and show that these exhibit hints of why they break down.

When the jump rate decreases exponentially or faster with distance, the spread is qualitatively similar to simple diffusive dispersal and the extent of the mutant population expands linearly in time. However, when the scale of the exponential falloff is long, the speed, v, is faster than the classic result for local dispersal (17, 18), $v = 2\sqrt{Ds}$, which depends only on the diffusion coefficient, $D = (1/2d) \int r^2 J(r) d^d r$: Specifically, consider $J(r) = c_d (D/2b^{d+2}) e^{-r/b}$ with coefficient c_d so that the diffusion constant D is independent of the characteristic length, b, of the jumps. A linear deterministic approximation for the mutant population density in a spatial continuum and a saddle point analysis to find the distance at which the population density becomes substantial yield, as for the conventional diffusive case, the correct asymptotic speed. The resulting expression for the speed is modified from the diffusive result by a function of the only dimensionless parameter, $b/\sqrt{D/s}$. For $b \ll \sqrt{D/s}$, $v \approx 2\sqrt{Ds}$, the

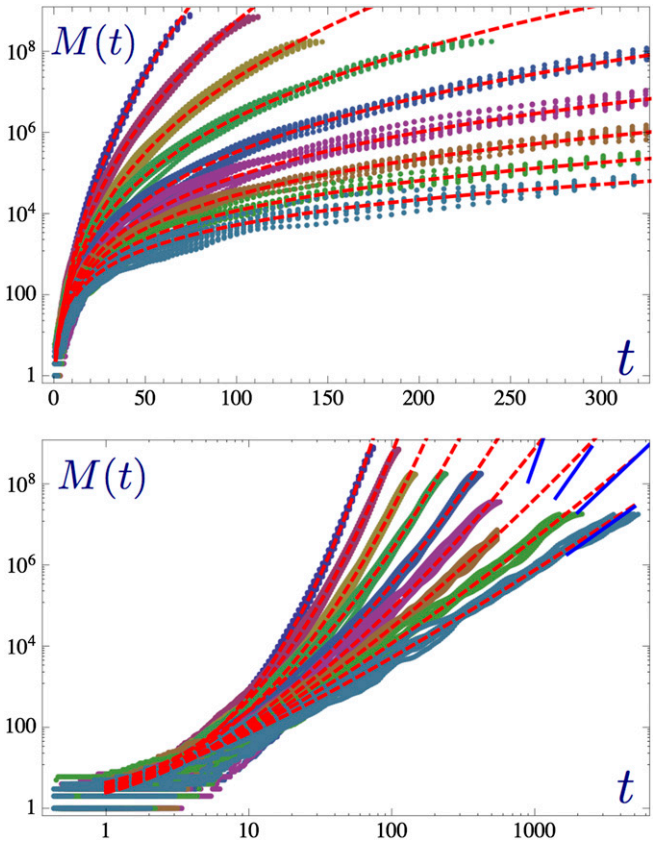

Fig. 3. Summary of the quantitative spreading dynamics in one spatial dimension. The total number, $M(t)$, of mutant sites is plotted as a function of time t, for various long-range jump kernels. Each colored cloud represents data obtained from 10 runs for a given jump kernel. The data are for μ chosen from $\{0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4\}$ from top to bottom. Red dashed lines represent predictions obtained from Eq. **8** with fitted magnitude scales for M and t. In the double-logarithmic plot at the bottom of each plot, short blue lines indicate the predicted asymptotic power-law behavior for $\mu > 1$. For $\mu = 1.1$ and $\mu = 1.2$, the dynamics are still far away from the asymptotic power law, which is indicative of the very slow crossover.

diffusive predictions obtain as expected, but for $b \gg \sqrt{D/s}$, $v \approx bs$, which is much larger.

We now provide a simple argument as to why the deterministic approximation drastically overshoots the stochastic spreading dynamics for broader than exponentially decaying jump kernels $J(r)$. The origin of the very rapid spread is the feeding of the populations far away by direct jumps from near the origin: This immediately produces a finite population density at any location, **R**. After a time of order $1/s$ has passed, the exponential growth of the local population near **R** takes off, proportional to $J(R)e^{st}$, and further jumps to that region no longer matter much. The time at which this local mutant population saturates suggests that the radius of the region taken over by the mutant population, $\ell(t)$, is given simply by $J[\ell(t)]e^{st} \sim 1$. In the deterministic approximation, this is a lower bound as seeding by intermediate jumps (as occurs especially for $b \sim \sqrt{D/s}$ or smaller) can only make the spread faster. For the marginal case of exponential dispersal discussed above, this simple approximation yields $v = bs$, correct for large b (i.e., $\gg \sqrt{D/s}$), implying that very long jumps directly to **R** indeed dominate for large distances R. (For small b, in contrast, the speed is much faster than bs because the spread is dominated by multiple small jumps: The diffusion approximation is then good.)

If $J(r)$ has a longer-than-exponential tail (42), in particular, $J(r) \sim 1/r^{d+\mu}$ (43), the spread in the deterministic approximation becomes exponentially fast: $J[\ell(t)] \approx e^{-st}$ yields $\ell(t) \sim \exp(st/(d+\mu))$. Thus, the total mutant population is $\sim \exp(dst/(d+\mu))$. This grows almost as fast as in a fully mixed population, with the population growth rate slower only by a factor of $d/(d+\mu)$. This factor approaches unity as $\mu \to 0$, the point at which the spatial structure becomes irrelevant as the jumps typically span the full system.

We can now understand why the deterministic approximation fails miserably for very long-range jumps. In the time, $\sim 1/s$, during which the jumps into a region from near the origin are supposed to lead to exponential growth of the local population even a large distance R, away, the expected total number of jumps to the whole region R or farther from the origin is only of order $1/(sr^{\mu})$: Thus the probability that any jumps have occurred is very small for large R and the deterministic approximation must fail (19, 44).

With local dispersal, the deterministic approximation is a good starting point with only modest corrections to the expansion speed at high population density, the most significant effect of stochasticity being fluctuations in the speed of the front (45). At the opposite extreme of jump rate independent of distance, the deterministic approximation is also good with the mutant population growing as ae^{st} and fluctuations causing only stochastic variability and a systematic reduction in the coefficient, a: These arise from early times when the population is small. It is thus surprising that in the regimes intermediate between these two, the deterministic approximation is not even qualitatively reasonable.

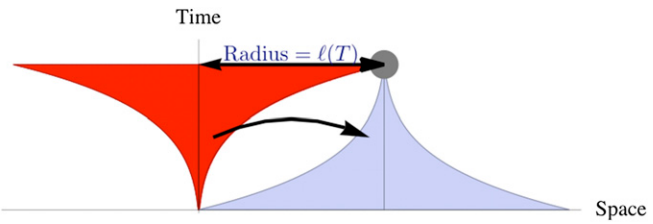

Fig. 4. Sketch of the growth of the compact core of a cluster (red) due to long-range jumps. For the (gray) point at distance $\ell(t)$ to be occupied at time t, a seed typically must become established somewhere in the blue space–time region ("target funnel") by means of a long-range jump (black arrow) from the red "source" region. This schematic leads to the iterative scaling approximation in Eq. **1**. Note that the concavity of the source-funnel geometry, leading to a gap between red and blue regions, is key to our arguments and enables neglecting effects of jumps into the gap region.

Iterative Scaling Argument. We assume that, at long times, most of the sites are filled out to some distance scale $\ell(t)$ and that the density decreases sufficiently steeply for larger distances, such that the total mutant population, $M(t)$, is proportional to $\ell^d(t)$. The validity of this assumption follows from more accurate analyses given in *SI Text*, section *SI3*. We call the crossover scale $\ell(t)$ the core radius or "size" of the mutant population.

In the dynamical regimes of interest, the core population grows primarily because it "absorbs" satellite clusters, which themselves were seeded by jumps from the core population. We now show that the rate of seeding of new mutant satellite clusters and the growth of the core populations by mergers with previously seeded clusters have to satisfy an iterative condition that enables us to determine the typical spreading dynamics of the mutant population.

It is convenient to illustrate our argument using a space–time diagram, Fig. 4, in which the growth of the core has the shape of a funnel. Now consider the edge of this funnel at time T (gray circle in Fig. 4). The only way that this edge can become populated is by becoming part of a population subcluster seeded by an appropriate long-range jump at an earlier time. To this end, the seed of this subcluster must have been established somewhere in the inverted blue funnel in Fig. 4. This "target" funnel has the same shape as the space–time portrait of the growing total population, but its stem is placed at $(\ell(T), T)$ and the mouth opens backward in time. Note that if $\ell(t)$ grows faster than linearly, space–time plots of the growing cluster and the funnel have concave boundaries: This necessitates a jump from the source to the funnel of length much longer than $\ell(T/2)$, as shown.

Now, we argue the consistency of growth and seeding requires that there is, on average, about one jump from the source to the target funnel: If it were unlikely that even one jump leads from the source to the target region, then the assumed shape for the source funnel would be too large and its edge (gray circle in Fig. 4) would, typically, not be occupied. Conversely, if the expected number of jumps was much larger than 1, then seeding would occur so frequently that a much larger funnel than the assumed one would typically be filled by the time T.

The condition of having, on average, about one jump from source to target region can be stated mathematically as

$$\int_0^T dt \int_{\mathcal{B}_{\ell(t)}} d^d\mathbf{x} \int_{\mathcal{B}_{\ell(T-t)}} d^d\mathbf{y}\, G(|\ell(T)\mathbf{e}+\mathbf{y}-\mathbf{x}|) \sim 1 \qquad [1]$$

in d dimensions, where $\mathcal{B}_\ell$ denotes a d-dimensional ball of radius ℓ centered at the origin, and we have taken the point of interest to be $\mathbf{R} = \mathbf{e}\ell(T)$ with $\mathbf{e}$ a unit vector in an arbitrary direction. The kernel $G(r)$ represents the rate per d-dimensional volume of (established) jumps of size r. Eq. **1** mathematizes the space–time picture of Fig. 4: To calculate the expected number of jumps from the red source to the blue target funnel, we need a time integral, $\int_0^T dt$, and two space integrals—one for the source funnel, $\int_{\mathcal{B}_{\ell(t)}} d^d\mathbf{x}$, and one for the target funnel, $\int_{\mathcal{B}_{\ell(T-t)}} d^d\mathbf{y}$—over the probability density $G(|\ell(T)\mathbf{e}+\mathbf{y}-\mathbf{x}|)$ to jump from the source point $\mathbf{x}$ to the target point $\ell(T)\mathbf{e}+\mathbf{y}$.

Note that the above intuitive picture is further sharpened in *SI Text*, section *SI2* and justified by developing rigorous bounds in *SI Text*, section *SI3*.

Asymptotic Results for Power-Law Jumps. We now show that the asymptotic growth dynamics are essentially constrained by the

above iterative scaling argument. Specifically, although the argument is more general, we consider a power-law jump distribution

$$G(r) \approx \frac{\epsilon}{r^{d+\mu}} \quad \text{with} \quad \mu < d+1 \qquad [2]$$

for large enough r. The prefactor ϵ in $G(r)$ amalgamates all of the factors that determine the rate density of successful jumps, including population density and establishment probability.

For the intermediate-range case $d < \mu < d+1$, Eqs. **1** and **2** exhibit the asymptotic scaling solution

$$\ell(t) \sim A_\mu (\epsilon t)^{1/(\mu-d)}, \qquad [3]$$

the form of which could have been guessed by dimensional analysis. Inserting the ansatz **[3]** into Eq. **1** determines the prefactor A_μ in this iterative scaling approximation, up to an order-unity coefficient: Details are in *SI Text*, section *SI2.B*. Interestingly, the value A_μ depends very sensitively on μ and runs from 0 to ∞ as μ passes through the interval from d to $d+1$. This can be seen in Fig. S3, where A_μ is plotted as a function of μ for $d=1$: It drops very steeply, as $A_\mu \sim 2^{-2/(\mu-1)^2}$, for $\mu \searrow 1$, and diverges as $A_\mu \sim 1/(2-\mu)$ for $\mu \nearrow 2$. As we discuss below, these singularities are a manifestation of intermediate asymptotic regimes that lead to very slow convergence to the asymptotic behavior.

We now turn to the (very) long-range case, $0 < \mu < d$, for which a direct solution to **[1]** cannot be found (and the dimensional analysis argument gives nonsense). However, much can be learned by approximating **[1]** using $G[|\ell(T)\mathbf{e} + \mathbf{y} - \mathbf{x}|] \approx G[\ell(T)]$, anticipating the very rapid growth and thus likely smallness of x and y compared with $\ell(T)$:

$$G[\ell(T)] \int_0^T dt\, \ell^d(t) \ell^d(T-t) \sim 1 \qquad [4]$$

(ignoring two factors from the angular integrations). With $\ell(t)$ growing subexponentially, the largest contributions will come from $t \approx (1/2)T$, which reflects the approximate time reversal symmetry of the source and target funnels in Fig. 4. Once we have found the form of $\ell(t)$, the validity of the ansatz can be tested by checking whether the $\ell(T)$ is much larger than $\ell(T/2)$. Indeed, for $\mu < d$ the solution to **[4]** is a rapidly growing stretched exponential,

$$\ell(t) \sim \exp\left(B_\mu t^\eta\right) \quad \text{with} \quad \eta = \frac{\log[2d/(d+\mu)]}{\log 2} < 1, \qquad [5]$$

which can be checked by direct insertion into Eq. **4**. Note that as $\mu \searrow 0$, $\eta \nearrow 1$ and $\ell(t)$ grows exponentially as for a flat distribution of long-range jumps that extends out to the size of the system: i.e., the globally mixed limit. In the opposite limit of $d-\mu$ small, $\eta \sim d-\mu$ and the coefficient B_μ diverges as shown below. We note that the asymptotic stretched-exponential growth for $\mu < d$ also arises in models of "chemical distance" and certain types of spatial spread for network models with a similar power-law distribution of long-distance connections: However, the prefactors in the exponent are different (24, 46). We discuss the connections between these in *SI Text*, section *SI3*.

For the marginal case, $\mu = d$, the asymptotic behavior is similarly found to be

$$\ell(t) \sim \exp\left[\frac{\log^2(t)}{4d\log(2)}\right]. \qquad [6]$$

We show below that this behavior also represents an important intermediate asymptotic regime that dominates the dynamics over a wide range of times for μ close to d: This is the source of the singular behavior of the coefficients A_μ and B_μ in this regime.

Our source-funnel argument obviously neglects jumps that originate from the not fully filled regions outside the core radius $\ell(t)$. An improved version of the funnel argument is presented in *SI Text*, section *SI2.E*, which also allows us to estimate the probability of occupancy outside the core region. Further, we present in *SI Text*, section *SI3* outlines of rigorous proofs of lower and upper bounds for the asymptotic growth laws in one dimension, including the slow crossovers near the marginal case. The linear growth for $\mu > 3$ in one dimension has been proved by Mollison (19). After the present paper was essentially complete, we became aware of a recent preprint by Chatterjee and Dey (26) who obtained rigorous bounds in all dimensions for the leading asymptotic behaviors in the three superlinear regimes. Our bounds are considerably tighter than theirs, including the coefficient and leading corrections to the asymptotic behavior in the marginal case and the absence of logarithmic prefactors in the power-law regime: Comparisons are discussed in *SI Text*, section *SI3*. More importantly, our bounds are explicitly for the iterative scaling analyses, thereby justifying them (and future uses of them). Our bounds hence include the full crossovers for μ near d (in one dimension), rather than just the asymptotic results: As we next show, understanding these crossovers is essential.

Crossovers and Beyond Asymptopia. Asymptotic laws are of limited value without some understanding of their regime of validity, especially if the approach to the asymptotic behavior is slow. And such knowledge is crucially needed to interpret and make use of results from simulations.

Assuming that long jumps are typically much rarer than short jumps, they will become important only after enough time has elapsed that there have been at least some jumps of order $\ell(t)$. This condition can be used to determine a crossover time and length scale from linear to superlinear growth (*SI Text*, section *SI2.A*). For the purpose of this section, it is convenient to measure time and length in units of these elementary crossover scales.

At times much longer than 1 (in rescaled time), we expect another slow crossover close to the boundary between the stretched-exponential and power-law regimes. Thus, we must take a closer look at the dynamics in the vicinity of the marginal case, $\mu = d$. As $\mu \searrow d$, the integrand in Eq. **4** develops a sharp peak at $t = T/2$: half-way between the bounds of the integral. Laplace's method can then be used to approximate the integral leading to a simplified recurrence relation for the rescaled core radius $\ell(t)$,

$$\ell^{d+\mu}(t) \sim t\, \ell\left(\frac{t}{2}\right)^{2d}. \qquad [7]$$

This is good only to a numerical factor of order unity that can be eliminated by rescaling t and to a larger logarithmic factor associated with the narrowness of the range of integration and its dependence on t and $\mu - d$. The associated subdominant corrections, analyzed in *SI Text*, section *SI2.C*, are negligible if we focus on the behavior on logarithmic scales in space and time—natural given their relationships. Defining $\varphi \equiv \log_2(\ell)$ and $z \equiv \log_2(t)$ and taking the binary logarithm, $\log_2$, of Eq. **7** yields a linear recurrence relation that can be solved exactly (*SI Text*, section *SI2.B*). Rescaling ℓ can be used to make $\varphi(0) = 0$, whence

$$\frac{\delta^2}{2d}\varphi(z) \approx \frac{\delta z}{2d} + \left(1 + \frac{\delta}{2d}\right)^{-z} - 1, \qquad [8]$$

where we introduced the variable $\delta = \mu - d$, which measures the distance to the marginal case.

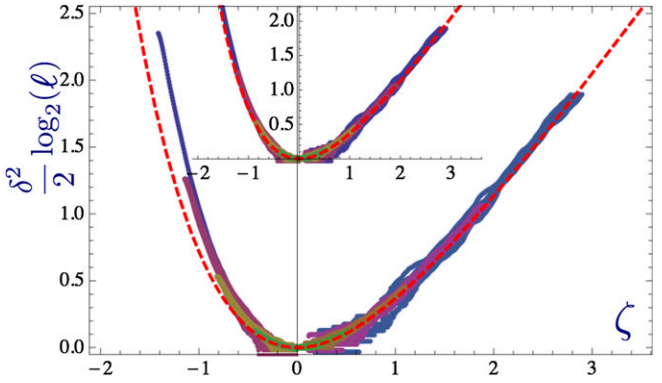

Fig. 5. Data for different $\mu = d + \delta$ are predicted to collapse on a scaling plot close to the marginal case $\mu = d$ between the stretched-exponential and power-law growth regimes, here demonstrated for the one-dimensional case ($d = 1$). The main plot shows the rescaled sizes $\delta^2 \log_2(\ell)/2$ of the mutant population vs. rescaled log-time $\zeta \equiv \delta \log_2(t)/2$. The differently colored datasets correspond to 10 realizations with power-law exponents μ chosen from {0.6, 0.7, 0.8, 0.9, 1.1, 1.2, 1.3, 1.4} (same colors as in Fig. 3, *Upper*). The simulated data collapse reasonably well with the dashed red line, representing the predicted scaling function $\chi(\zeta) = \exp(-\zeta) + \zeta - 1$. *Inset* depicts the same data scaled slightly differently away from the scaling regime such that the horizontal axis for $\delta < 0$ shows $\eta \log(t)$ (with η defined in [5]), upon which the data collapse improves. Note that $\eta \log(t) \approx -\zeta$ as $\delta \to 0$ so that the stretched exponential form is recovered for ζ large and negative.

The asymptotic scaling for $\delta > 0$ reproduces the earlier predicted power-law regime [3] and yields the prefactor $\log A_\mu \approx -2d \log(2)/\delta^2$, up to correction that is subdominant for small δ (compare Fig. S3). For $\delta < 0$, the asymptotics yields the stretched exponential in [5] and fixes the prefactor $B_\mu \approx 2d \log(2) \delta^{-2}$, which could not be obtained from the basic asymptotic analysis carried out above.

The singular prefactors for $\delta \to 0$ give warnings of breakdown of the asymptotic results except at very long times. This peculiar behavior is the consequence of an intermediate asymptotic regime that dominates the dynamics close to the marginal case. This leads to slow convergence to the eventual asymptotic behavior for μ near d. The asymptotic scaling can be observed only on times and length such that

$$\log_2(t) \gg \frac{2d}{|\delta|} \quad \text{and} \quad \log_2(\ell) \gg \frac{2d}{\delta^2}. \qquad [9]$$

On smaller times, the dynamics are similar to those of the marginal case [6]. The rapid divergence of the logarithm of the time after which the asymptotic results obtain makes it nearly impossible to clearly observe the asymptotic limits: In one-dimensional simulations this problem occurs when $|\delta| < 0.3$, as is clearly visible in Fig. 3, and it is likely even harder to observe in natural systems. This underscores the need for the much fuller analysis of the spreading dynamics as in [8].

Although the dynamics at moderate times will be dominated by the initial growth characteristic of the marginal case, we expect [8] to be a good description of the universal dynamics at large $z = \log_2 t$ even when δ is small. The limit $z \to \infty$ while $\zeta \equiv \delta z/2d$ is fixed is particularly interesting, as the solution [8] then reduces to a scaling form

$$\frac{\delta^2 \varphi}{2d} \approx \chi\left(\frac{\delta z}{2d}\right) \qquad [10]$$

with

$$\chi(\zeta) = \exp(-\xi) + \xi - 1. \qquad [11]$$

This scaling form allows us to test by simulations our analytic results across all intermediate asymptotic regimes, by plotting data obtained for different δ in one scaling plot (Fig. 5). To make the approximation uniformly valid both in the scaling regime and at asymptotically long times outside of it, we can simply replace, for $\delta < 0$, the scaling variable by $\zeta \equiv -\log(t)\eta$ (with η defined in [5]). The scaling form [10] will then be valid up to corrections that are small compared with the ones given in all regimes: We thus use this form for the scaling fits in Fig. 5, *Inset*, plotting data obtained for different δ in one scaling plot, thereby testing our solution across all intermediate asymptotic regimes.

Heterogeneities and Dynamics on Networks. Thus far we have considered spatially uniform systems, in which the jump probability between two points depends only on their separation. However, long-distance transport processes may be very heterogeneous. An extreme example is airplane travel, which occurs on a network of links between airports with mixtures of short- and long-distance flights, plus local transportation to and from airports. A simple model is to consider each site to have a number of connections from it, with the probability of a connection between each pair of sites a distance r away being $C(r)$, independently for each pair: Note that although the network is heterogeneous, statistically the system is still homogeneous as the connection probability does not depend on position. If the rate at which jumps occur across a connection of length r is $H(r)$, then, averaged over all pairs of sites, the rate of jumps of distance r is $J(r) = C(r)H(r)$. How similar is this to a homogeneous model with the same $J(r)$, in particular if $J(r) \sim 1/r^{d+\mu}$? If there are a large number of possible connections along which the key jumps can occur to get from a source region of size $\sim \ell(T/2)$ to a funnel of similar size a distance $\ell(T)$ away, then the fact that these occur to and from only a small fraction of the sites should not matter for the large length-scale behavior. The number of such connections is $n_c \sim \ell(T/2)^{2d} C(\ell(T))$ with, making the ansatz that, as in the homogeneous case, $T\ell(T/2)^{2d} J(\ell(T)) \sim 1$ (ignoring subdominant factors) we have $n_c \sim 1/[TH(\ell(T))]$. Thus, the condition for our results to be valid asymptotically is that $H(r) \ll 1/\tau(r)$ with $\tau(r)$ the inverse of the function $\ell(t)$: i.e., in the exponential, marginal, and power-law cases, respectively, that $H(r) \ll \log(r)^{-1/\eta}$, $H(r) \ll \exp(-\sqrt{4d \log 2 \log r})$, and $H(r) \ll 1/r^{\mu - d}$.

If there are insufficient numbers of connections for the heterogeneity of the network to be effectively averaged over, the behavior changes. The extreme situation is when there is a distance-independent rate for jumps along the longest connection out of a site: i.e., $H(r) \to$ const. In this case, jumps along the path with the shortest number of steps, S, to get from the origin to a point R will reach that point in a time proportional to S; i.e., $\tau(R) \sim S(R)$. The geometrical problem of obtaining the statistics of $S(R)$ has been analyzed by Biskup (24, 46). For $\mu > d$, $S \sim R$ and long jumps do not matter, whereas for $\mu < d$, $S \sim (\log R)^{1/\eta}$ with the same exponent η as in the homogeneous case we have analyzed. The difference between this result and ours is only in the power-law-of-T prefactor of $\ell(T)$ arising from the integral over time: This does not exist in the extreme network limit. In the marginal case, $\mu = d$, $T \sim SR^\alpha$ with α dependent on the coefficient of the power-law decay of the connection probability.

If the probability of a jump along a long-distance connection decays with distance but more slowly than $1/\tau(r)$, the behavior is similar: For $\mu < d$ again the ubiquitous stretched exponential behavior occurs, whereas for $\mu > d$ there are too few connections only if $C(r) < 1/r^{2d}$ in which case the number of steps and the time are both proportional to the distance. The marginal cases we have not analyzed further.

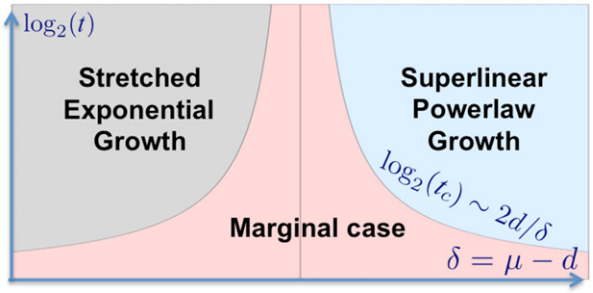

Fig. 6. Close to the marginal case, $\mu = d$, the spreading dynamics exhibit three behaviors. On asymptotically large times, either stretched exponential growth for $-d < \delta \equiv \mu - d < 0$ or superlinear power-law growth for $0 < \delta < 1$ occurs. However, the approach to the asymptotic regime is extremely slow for μ close to d. For log-times $\log_2(t) \ll 2d/|\delta|$, the behavior is controlled by the dynamics of the marginal case. Note that crossover behavior also obtains near the borderline between linear and superlinear behavior at $\mu = d + 1$; compare *SI Text*, section *SI2.D* and Fig. S4.

For natural transport processes, the probabilities of long dispersal events will depend on both the source and the destination. If the heterogeneities are weak on large length scales, our results still obtain. However, if there are sufficiently strong large-scale heterogeneities, either in a spatial continuum or in the network structure (i.e., location of nodes and links and the jump rates along these or hub-spoke structure with multiple links from a small subset of sites), then the spatial spread will be heterogeneous even on large scales: How this reflects the underlying heterogeneities of the dispersal has to be analyzed on a case-by-case basis.

Discussion

Modeling Evolutionary Spread. We have studied the impact of long-range jumps on evolutionary spreading, using the example of mutants that carry a favorable genetic variant. To this end, we analyzed a simple model in which long-range jumps lead to the continual seeding of new clusters of mutants, which themselves grow and send out more migrant mutants. The ultimate merging of these satellite clusters limits the overall growth of the mutant population, and it is a balance of seeding and merging of subclusters that controls the spreading behavior.

To classify the phenomena emerging from this model, we focused on jump distributions that exhibit a power-law tail. We found that, with power-law jumps, four generic behaviors are possible on sufficiently long times: The effective radius of the mutant population grows at constant speed, as a superlinear power law of time, as a stretched exponential, or simply exponentially depending on the exponent, $d + \mu$, of the power-law decay of the jump probability. These predictions are in contrast to deterministic approximations that predict exponential growth for power-law decaying jump kernels (27, 28, 30, 31). In dimensions more than one, the results also contradict the naive expectation from dynamics of neutral dispersal, that a finite diffusion coefficient is sufficient for conventional behavior (in this context, finite speed of spreading): Specifically, the variance in dispersal distances is finite for $d + 1 > \mu > 2$, but the spread is superlinear—indeed stretched exponential for $d > \mu > 2$. That superlinear dynamics can occur for $\mu < 2$ is not surprising as even a migrating individual undergoes a Levy flight: More surprising is that this occurs even when the dynamics of individuals are, on large scales, like a normal random walk.

The breakdowns of both deterministic and diffusive expectations are indicative of the importance of fluctuations: The dynamics are dominated by very rare—but not too rare—jumps: roughly, the most unlikely that occur at all up to that time. One of the consequences of this control by the rare jumps is the relatively minor role played by the linear growth speed v of individual clusters due to short-range migration: In the regime of power-law growth, the asymptotic growth of the mutant population becomes (to leading order) independent of v although when individual clusters grow more slowly, the asymptotic regime is reached at a later time. In the stretched exponential regime, the growth of subclusters sets the crossover time from linear to stretched exponential and thus determines the prefactor in the power law that characterizes the logarithm of the mutant population size (*Crossovers and Beyond Asymptopia*).

An important feature of the spreading dynamics is that the approach to the asymptotic behavior is very slow in the vicinity of the marginal cases, as illustrated in Fig. 6. Consider, for instance, the 2D case, where we have asymptotically stretched exponential growth for $\mu < 2$: For $\mu = 1.8$ ($\mu = 1.6$), the epidemic has to run for times $t \gg 10^6$ (10^3) to reach the asymptotic regime. By that time, the mutant population with $\ell \gg 10^{30}$ (10^7) would have certainly spread over the surface of the earth, so that the asymptotic laws alone are in fact of limited value.

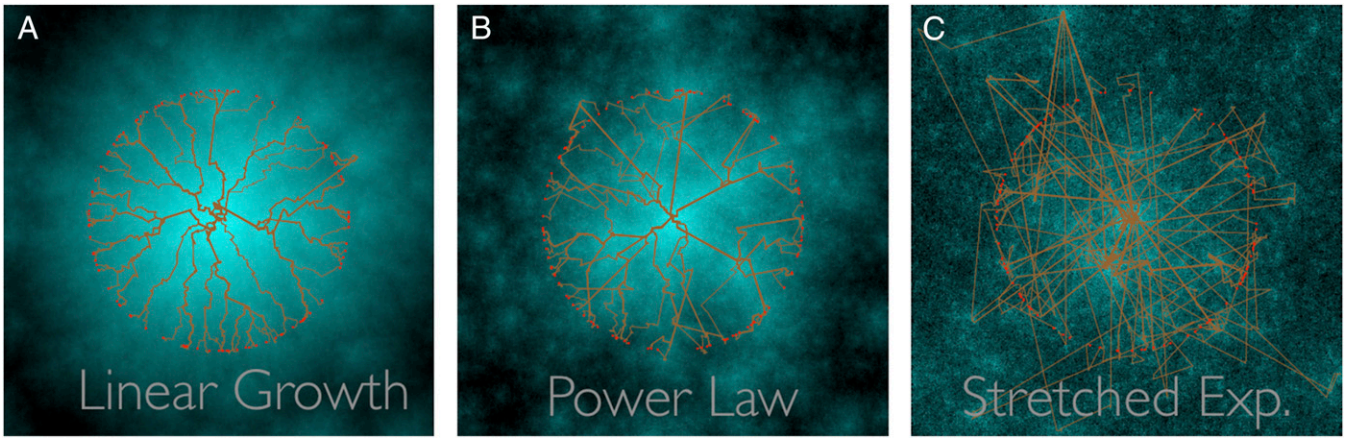

Fig. 7. (*A*–*C*) Coalescent trees, or infection trees in epidemiology, generated by power-law dispersal. *A*, *B*, and *C* each represent one simulation run with parameter $\mu = 3.5$, $\mu = 2.5$, and $\mu = 1.5$, respectively. For each run, 100 lattice sites indicated as red points equidistant from the start point at the origin were sampled. For each labeled site the path that led to its colonization is plotted. The resulting coalescence trees are characteristic of the three different regimes and reveal the long-range jumps that drove the colonization process. The background represents the colonization process in time with the color of each site indicating its colonization time (light, early; dark, late).

However, based on a simple geometrical argument illustrated in Fig. 4, we could show that the full crossover dynamics can be understood from a trade-off between frequency and potential effectiveness of long-distance jumps: Jumps of a given size are more abundant at late times (source funnel is large) but they are most effective at early times (blue funnel is large). As a result, the key jumps that dominate the filled regions at time T predictably occur near half that time. This led to a simple recurrence relation for the spread at time T in terms of its behavior at time $T/2$ (Eq. **7**). Its exact solution predicts a universal crossover function for the transient dynamics near $\mu = d$ that could be reproduced by simulations for different power-law exponents collapsed onto one scaling plot; compare Fig. 5. The good agreement of the predicted scaling function and simulation results provides strong support for the iterative scaling approximation. The rigorous bounds whose proofs are outlined in *SI Text*, section *SI3* provide further support. Understanding this crossover is essential for making sense of, and extrapolating from, simulations as asymptotic behavior is not visible until enormous system sizes even when the exponent μ is more than 0.2 from its marginal value.

Another benefit of the ability of the simple iterative scaling argument to capture well nonasymptotic behavior is that it can be used in cases in which the dispersal spectrum of jumps is not a simple power law, e.g., with a crossover from one form to another as a function of distance. And the heuristic picture that it gives rise to—an exponential hierarchy of timescales separated by roughly factors of 2—is suggestive even in more complicated situations. That such a structure should emerge without a hierarchal structure of the underlying space or dynamics is perhaps surprising.

Potential Applications and Dynamics of Epidemics. Our primary biological aim is the qualitative and semiquantitative understanding that emerges from consideration of the simple models and analyses of these, especially demonstrating how rapid spatial spread of beneficial mutations or other biological novelties can be even with very limited long-range dispersal. As the models do not depend on any detailed information about the biology or dispersal mechanisms, they can be considered as a basic null model for spreading dynamics in physical, rather than more abstract network, space.

The empirical literature suggests that fat-tailed spectra of spatial dispersal are common in the biological world (8, 13, 36–41). Because most of these are surely neither a constant power law over a wide range of scales nor spatially homogeneous, our detailed results are not directly applicable. However, as discussed above, our iterative scaling argument is more general and can be applied with more complicated distance dependence, anisotropy, or even directional migration (47). Furthermore, some of the heterogeneities of the dispersal will be averaged out for the overall spread, while affecting when mutants are likely to arrive at particular locations.

For dispersal via hitchhiking on human transport, either of pathogens or of commensals such as fruit flies with food, the apparent heterogeneities are large because of the nature of transportation networks. Nevertheless, data from tracking dollar bills and mobile phones indicate that dispersal of humans can be reasonably approximated by a power law with $\mu \approx 0.7 \pm 0.15$ (3, 4), cut off by an exponential tail for jump distances larger than 400 km (4) (as obtained from dataset $D1$ in ref. 4, tracking 10^5 mobile phone users). The exponent of the power-law part falls into the asymptotic regime of stretched-exponential growth with exponent $\eta \approx 0.57 \pm 0.1$. However, due to the exponential cutoff and the slow crossovers, the asymptotics are of little predictive value. Given the inferred kernel, we would then expect the growth dynamics to follow Eq. **8** until the key jumps fall into the exponential tail of the truncated power-law kernel, upon which linear growth (at a speed set by the exponential tail) would ensue.

Whether transport via a network with hubs at many scales fundamentally changes the dynamics of an expanding population of hitchhikers from that with more homogeneous jump processes depends on the nature of how the population expands. For spread of a human epidemic, there are several possible scenarios. If the human population is reasonably uniform spatially, and the chance that a person travels from, say, his or her home to another person's home is primarily a function of the distance between these rather than the specific locations, then whether the properties of the transportation network matter depends on features of the disease. If individuals are infectious for the whole time the outbreak lasts and if transmission is primarily at end points of journeys rather than en route—for example, HIV—then the transportation network plays no role except to provide the spatial jumps. At the other extreme is whether individuals living near hubs are more likely to travel (or even whether destinations near hubs are more likely) and, more so, whether infections are likely to occur en route, in which case the structure of the transportation network—as well as of spectra of city sizes, etc.—matters a great deal. In between these limits the network (or lack of it in places) may matter for initial local spread but at longer times the network structure may effectively average out and the dynamics be more like those of the homogeneous models. The two opposite limits and behavior in between these, together with the specific network model we analyzed with jumping probabilities depending on distance even in the presence of a connection—as is true from airports—all illustrate an important point: Geometrical properties of networks alone rarely determine their properties; quantitative aspects, such as probabilities of moving along links and what exists at the nodes, are crucial.

More complicated epidemic models can be discussed within the same framework: The model discussed thus far corresponds to an SI model, the most basic epidemic model, which consists of susceptible and infected individuals only. Many important epidemics are characterized by rather short infectious periods, so that one has to take into account the transition from infected to recovered: SIR models of the interaction between susceptible (S), infected (I) and recovered (R) individuals. This changes fundamentally the geometry of the space–time analysis in Fig. 4. Whereas the target funnel remains a full funnel, the source funnel becomes hollow: The center of the population consists mostly of fully recovered individuals, whose long-range jumps are irrelevant. The relevant source population of infected individuals is primarily near the boundaries of the funnel. This leads to a break in the time symmetry of the argument. As a result, the spreading crosses over from the behavior described above to genuine SIR behavior. The SIR dynamics are closely related to the scaling of graph distance in networks with power-law distributions of link lengths (24, 46), as recently shown by 1D and 2D simulations (48, 49). In particular, the limited time of infectiousness causes wavelike spreading at a constant speed for $\mu > d$. However, importantly, the spreading velocity is controlled by the SI dynamics we have studied until a time of order of the infectious period. Analogous crossovers from SI to SIR occur also with longer-range jumps. More generally, other complications can be discussed in our framework, and we expect new behaviors, depending on how they modify the geometry of the source to target-funnel picture.

We expect our approach to be useful also for investigating genetic consequences of range expansions and genetic hitchhiking on spatial selective sweeps with long-distance dispersal (50–53). A generalization of the analysis of the phenomenon of "allele surfing," which has been mainly analyzed assuming short-range migration so far (54–56), could be used to clarify the conditions under which long-distance dispersal increases or decreases genetic diversity—both effects have been seen in simulations (57–59). New effects arise due to the "patchiness" (60) generated by the proliferation of satellite clusters that are seeded by long-range jumps, as, e.g., observed in many plant species (61). A needed further step is the analysis of the interaction between multiple

spreading processes, for instance generated by the occurrence of multiple beneficial variants in a population, which leads to soft sweeps or clonal interference in space (62, 63). Whereas colliding clones strongly hinder each other's spread with only short-range migration, rare long-distance jumps may overcome these constraints (62), leading to irregular global spreading as we have found for a single clone.

Finally, our analyses naturally provide information on the typical structure of coalescent trees backward in time. For a given site, the path of jumps by which the site was colonized can be plotted as in Fig. 7. Doing this for many sites yields coalescent trees, which can reveal the key long-range jumps shared by many lineages. In general, such genealogical information is helpful for reconstructing the demographic history of a species. In the particular context of epidemics, combining spatiotemporal sampling of rapidly evolving pathogens with whole-genome sequencing is now making it possible to construct corresponding "infection" trees, and their analysis is used to identify major infection routes (64). For such inference purposes, it would be interesting to investigate the statistical properties of coalescence trees generated by simple models such as ours and how they depend on the dispersal properties, network structure, and other features of epidemic models.

ACKNOWLEDGMENTS. We wish to thank Rava da Silveira for useful discussions. This work was partially supported by a Simons Investigator award from the Simons Foundation (O.H.), the Deutsche Forschungsgemeinschaft via Grant HA 5163/2-1 (to O.H.), and the National Science Foundation via Grants DMS-1120699 (to D.S.F.) and PHY-1305433 (to D.S.F.).

1. Ruiz GM, et al. (2000) Global spread of microorganisms by ships. *Nature* 408(6808):49–50.
2. Suarez AV, Holway DA, Case TJ (2001) Patterns of spread in biological invasions dominated by long-distance jump dispersal: Insights from argentine ants. *Proc Natl Acad Sci USA* 98:1095–1100.
3. Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465.
4. Gonzalez M, Hidalgo C, Barabási A (2008) Understanding individual human mobility patterns. *Nature* 453:779–782.
5. Rhee I, et al. (2011) On the levy-walk nature of human mobility. *IEEE/ACM Trans Netw* 19:630–643.
6. Brockmann D, Helbing D (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342:1337–1342.
7. Clark JS (1998) Why trees migrate so fast: Confronting theory with dispersal biology and the paleorecord. *Am Nat* 152:204–224.
8. Clark JS, Silman M, Kern R, Macklin E, HilleRisLambers J (1999) Seed dispersal near and far: Patterns across temperate and tropical forests. *Ecology* 80:1475–1494.
9. Brown JK, Hovmoller MS (2002) Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science* 297:537–541.
10. McCallum H, Harvell D, Dobson A (2003) Rates of spread of marine pathogens. *Ecol Lett* 6:1062–1067.
11. D'Ovidio F, Fernández V, Hernández-García E, López C (2004) Mixing structures in the Mediterranean sea from finite-size lyapunov exponents. *Geophys Res Lett* 31(17):1–4.
12. Martin A (2003) Phytoplankton patchiness: The role of lateral stirring and mixing. *Prog Oceanogr* 57(2):125–174.
13. Levin SA, Muller-Landau HC, Nathan R, Chave J (2003) The ecology and evolution of seed dispersal: A theoretical perspective. *Annu Rev Ecol Evol Syst* 34:575–604.
14. Nathan R (2006) Long-distance dispersal of plants. *Science* 313:786–788.
15. Perlekar P, Benzi R, Nelson D, Toschi F (2010) Population dynamics at high Reynolds number. *Phys Rev Lett* 105:144501.
16. Gillespie RG, et al. (2012) Long-distance dispersal: A framework for hypothesis testing. *Trends Ecol Evol* 27(1):47–56.
17. Fisher RA (1937) The wave of advance of advantageous genes. *Ann Eugen* 7:355–369.
18. Kolmogorov A, Petrovsky I, Piskunov N (1937) Study of the diffusion equation with growth of the quantity of matter and its application to a biological problem. Bull Univ Moscow Ser Int *Sec A* 1:1–25.
19. Mollison D (1972) The rate of spatial propagation of simple epidemics. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, eds Le Cam LM, Neyman J, Scott EL (Univ of California Press, Berkeley, CA), Vol 3, pp 579–614.
20. Clark JS, Lewis M, McLachlan JS, HilleRisLambers J (2003) Estimating population spread: What can we forecast and how well? *Ecology* 84:1979–1988.
21. Filipe J, Maule M (2004) Effects of dispersal mechanisms on spatio-temporal development of epidemics. *J Theor Biol* 226(2):125–141.
22. Cannas SA, Marco DE, Montemurro MA (2006) Long range dispersal and spatial pattern formation in biological invasions. *Math Biosci* 203(2):155–170.
23. Marco DE, Montemurro MA, Cannas SA (2011) Comparing short and long-distance dispersal: Modelling and field case studies. *Ecography* 34:671–682.
24. Biskup M (2004) On the scaling of the chemical distance in long-range percolation models. *Ann Probab* 32(4):2938–2977.
25. Aldous DJ (2010) When knowing early matters: Gossip, percolation and Nash equilibria. *arXiv:*1005.4846.
26. Chatterjee S, Dey PS (2013) Multiple phase transitions in long-range first-passage percolation on square lattices. *arXiv:*1309.5757.
27. Mollison D (1977) Spatial contact models for ecological and epidemic spread. *J R Stat Soc B* 39(3):283–326.
28. Kot M, Lewis M, Van Den Driessche P (1996) Dispersal data and the spread of invading organisms. *Ecology* 77:2027–2042.
29. Neubert MG, Caswell H (2000) Demography and dispersal: Calculation and sensitivity analysis of invasion speed for structured populations. *Ecology* 81:1613–1628.
30. Mancinelli R, Vergni D, Vulpiani A (2002) Superfast front propagation in reactive systems with non-Gaussian diffusion. *Europhys Lett* 60:532.
31. del Castillo-Negrete D, Carreras B, Lynch V (2003) Front dynamics in reaction-diffusion systems with Levy flights: A fractional diffusion approach. *Phys Rev Lett* 91:18302.
32. Medlock J, Kot M (2003) Spreading disease: Integro-differential equations old and new. *Math Biosci* 184:201–222.
33. Brockmann D, Hufnagel L (2007) Front propagation in reaction-superdiffusion dynamics: Taming Levy flights with fluctuations. *Phys Rev Lett* 98:178301.
34. Coville J, Dupaigne L (2007) On a non-local equation arising in population dynamics. *Proc R Soc Edinb Sec A Math* 137:727–755.
35. del Castillo-Negrete D (2009) Truncation effects in superdiffusive front propagation with Lévy flights. *Phys Rev E Stat Nonlin Soft Matter Phys* 79:031120.
36. Levandowsky M, Schuster FL, White BS (1997) Random movements of soil amebas. *Acta Protozool* 36:237–248.
37. Klafter J, White BS, Levandowsky M (1990) Microzooplankton feeding behavior and the Levy walk. *Biological Motion*, Lecture Notes in Biomathematics, eds Alt W, Hoffmann G (Springer, Berlin), Vol 89, pp 281–293.
38. Atkinson R, Rhodes C, Macdonald D, Anderson R (2002) Scale-free dynamics in the movement patterns of jackals. *Oikos* 98(1):134–140.
39. Fritz H, Said S, Weimerskirch H (2003) Scale–dependent hierarchical adjustments of movement patterns in a long–range foraging seabird. *Proc R Soc Lond B Biol Sci* 270:1143–1148.
40. Ramos-Fernandez G, et al. (2004) Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). *Behav Ecol Sociobiol* 55:223–230.
41. Dai X, Shannon G, Slotow R, Page B, Duffy K (2007) Short-duration daytime movements of a cow herd of African elephants. *J Mammal* 88(1):151–157.
42. Garnier J (2011) Accelerating solutions in integro-differential equations. *SIAM J Math Anal* 43:1955–1974.
43. Cabré X, Roquejoffre J-M (2009) Propagation de fronts dans les équations de fisher–kpp avec diffusion fractionnaire [Front propagation in Fisher–KPP equations with fractional diffusion]. *C R Acad Sci Paris* 347:1361–1366.
44. Marvel SA, Martin T, Doering CR, Lusseau D, Newman MEJ (2013) The small-world effect is a modern phenomenon. arXiv:1310.2636.
45. Van Saarloos W (2003) Front propagation into unstable states. *Phys Rep* 386:29–222.
46. Biskup M (2004) Graph diameter in long-range percolation. *arXiv*:0406379.
47. Armsworth PR, Roughgarden JE (2005) The impact of directed versus random movement on population dynamics and biodiversity patterns. *Am Nat* 165:449–465.
48. Grassberger P (2013) Sir epidemics with long-range infection in one dimension. *J Stat Mech* 2013:P04004.
49. Grassberger P (2013) Two-dimensional SIR epidemics with long range infection. *J Stat Phys* 153:289–311.
50. Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annu Rev Ecol Evol Syst* 40:481–501.
51. Novembre J, Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* 10:745–755.
52. Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genet Res* 71(02):155–160.
53. Barton N, Etheridge A, Kelleher J, Veber A (2013) Genetic hitchhiking in spatially extended populations. *Theor Popul Biol* 87:75–89.
54. Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 101:975–979.
55. Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 23:482–490.
56. Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theor Popul Biol* 73(1):158–170.
57. Nichols RA, Hewitt GM (1994) The genetic consequences of long distance dispersal during colonization. *Heredity* 72:312–317.
58. Bialozyt R, Ziegenhagen B, Petit R (2006) Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *J Evol Biol* 19(1):12–20.
59. Berthouly-Salazar C, et al. (2013) Long-distance dispersal maximizes evolutionary potential during rapid geographic range expansion. *Mol Ecol* 22:5793–5804.
60. Lewis M, Pacala S (2000) Modeling and analysis of stochastic invasion processes. *J Math Biol* 41:387–429.
61. Moody ME, Mack RN (1988) Controlling the spread of plant invasions: The importance of nascent foci. *J Appl Ecol* 25:1009–1021.
62. Ralph P, Coop G (2010) Parallel adaptation: One or many waves of advance of an advantageous allele? *Genetics* 186:647–668.
63. Martens E, Hallatschek O (2011) Interfering waves of adaptation promote spatial mixing. *Genetics* 189:1045–1060.
64. Matthews L, Woolhouse M (2005) New approaches to quantifying the spread of infection. *Nat Rev Microbiol* 3:529–536.

Supporting Information

Hallatschek and Fisher 10.1073/pnas.1404663111

SI Text

SI1. Simulation Details

A. Simulation Algorithm. For one-dimensional simulations, the state of the population is described by a linear array of N sites with periodic boundary conditions. N is chosen large enough so that end effects can be ignored (typically between 10^7 and 10^8 sites). Each site has the identity of either mutant or wild type. Initially, the whole population is wild type except for the central site, which is occupied by mutants.

In each computational time step, a source site A and target site B are chosen randomly such that their distance r is sampled from a probability density function with a tail $\mu r^{-(1+\mu)}$ at large r (see below). If A is mutant and B a wild type, then B turns into mutants and seeds a new mutant cluster. We use the convention that N time steps—i.e., an average of one jump attempt per site—comprise 1 unit of time or effective "generation." The rate of long-range jumps should be thought of as representing the product of the probability to establish a new cluster per jump and the jump rate per generation per site.

B. Jump Size Distribution. In our simulations, the distance X of a long-range jump was generated as follows. First, draw a random number Y within $(0,1)$ and calculate the variable

$$X = [Y(L^{-\mu} - C^{-\mu}) + C^{-\mu}]^{-1/\mu}, \qquad \text{[S1]}$$

where C is a cutoff (see below) and L is the system size. This generates a continuous probability density function

$$\Pr(X = x) = x^{-(\mu+1)} \frac{\mu (CL)^{\mu}}{L^{\mu} - C^{\mu}} \qquad \text{[S2]}$$

with x values in (C, L). The actual jump distance is obtained from X by rounding down to the next integer. [Note that, because the distribution has a tail $\mu r^{-(1+\mu)}$, we have to choose $\epsilon = \mu$ in Eq. **2** of the main text.]

For the one-dimensional data in all graphs of this paper, we used $C = 1$ and system sizes ranging from $L = 10^9$ for $\mu = 0.6$ to $L = 10^8$ for $\mu = 1.4$. For such large systems, the tail of the distribution is well approximated by $p(x) \sim \mu x^{-(\mu+1)}$, as stated in the main text. We also tested variations in the cutoff C. Using $C = 10$ or $C = 100$ affected only the short-time dynamics and had very little influence on the intermediate asymptotic or long-distance behavior of the system.

For our 2D simulations, we draw jump sizes from the same distribution as the one described above and round down to the next possible site. We set the lower cutoff to $C = 1.5 > \sqrt{2}$ to make sure that jumps reach out of the source lattice point. After the jump size is drawn, the jump direction is chosen at random.

SI2. Iterative Scaling Approximation: Details and Extensions

In the main text, *Crossovers and Beyond Asymptopia*, we used a saddle-point approximation (Laplace method) to obtain a recurrence relation, Eq. **7**, for the core radius of the mutant population as a function of time. For this purpose, we assumed that length and time were measured in units of elementary crossover scales at which the growth law changes from linear to superlinear in time.

Here, we first discuss how the crossover scales can be determined, solve the recurrence relation explicitly, and work out corrections to the saddle-point approximation. Finally, we present an improved version of our geometrical argument of the source and target funnels (Fig. 4), which also allows us to estimate the probability of occupancy outside the core region.

A. Crossover Scales. The crossover scales $\ell_\times$ and $t_\times$ from linear to superlinear growth can be obtained explicitly when we assume that the total rate of long jumps is small compared with that of the short-range jumps: i.e., when ε, the coefficient of $G(r) \approx \epsilon / r^{d+\mu}$, is small. Short jumps result in diffusive motion and linear growth $\ell(t) \approx v_0 t$ with v_0 determined by the details of the selective and diffusive dynamics. Long jumps start to become important after enough time has elapsed that there have been at least some jumps of lengths of order $\ell(t)$: i.e., when $\epsilon t \ell(t)^d / \ell(t)^{\mu} \gg 1$; this occurs after a crossover time $t_\times \sim [v_0^{\mu-d} / \epsilon]^{1/(d+1-\mu)}$ at which point $\ell \sim \ell_\times \sim [v_0/\epsilon]^{1/(d+1-\mu)}$. At longer times and distances, we can measure lengths and times in units of these crossover scales, defining $\lambda \equiv \ell/\ell_\times$ and time $\theta = t/t_\times$, and expect that the behavior in these units will not depend on the underlying parameters. Note that this separation in short-time linear growth and long-time regimes can also be done for more general $G(r)$ although then the behavior will depend on the whole function—the crossover on distances of order $\ell_\times$ and the superlinear behavior on the longer distance form.

B. Solving the Recurrence Relation for the Core Radius. The recurrence relation Eq. **7** for the rescaled core radius $\lambda(t) \equiv \ell(t)/\ell_\times$ of the mutant population in terms of the rescaled time $\theta \equiv t/t_\times$,

$$\lambda^{2d+\delta}(\theta) \sim \theta \lambda (\theta/2)^{2d}, \qquad \text{[S3]}$$

is valid in the vicinity of the phase boundary $\delta \equiv \mu - d = 0$. We now show how the solution, quoted in [**8**], and the associated prefactors of the asymptotic growth laws can be obtained.

Defining $\varphi \equiv \log_2(\lambda)$ and $z \equiv \log_2(\theta)$ and taking the binary logarithm, $\log_2$, of Eq. **S3**, we obtain the linear recurrence relation

$$(2d + \delta)\varphi(z) \approx 2d\ \varphi(z-1) + z. \qquad \text{[S4]}$$

Note that in our conventions we use $\sim$ (as in Eq. **S3**) for asymptotically goes as, with unknown coefficient [i.e., loosely similar to $O(\ldots)$ notation], and $\approx$ (as in Eq. **S4**) if we know the coefficient; i.e., the ratio goes to unity.

Now, it is straightforward to see that

$$Q(z) \equiv \frac{z}{\delta} - \frac{2d}{\delta^2} \qquad \text{[S5]}$$

is a special solution of [**S4**]. Substituting $\varphi(z) = \tilde{\varphi}(z) + Q(z)$, we are left with the homogeneous problem

$$(2d + \delta)\tilde{\varphi}(z) \approx 2d\ \tilde{\varphi}(z-1), \qquad \text{[S6]}$$

which is easily solved by

$$\tilde{\varphi}(z) \approx \left(1 + \frac{\delta}{2d}\right)^{-z} \tilde{\varphi}(0). \qquad \text{[S7]}$$

Reinserting $\varphi(z)$ and imposing the initial condition $\varphi(0) = 0$ finally yields Eq. **8** of the main text,

$$\frac{\delta^2}{2d}\varphi(z)\approx\frac{\delta z}{2d}+\left(1+\frac{\delta}{2d}\right)^{-z}-1. \qquad \textbf{[S8]}$$

In the limit $z\to\infty$, the second and first terms dominate for $\delta>0$ and $\delta<0$, respectively. The resulting asymptotics

$$\log[\lambda(t)]\approx\begin{cases} B_\mu t^\eta, & \eta=\log\dfrac{2d/(d+\mu)}{\log 2}, \quad \delta>0 \\ \log\left[A_\mu t^\beta\right], & \beta=\dfrac{1}{\mu-d}, \quad \delta<0 \end{cases} \qquad \textbf{[S9]}$$

reveal the prefactors $B_\mu=2d\log(2)\delta^{-2}$ and $\log A_\mu=-2d\log(2)\delta^{-2}$ quoted in the main text. Note that we use the variable β throughout the *SI Text* as a shorthand for the power-law exponent $\beta=1/(\mu-d)$.

C. Subdominant Corrections from Time Integrals. In analyzing the iterative scaling approximation in the main text, we have effectively replaced the integrand in the time integral of Eq. **4** by its value at the half-time peak that dominates the probabilities of occupation at time T. (This procedure resulted in the simple recursion relation **[8]**, which we have explicitly solved in the previous section *SI2.B*.) In doing so, we have ignored effects associated with the parameter-dependent width of the peak around $T/2$ that contributes substantially to the integral. This is valid for obtaining the leading behaviors of $\log\ell(t)$ in the large time limit, but there are corrections to these that can be larger than those that arise from the short-time small-length-scale crossovers that we discussed in the main text and we discuss them further below. For μ not much smaller than $d+1$, when the growth of ℓ is a modest power of time, the factor from the range of the time integral is only of order unity and hence no worse than other factors—including from the stochasticity—that we have neglected. However, when $\ell(t)$ grows very rapidly, the range of $(1/2)T-t$ that dominates is much smaller than T and the corrections are larger.

With rapid growth of $\ell(t)$, a saddle-point approximation to the time integral is valid: $\int_0^t dt\ell(t)\ell(T-t)\approx Tc_{T/2}\ell^2(T/2)$ with the prefactor given by

$$c_t\approx\sqrt{\frac{2\pi}{-2t^2\partial_t^2\log\ell(t)}}, \qquad \textbf{[S10]}$$

from which, with the second derivative absorbing the t^2 factor, the derivative part can be rewritten as $t^2\partial_t^2\log\ell(t)=-\partial_{\log t}\log\ell+\partial_{\log t}^2\log\ell$. With the asymptotic growth laws we have derived, this gives $c\sim\sqrt{\mu-d}$ for $\mu-d$ small and positive, $c_t\sim1/\sqrt{\log t}$ for $\mu=d$, and $c_t\sim t^{-\eta/2}$ for $\mu<d$. Integrating up the effects of this over the scales yields the following corrections in the various regimes: For $\mu>d$, the coefficient, A_μ, of t^β is changed by a multiplicative factor that is much less singular for $\mu\searrow d$ than that already obtained. For $\mu=d$,

$$\log[\ell(t)]\approx C\left[\log^2 t-\log t\log\log t+\mathcal{O}(\log t)\right] \qquad \textbf{[S11]}$$

with $C=1/4d\log 2$, the second term being new and the smaller correction term including the effects of the small time crossover. For $0<\mu<d$,

$$\log[\ell(t)]\approx B_\mu t^\eta-\frac{1-\eta/2}{d-\mu}\log t \qquad \textbf{[S12]}$$

with the second term for $\mu\nearrow d$ just what occurs in the crossover regime analyzed in *Crossovers and Beyond Asymptopia* (main text), where we showed that the coefficient B_μ diverges proportional to $1/(d-\mu)^2$.

D. Prefactors in Power-Law Regime. In the main text, we mainly focused on regimes in which the mutant growth is very much faster than linear; i.e., $\mu\lesssim d+1/2$. This allowed us to approximate the integrals in the iterative scaling approximation of Eq. **1** (main text) by the use of Laplace's method. This saddle-point approximation yields the correct scaling for all exponents $\mu<d$, but (as we see below) incorrect prefactors in regimes where the actual growth is close to linear, i.e., in the power-law growth regime with $d+1>\mu\gtrsim d$.

To obtain a better estimate of the prefactors in this power-law regime, it is helpful to directly solve for the asymptotics of the iterative scaling argument in Eq. **1** (main text). Here, we demonstrate how this can be done for $d=1$: Assume that most of the weight in the integral comes from regions where the jump kernel is well approximated by its power-law tail described in Eq. **2** (main text). Given $\mu<2$ (for $d=1$), this always holds at sufficiently long times. Then, we have

$$\epsilon\int_0^t dt'H(t')\sim1, \qquad \textbf{[S13]}$$

where

$$\begin{aligned}\mu(\mu-1)H(t')\equiv&\left(\ell(t)-\ell(t')-\ell(t-t')\right)^{1-\mu}\\&-\left(\ell(t)-\ell(t')+\ell(t-t')\right)^{1-\mu}\\&+\left(\ell(t)+\ell(t')+\ell(t-t')\right)^{1-\mu}\\&-\left(\ell(t)+\ell(t')-\ell(t-t')\right)^{1-\mu}.\end{aligned} \qquad \textbf{[S14]}$$

For $1<\mu<2$, Eq. **S13** exhibits an asymptotic power-law solution

$$\ell(t)=A_\mu(\epsilon t)^{1/(\mu-1)}. \qquad \textbf{[S15]}$$

By inserting this ansatz into Eqs. **S13** and **S14**, we obtain the following result for the numerical prefactor,

$$A_\mu^{\mu-1}\approx\int_0^1 dz\tilde{R}(z), \qquad \textbf{[S16]}$$

with $\tilde{R}(z)$ being equal to $H(t)$ in Eq. **S14** with $\ell(t)$ replaced by $z^{1/(\mu-1)}$.

The resulting prefactor is plotted as a function of $\mu-1$ in Fig. S3. Note that A_μ strongly depends on the exponent μ. It sharply drops for μ approaching 1, where it follows the asymptotics $2^{-2(\mu-1)^{-2}}$. On the other hand, as μ approaches the other marginal case at $\mu=2$, the prefactor diverges as $A_\mu\sim(2-\mu)^{-1}$, indicating the importance of intermediate asymptotic regimes, as discussed in *Crossovers and Beyond Asymptopia* (main text).

Although we have focused on the marginal case near $\mu=d$ in this article, it is clear that another case of marginality controls the crossovers near $\mu=d+1$. Simulation results reported in Fig. S4 indicate that $\ell(t)/t\sim\log(t)$ for $\mu=2$ in one dimension. This is consistent with our funnel argument: With nearly constant speed, the gap between the funnels remains roughly constant for a time of order t. To ensure that the source emits about one jump to the target funnel, we must have that, per unit of time, the probability of a seed jumping over the gap is of order $1/t$. Thus, the gap size ΔE should be such that $\Delta E^{-\mu+1}\sim1/t$. For $\mu=2$, we thus have $\Delta E\sim t$; i.e., the key jumps span distances of order t. This is ensured when $(\ell(t)-2\ell(t/2))/t\sim$ const., i.e., if $\ell(t)/t\sim\log t$. Note that a rigorous upper bound of this form follows from the arguments

presented in *SI Text*, section *SI3.A.4* for the regime $\mu > d$. The jumps of order $O(t)$ that drive the logarithmic increase in spreading velocity might be the "leaps forward" (1) recognized by Mollison in one of the earliest studies on spreading with long-range jumps.

E. Occupancy Profiles and Relevance of Secondary Seeds. In the main text, we introduced the notion of a nearly occupied core of the population [of size $\ell(t)$] as the source of most of the relevant seeds in the target funnel. However, it is clear that outside of this core there is a region of partial occupancy. This region is potentially broad, in particular for $\mu \to 0$, and may therefore lead to a significant fraction of relevant seeds. An improved theory should account for those secondary seeds and should also be able to determine the profile of mean occupancy or, equivalently, the probability that a site is occupied. Whereas we give rigorous bounds on how the total population grows in *SI Text*, section *SI3*, we first use an improved version of our funnel argument to describe the occupancy profiles.

We focus on the probability, $q(r,t)$, that at time t after a mutant establishes, it will have taken over the population a distance r away. We expect that $q(r,t)$ will be close to unity out to some core radius $\ell(t)$ and then decrease for larger r, with the average total mutant population proportional to $\ell(t)^d$. With only short-range dispersal, $\ell(t) \approx vt$ and the core is clearly delineated but when long jumps are important, the crossover from mostly occupied core to sparsely occupied halo will not be sharp. The more important quantity is the average of the total area (in two dimensions or linear extent or volume in one or three dimensions) occupied by the mutant population; we denote this $M(t) = \int d^d r q(r,t)$.

To find out when long jumps could be important, we first ask whether there are likely to be any jumps longer than $\ell(t)$ that occur up to time t. The average number of such long jumps is of order $t\ell(t)^d \int_{\ell(t)}^{\infty} r^{d-1} dr G(r)$. If $G(r)$ decreases more rapidly than $1/r^{2d+1}$, this is much less than $t/\ell(t)$ for large t. As $\ell(t)$ increases at least linearly in time, the probability that there have been any jumps longer than $\ell(t)$ is very small. The guess that ℓ indeed grows as vt, and consideration of jumps that could advance the front fast enough to contribute substantially to v, leads, similarly, to the conclusion that there is a maximum t-independent jump length beyond which the effects of jumps are negligible; indeed, their effect decreases more rapidly than $G(r)$. This reinforces the conclusion that there is only linear growth with $\mu > d+1$: a very strong breakdown of the deterministic approximation that yielded exponential growth for any power law.

When G is longer range, in particular if $G(r) \sim 1/r^{d+\mu}$ with $\mu < d+1$, many jumps longer than $\ell(t)$ will have occurred by time t. We now study the effects of such long jumps on the density profile. To do so, we investigate the behavior of $q(R,T)$ for large R and T in terms of the $\{q(r,t)\}$ at shorter times and—primarily—corresponding distances $r \sim \ell(t)$ that can be much less than R. Mutants can get to a chosen point, $\mathbf{R}$, by one making a long jump at time t from a starting point $\mathbf{x}$ to an end point, $\mathbf{y}$, and subsequently spreading from there to $\mathbf{R}$ during the remaining time interval of duration $T-t$. The rate (per volume elements) of this occurring is $q(x,t)G(|\mathbf{x}-\mathbf{y}|)q(|\mathbf{R}-\mathbf{y}|, T-t)$. In the approximation that these are independent, the probability that this does not occur at any $t < T$ from any $\mathbf{x}$ to any $\mathbf{y}$ is simply Poisson so that

$$q(R,T) \approx 1 - e^{-Q(R,T)} \qquad \textbf{[S17]}$$

with

$$Q(R,T) \approx \int_0^T dt \int_{z>\ell(T/2)} d^d z \int d^d x \, q(x,t) G(z) q(|\mathbf{R}-\mathbf{x}-\mathbf{z}|, T-t). \qquad \textbf{[S18]}$$

Here, we substituted the final point $\mathbf{y} = \mathbf{x} + \mathbf{z}$ by the sum of the jump start site and a jump vector $\mathbf{z}$, over which we integrate. Note that a lower cutoff in the z integral is necessary to exclude the many very short jumps that lead to strongly correlated establishments. The cutoff is also necessary to not count mutants that result from growth in the target area, rather than seeding from the source funnel. Our main assumption here is that if a single seed is sufficiently far from other seeds or occupied regions, then the growth from the seed is independent of the rest of the system as long as collisions are unlikely.

When the jump integral is strongly peaked at $z = R$, as is the case in or close to the stretched exponential regime, the final results will be independent of this cutoff to leading order. Then, we can approximate $Q(R,t)$ as

$$Q(R,t) \approx G(R) \int_0^t dt' M(t') M(t-t'), \qquad \textbf{[S19]}$$

where $M(t)$ is the expected total size of a population at a time t,

$$M(t) = \int d^d x q(x,t). \qquad \textbf{[S20]}$$

For a power-law kernel $G(R) = G(1) R^{-(1+\mu)}$, we make the ansatz that Eqs. **S19**, **S20**, and **S17** can be approximately solved by a scaling form

$$q(R,t) = \Xi\left(\frac{R}{\lambda(t)}\right) \qquad \textbf{[S21]}$$

with

$$\Xi(\xi) = 1 - \exp\left(-\xi^{-(d+\mu)}\right), \qquad \textbf{[S22]}$$

which leads to the condition

$$Q(\xi\lambda(t), t) = \xi^{-(d+\mu)} \kappa^2 G(1) \lambda^{-(d+\mu)} \int_0^t dt' \lambda^d(t') \lambda^d(t-t'), \qquad \textbf{[S23]}$$

where $M(t) = \kappa \lambda(t)^d$ and κ_μ is given by

$$\kappa = \int d\xi^d \Xi(\xi). \qquad \textbf{[S24]}$$

Thus, the above scaling form is a valid solution if the characteristic scale $\lambda(t)$ satisfies

$$\kappa^2 G(1) \lambda^{-(1+\mu)} \int_0^t dt' \lambda(t') \lambda(t-t') = 1. \qquad \textbf{[S25]}$$

The resulting condition is similar to our condition in the main text but differs by the numerical factor $G(1)\kappa_\mu^2$. In one dimension,

$$\kappa_\mu = 2\Gamma\left(\frac{\mu}{1+\mu}\right). \qquad \textbf{[S26]}$$

The divergence $\kappa_\mu^2 \sim \mu^{-2}$ as $\mu \to 0$ indicates the importance of seeds from the tail regions for small μ.

Note that at long distances, $R \gg \ell(T)$, the lengths of the jumps that dominate are close to R so that our approximation for $Q(R,T)$ should be correct even if it is a poor approximation for $R \approx \ell(T)$. Thus, the decrease in q at large distances is simply proportional to $G(R)$; more specifically,

$$q(R,T) \approx \frac{G(R)}{G(\ell(T))} \qquad \text{[S27]}$$

(as predicted by the scaling form) so that it is of order unity at $R \approx \ell(T)$. Note that this implies that, because $G(r)$ is integrable, $\int d^d r q(r,t)$ is indeed dominated by $r \sim \ell(t)$ as we have assumed. This long-distance form for the density profile is also found in the analyses of upper and lower bounds in the next sections: Thus it can be readily proved along the same lines.

SI3. Rigorous Bounds and Outlines of Routes to Proofs

As most of our results are based on approximate analyses and heuristic arguments, it is useful to supplement these by some rigorous results. We focus on the one-dimensional case: Extensions to higher dimensions can be done similarly, although with a few complications that will require some care. We sketch here the arguments that can lead to proofs without all of the details filled in.

As via the heuristic arguments, we want to obtain the behavior at longer times in terms of the behavior at shorter times, in particular times around half as long.

We want to prove that there exist time-dependent length scales, $\ell_<(t)$ and $\ell_>(t)$, and functions, $F_<(r,t)$ and $F_>(r,t)$, such that the probability, $q(r,t)$, that a site at r from the origin is occupied at time t, is bounded above and below by

$$F_<(r,t) < q(r,t) < F_>(r,t) \quad \text{with} \quad M_< \equiv \int dr F_<(r,t) \propto \ell_<(t)$$
$$\text{and} \quad M_> \equiv \int dr F_>(r,t) \propto \ell_>(t) \qquad \text{[S28]}$$

for all times. Then $\ell_<$ and $\ell_>$ are lower and upper bounds for $\ell(t)$—with some appropriately chosen definitions of $\ell(t)$ that differ somewhat, although not significantly, for the upper and lower bounds. Although we want the upper and lower bounds on $\ell(t)$ to be as close as possible to each other, in practice, we have obtained bounds that are good on a logarithmic scale: i.e., for $\log \ell(t)$, rather than on a linear scale. Similarly, we want to have the bounds be close to the actual expected form of q, with $F_<$ very close to unity for $r \ll \ell_<$ and proportional to $[\ell_</r]^{\mu+1}$ for $r \gg \ell_<$ and similar for the upper bounds.

It is often more convenient to consider the typical time to occupation as a function of the distance, $\tau(r)$, and derive upper and lower bounds for this, $\tau_>(r)$ and $\tau_<(r)$, respectively, such that

$$\ell_<(t = \tau_>(r)) = r \quad \text{and} \quad \ell_>(t = \tau_<(r)) = r \qquad \text{[S29]}$$

with

$$\tau_>(r) > \tau(r) > \tau_<(r). \qquad \text{[S30]}$$

Because of the faster than linear growth, bounds on $\tau(r)$ are generally much closer than those on $\ell(t)$.

As we want to justify the use of the heuristic iterative scaling arguments more generally, it is especially useful to obtain iterative bounds directly of the form used in those heuristic arguments: $\ell(T)$ in terms of $\{\ell(t)\}$ for t in a range near $T/2$. As the heuristic arguments do, in any case, give $\ell(T)$ only up to a multiplicative coefficient of order unity, we will generally ignore such order-unity coefficients in length scales except for coefficients that diverge or vanish exponentially rapidly as $\mu \to d$, in particular in the intermediate-range regime the coefficient, A_μ in $\ell(t) \sim A_\mu t^{1/(\mu-d)}$, which vanishes as $\log(A_\mu) \approx -\log 4/(\mu-d)^2$ as $\mu \searrow d$.

A. Upper Bounds.

1. Simple power-law bound. The simplest bound to obtain is an upper bound for $\ell(t)$ in the short- and intermediate-range regimes: i.e., in one dimension, $\mu > 1$. Define $E(t)$ to be the rightmost edge of the occupied region at time t; i.e., $c(x,t) = 0$ for $x > E(t)$. The probability of a jump that fills a position $y > E(t)$ in $(t, t+dt)$ is less than $\int_{-\infty}^{E(t)} dx G(y-x) \sim 1/(y-E(t))^\mu$. For $\mu > 1$, the lower extent of the integral can be taken to $-\infty$ as the jumps arise, predominantly, from points that are not too far from the edge. [In contrast, for $\mu < 1$ jumps from the whole occupied region are important and this bound would yield a total jump probability to long distances that diverged when integrated over y, and we would have to instead use a lower extent of the integral of $-E(t)$ for the left edge.]

The advancement of the edge is bounded by a translationally and temporally invariant process of jumps of the position of the edge by distances, ΔE, whose distribution has a power-law tail. For $\mu > 2$, the mean $\langle \Delta E \rangle < \infty$, implying that the edge, and hence $\ell(t)$, cannot advance faster than linearly in time. However, for the intermediate regime, $\langle \Delta E \rangle > \infty$ so that $E(t)$ could advance as fast as a one-sided Levy flight with $E(t)$ dominated by the largest advance. As this process would yield $E \sim t^{1/(\mu-1)}$, this implies that $\ell(t)$ is bounded above by the same form as the heuristic result.

Although the simple bound captures some relevant features, in particular the dominance of jumps of length of order r to fill up a point at distance r, it is otherwise rather unsatisfactory. First, the coefficient does not vanish rapidly as $\mu \searrow 1$. And second, it suggests that the probability that an anomalously distant point, $r \gg \ell(t)$, is occupied, is, in this crude approximation of full occupancy out to the edge, simply the probability that $E(t) > r$, which falls off only as $1/r^{\mu-1}$—much more slowly than the actual $q(r,t) \sim r^{-1-\mu}$.

Nevertheless, for proving better upper bounds, the Levy-flight approximation for the dynamics of the edge is quite useful.

2. Upper bounds from source–jump–target picture. As discussed earlier, we want to make the heuristic argument of a single long jump from a source region to a target funnel region include also—or provide solid reasons to ignore—the effects of jumps from the partially filled region outside the core of the source. Very loosely, we want to write the probability that a point, R, is not occupied at time T, as

$$1 - q(R,T) \approx \exp\left[-\int_0^T dt \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy q(x,t) G(|y-x|) q(R-y, T-t) \right] \qquad \text{[S31]}$$

with x in the source region and y in the funnel of R. However, for any positive μ, the spatial integral is dominated by $y - x$ small, so that this does not properly represent the process: There is a drastic overcounting of short jumps.

We can do much better by trying to separate the long jumps from the short ones and the source region from the funnel (in the crude approximation these overlap). To do this, we choose, for the R and T of interest, a spatiotemporal source region, S, around the origin that has a boundary at distance $B_S(t)$ that loosely reflects the growing source: $dB_S/dt \geq 0$. We then separate the process of the set of jumps that lead to R into three parts: first, jumps solely inside S, which lead to a spatiotemporal configuration of occupied sites, $\{c_S(x,t)\}$; second, bridging jumps from these out of S, say at time t from x in S to a point y in the rest of space–time, $\overline{S}$; and third, all of the subsequent dynamics from such seeds in $\overline{S}$, including inside and outside S and between these. This overcounts the possible spatiotemporal routes to R, T—especially as returns to inside S from outside are included—and thus provides an upper bound for $q(R,T)$. The probability, p_a, that a single seed, a, to y_a

at t_a leads to R being filled by T is $q(R-y_a, T-t_a)$. However, the probability that a second seed, b, leads to R filled by T is not independent as the fate of these seeds involves overlapping sets of jumps: Indeed, they are positively correlated so that

$$\begin{aligned}\mathcal{P}[c(R,T)=0|\text{seeds } a \text{ and } b] &\geq \mathcal{P}[c(R,T)=0|\text{seed } a] \times \mathcal{P}[c(R,T)=0|\text{seed } b].\end{aligned} \quad \text{[S32]}$$

Because for a given occupancy profile $\{c_S(r,t)\}$, the probability density of a seed at y,t is $dydt\int_{|x|<B_S(t)} c_S(x,t)G(y-x)$, and using the generalization of the above bound to many seeds, we have

$$q(R,T) \leq 1-\exp\left[-\int_0^T dt \int_{|x|<B_S(t)} dx\, q_S(x,t) \int_{|y|>B_S(t)} dy\, G(y-x)q(R-y,T-t))\right], \quad \text{[S33]}$$

where $q_S(x,t) \equiv \langle c_S(x,t)\rangle$ and we have used $\langle \exp(X)\rangle \geq \exp(\langle X\rangle)$ for any random variable.

To derive a useful upper bound on $q(R,T)$ we need to choose appropriately the boundary, $B_S(t)$, of the source region and put a sufficiently stringent upper bound on $q_S(x,t)$.

3. Long-range case. For the long-range case, $\mu<1$, the integrals over $x<B_S$ and $y>B_S$ of $G(y-x)$ are dominated by long distances. Thus, the short jumps from inside to outside S do not contribute significantly. We can then simply replace q_S by the larger q to obtain a slightly weaker bound that is of exactly the form of the naive estimate except for the strict delineation of the source region, which prevents the most problematic overcounting of the effects of short jumps. A particularly simple choice is $B_S=(1/2)R$ independent of t.

We now proceed by induction. Take the bound on the scaling function to have the form $F_>(r,t)=1$ for $r<\ell_>(t)$ whereas $F_>(t)=[\ell_>/r]^{\mu+1}$ for $r\gg\ell_>(t)$ and assume that for some appropriate $\ell_>(t)$, this is indeed an upper bound for all $t<T$; i.e., $q(r,t)\leq F_>[r/\ell_>(t)]$. We can now use [S33] with q_S and q both replaced by $F_>$.

When $\ell_>(t)$ and $\ell_>(T-t)$ are both much less than R, the integrals over x and y will be dominated by the regions near the origin and R, respectively, yielding the spatial convolution $F_>\circ G\circ F_> \sim \ell_>(t)\ell_>(T-t)/R^{\mu+1}$. There are small positive corrections to this from two sources: first, from the regions near 0 and R, which, by expanding $y-x$ in x and $R-y$, are seen to be of order $[\ell_>(t)^2+\ell_>(T-t)^2]\ell_>(t)\ell_>(T-t)/R^{\mu+3}$; and second, from $y-x\ll R$, the regions near the source boundary, which are of order $[\ell_>(t)/R]^{\mu+1}[\ell_>(T-t)/R]^{\mu+1}R^{1-\mu}$ with the last part from the integrals over x and y. As the dominant part is exactly of the form in the heuristic treatment, integrating it over time is strongly peaked at $t\approx T/2$ (note that for either t or $T-t$ much smaller than T, one of the $F_>$ factors will be close to unity near the boundary, but these ranges of time contribute only weakly). If we use $1-e^{-Q}\leq\min(Q,1)$, then $\ell_>(T)$ can be chosen as the value of R for which $Q=1$, and for $R\gg\ell_>(T)$ the desired $F_>\sim[\ell_>/R]^{\mu+1}$ is obtained. Including the small correction factors in the convolutions necessitates slight modifications of the recursion relations for $\ell_>$ but these are negligible at long times.

4. Intermediate-range case. Obtaining an upper bound in the intermediate-range case is somewhat trickier. If we again replaced q_S by q, then the integrals over x and y would have a part dominated by both points being near the boundary: With $B_S\sim R$, this contribution to the convolution would be of order $[\ell_>(t)/R]^{\mu+1}[\ell_>(T-t)/R]^{\mu+1}$. With $t\sim T/2$ and $R\sim\ell(T)$, all of the lengths should be of order t^β with $\beta=1/(\mu-d)$, so that this boundary piece is larger by a factor of $[\ell(T)^\beta]^{\mu-1}\sim T$ than what should be the dominant part from x and y near 0 and R, respectively. Thus, we need a better upper bound on the restricted-source $q_S(r,t)$, which vanishes as $r\nearrow B_S(t)$.

To bound q_S, we can make use of the simple bound for the edge of the occupied region derived above, combined with the restrictive effects of the boundary, $B_S(t)$. Instead of choosing B_S to be constant, we choose it to have constant slope, $U\equiv dB_S/dt$, of order $\ell(T)/T$. As jumps that contribute to q_S are not allowed to cross the boundary, the distribution of jumps of the edge $E(t)$ is cut off at $\zeta(t)\equiv B_S(t)-E(t)$. Because $U>v_0$, the speed of spread in the absence of jumps beyond nearest neighboring sites, typically the gap, $\zeta(t)$, will increase with time, decreasing only by jumps. The sum of all of the jumps of E in a time interval Δt is dominated by the largest, which is of order $(\Delta t)^{1/(\mu-1)}$. This would result in the edge moving faster than U except for the cutoff. The typical gap, $\tilde{\zeta}$, is then obtained by balancing its steady decrease against the dominant jump: $U\Delta t\sim(\Delta t)^{1/(\mu-1)}$, yielding $\Delta t\sim U^{(\mu-1/)(2-\mu)}$ and hence

$$\tilde{\zeta}\sim U^{1/(2-\mu)}\sim\left[\frac{\ell(T)}{T}\right]^{1/(2-\mu)}\sim\ell(T)A_\mu^{(\mu-1)/(2-\mu)}, \quad \text{[S34]}$$

using $\ell(t)\sim A_\mu t^{1/(\mu-1)}$. In the limit of $\mu\searrow 1$, $\tilde{\zeta}/\ell\sim 4^{-1/(\mu-1)}$, vanishing rapidly—a reflection of the strong failure of the simple edge bound in this limit but sufficient for our present purposes. In a time $\Delta t\ll T$, the distribution of ζ in this approximation will reach a steady state. The probability that $\zeta\ll\tilde{\zeta}$ is controlled by the balance between jumps of E to near the boundary, and the steady increase in ζ from the boundary motion: Its probability density is hence of order $\zeta/\tilde{\zeta}$, which, because in this approximation all sites are occupied up to E, implies that q_S vanishes at least quadratically for small gap ζ. Combining this with the trivial bound of $q_S<q$ and choosing a convenient normalization of $\tilde{\zeta}$, we thus have

$$q_S(r,t)\leq\min\left[F_>(r,t),\frac{(B_S(t)-r)^2}{\tilde{\zeta}^2}\right]. \quad \text{[S35]}$$

It remains to choose $B_S(t)$ so that the bound on q_S remains sufficiently good for t small enough that the steady-state distribution of $\zeta(t)$ has not yet been reached. To keep $E(t)$ typically of order $\tilde{\zeta}$ from $B_S(t)$, we can simply chose $B_S(0)=\tilde{\zeta}$ and $U=(R-2\tilde{\zeta})/T$.

With our improved bound on q_S, for the convolution $q_S\circ G\circ q$, the small B_S-x parts are no longer dominated by B_S-x of order unity, but by B_S-x near the crossover point between the two bounds on q_S. This yields a contribution to the convolution of order $\ell_>(t)^{\alpha_S}\ell_>(T-t)^{\alpha_F}/R^{\alpha_S+\alpha_F+\mu-1}$ with $\alpha_F=(\mu+1)(2-\mu)$ and $\alpha_S=\alpha_F(3-\mu)/2$ and a multiplicative coefficient that does not depend exponentially on $1/(\mu-1)$ because the integral over x scales as $1/\tilde{\zeta}^{\mu-1}$. As $\mu\searrow 1$, $\alpha_F\to\alpha_S\to 1$, and the boundary contribution is less than the dominant part uniformly in t. Note that for $\mu>\mu_B\cong 1.5$, the bound on the near-boundary contribution can be somewhat larger for $t<T/2$ than the dominant parts, but it scales in the same way with T and thus only weakens the upper bound on the coefficient, A_μ, which is in any case of order unity in this regime.

Once the overcounting of short jumps has been sufficiently reduced, as we have now done, the rest of the analysis, in particular the large $R/\ell_>$ form of $F_>$, follows as in the long-range case.

The marginal case $\mu=1$ can be analyzed similarly to the intermediate-range case, resulting in an additional logarithmic dependence on R of the near-boundary contribution, which is, nevertheless, still much smaller than the dominant part.

The upper bounds that we have obtained are, except for modifications at small scales and for μ not much smaller than 2, essentially the same as given by the heuristic arguments, thus differing at long scales only by order-unity coefficients that, in any case, we did not expect to get correctly. All of the crossover behavior near $\mu = 1$ is in the upper bounds, although that near $\mu = 2$ is not.

B. Lower Bounds. To obtain lower bounds on the growth of the characteristic length scale $\ell(t)$ and the occupation probability, $q(r,t)$, a different strategy needs to be used. One of the difficulties is the dependence on the behavior at each timescale on all of the earlier timescales: For the filled region to grow typically between times T and $2T$, the stochastic processes that lead to the configuration $c(x,T)$ must not have been atypically slow or ineffective. As this applies iteratively scale by scale, we must allow for some uncertainty in whether the smaller-scale regions are typical, leading to some uncertainty at all scales, which, nevertheless, we need to bound. Because of the stochastic heterogeneity of $c(x,t)$, it is better to focus on a coarse-grained version of the occupation profile rather than on $c(x,t)$ itself, as integrations over c at time t are what act as the sources of future occupation at larger distances.

1. Mostly filled in: Marginal and long-range regimes. We consider the probability that a region is almost full; in particular, with a seed at the origin, we consider the region to one side of the origin and define

$$P_F(r,t;\Phi) \equiv \mathcal{P}\left[\frac{1}{r}\int_0^r dx\, c(x,t) > \Phi\right] \qquad \text{[S36]}$$

with Φ close to or equal to unity being of particular interest. To keep events sufficiently independent, we consider, as for the upper bounds, the probability of events that do not involve any jumps out of some region. In particular, we define $P_S(r,t,\Phi)$ similarly to P_F, but with the restriction that jumps do not go out of the interval $(0,r)$. For the long-range and marginal cases, we focus on partial filling, but for the intermediate range case the scale invariance mandates different treatment so we instead analyze full filling—i.e., $\Phi = 1$.

The basic strategy is to start with a particular deterministic approximation to $\ell(t)$, $\tilde{\ell}(t)$ with corresponding times $\tilde{\tau}(r)$, and then show that at time not too large a multiple of $\tilde{\tau}(r)$, the region out to r will be nearly filled with high probability: i.e., that $P_S(r,\tau_>(r);\Phi)$ is close to unity for $\tau_>(r)/\tilde{\tau}(r)$ sufficiently large. We are interested in large scales as, in any case, fluctuations at the small scales can change coefficients only by order unity. We can thus be sloppy with some of the bounding inequalities: These could be improved to include the ignored corrections to the large-scale effects to make fully rigorous bounds.

As the range of time over which the typical $\ell(t)$ expands significantly plays an important role, it is useful to define

$$\tilde{D}(t) \equiv \left[\frac{d\log\tilde{\ell}(t)}{d\log t}\right]^{-1}, \qquad \text{[S37]}$$

which is small except for μ substantially larger than one. The dominant jumps from source to funnel involve an integral over time of $\tilde{\ell}(t)\tilde{\ell}(T-t)$, which is primarily from a range of order $T\sqrt{\tilde{D}}$ around T as discussed above. The deterministic-iterative approximation that we use as a base for the lower bounds is the solution to the iterative relation [ambiguous up to an $\mathcal{O}(1)$ multiplicative factor, which we ignore throughout]

$$\left[\tilde{\ell}(2T)\right]^{\mu+1} = T\left[\tilde{\ell}(T)\right]^2\sqrt{\tilde{D}(T)}, \qquad \text{[S38]}$$

corresponding to roughly one seed into a funnel of width $\tilde{\ell}(T)$ from a jump of distance $\tilde{\ell}(2T)$ from the source up to time T. For convenience, we use only half the source—x from 0 to $\tilde{\ell}(T)$. The results, $\tilde{\ell}(t)$, of this iterative approximation are, up to numerical factors that arise from these modifications and from other from nonasymptotic effects at small scales, equivalent to the upper bounds, $\ell_>(t)$ from the above. In particular, we expect the ratio between the corresponding times, $\tilde{\tau}(r)$ and $\tau_<(r)$, to approach constants that are not singular near the marginal case $\mu = d = 1$.

For the lower bounds it is convenient to work with a specific set of length scales, $\tilde{\ell}_n = \tilde{\ell}(\tilde{\tau}_n)$, corresponding to a series of timescales, $\tilde{\tau}_n = 2^n$ (dropping a prefactor). To mostly fill out to $\tilde{\ell}_{n+1}$ without jumps going out of $(0,\tilde{\ell}_{n+1})$ from the source of size $\tilde{\ell}_n$, most of the $K_n = \tilde{\ell}_{n+1}/\tilde{\ell}_n$ bins of size $\tilde{\ell}_n$ must be mostly filled. To get a lower bound on how long this takes and how likely it is, we make several simplifications, each of which leads to underestimates of the probability that the desired filling has occurred. First, consider only jumps into each bin that come directly from the source (rather than from other bins as can occur later). Second, ignore all but the first seed jump from the source into the bin (the effects of later jumps are not independent of those of the first). And third, include only jumps that lead from the seed in a bin that do not go outside that bin during the time during which the probability of it being mostly filled is considered. The last two conditions mean that the probability that the bin is filled to a fraction Φ by a given time, t, after the seeding jump, is at least as large as $P_S(\tilde{\ell}_n,t;\Phi)$ because a seed at the edge of the bin, which corresponds to the definition at the source, is less likely to mostly fill the bin than a seed away from the edge.

At large scales for $\mu \le 1$, the number of bins, K_n, grows with scale $K_n \sim \sqrt{\tilde{\tau}_n}$ for the marginal case and larger for the long-range case. Thus, if the probability that the furthest bin from the source is mostly filled is f_n, with the filling of the others being more probable as they are closer, it is likely that the number that are similarly mostly filled is close to $K_n f_n$, with significant deviations from this being very unlikely at large scales. To iterate while not losing too much in the filling fraction, we chose a series of partial filling fractions, $\{\phi_n\}$, such that $\Phi_N \equiv \prod_{n=1,N-1}\phi_n$ converges to the desired overall filling fraction, Φ, at large N, and chose conditions such that f_n is sufficiently large that the fraction of the K_n bins filled to Φ_n is greater than ϕ_n with high probability: This then implies that the region from the origin to $\tilde{\ell}_{n+1}$ will be filled to greater than Φ_{n+1} with high probability. A convenient choice is $\phi_n = 1 - \Delta/n^{1+\alpha}$ with any positive α and $\Delta\sum_n n^{-1-\alpha} < 1-\Phi$. For convenience in dropping $\log\Delta$ factors that otherwise appear in many places, we restrict consideration to Δ not very small and do not keep careful track of α factors that also appear as we can take $\alpha \to 0$ at the expense of corrections that are down by one extra logarithm.

The filling probability of a bin is at least as large as that obtained from the requirement of the occurrence of both of two independent events: a jump into the bin from the source that occurs before some chosen initial time, T_I, and the bin being filled from that single seed by a time, $T_B + T_I$. The probability of a jump into a bin is at least $1 - e^{-W_n}$ in terms of a conveniently chosen lower bound, W_n, on the expected number of jumps from the source into the farthest away bin, and the probability of the bin being filled from the single seed is at least $P_S(\tilde{\ell}_n, T_B; \Phi_n)$. We find iterative bounds on P_S that are convenient to write in the form

$$P_S(r,t;\Phi) \ge 1 - e^{-\Lambda(r,t;\Phi)} \qquad \text{[S39]}$$

so that

$$1 - f_n \le e^{-W_n} + e^{-\Lambda_B} \quad \text{with} \quad \Lambda_B \equiv \Lambda\left(\tilde{\ell}_n, T_B; \Phi_n\right). \qquad \text{[S40]}$$

For convenience we chose conditions so that $\Lambda_B \ge W_n$ and $1 - f_n \le (1/2)(1-\phi_n)$, which, for K_n large, makes the probability that a fraction ϕ_n of the bins are not filled exponentially small.

We henceforth ignore this factor in the probability as it does not matter except on small scales: Adjustments to take it into account are straightforward. We thus require that

$$\Lambda_B \geq W_n \geq \log\left[\frac{(1-\phi_n)}{4}\right] = (1+\alpha)\log n + \mathcal{O}(1). \qquad \textbf{[S41]}$$

To obtain a bound on $P_S(\tilde{\ell}_{n+1}, T; \Phi_{n+1})$, we must show that a source that can give rise to an average number at least W_n of jumps into the farthest bin by time T_I occurs with probability that is somewhat larger than the desired bound at the next scale. The expected effective number of jumps out of the source of size $\tilde{\ell}_n$ into a bin of the same size a distance up to $\tilde{\ell}_{n+1}$ away before time $\tilde{\tau}_n$ was assumed in the deterministic iterative approximation to be of order $\sqrt{\tilde{D}_n}\tilde{\tau}_n$. To ensure that the average number from the actual source is sufficiently large, we can require that it be almost filled by some time T_S and include only jumps that occur between T_S and T_I as the rate of these is bounded below by the filling at T_S. The required range is

$$T_I - T_S = \frac{W_n \tilde{\tau}_n \sqrt{\tilde{D}_n}}{\Phi_n}. \qquad \textbf{[S42]}$$

We now proceed by induction and show that if

$$\Lambda\left(\tilde{\ell}_n, t; \Phi_n\right) \geq \gamma_n (t - U_n \tilde{\tau}_n) \qquad \textbf{[S43]}$$

for t in a range such that Λ is relatively large—the precise range is not crucial but minor modifications are needed to extend out to arbitrary large t—then a similar bound holds at the next scale with coefficients γ_{n+1} and U_{n+1} with both these varying slowly with n at large scales. Note that at the smallest scale the probability that a site is filled by a jump directly from the origin by time t converges exponentially to unity for long t, and thus at the smallest scales there is a trivial bound of this form. As the scale is increased, γ_n will initially change, but once the scale becomes large enough that the width of the distribution of the fraction of the bins mostly filled is small, then γ_n saturates and becomes weakly dependent on n. In the analysis below, it can be replaced by a constant.

Consider a total time T to mostly fill out to $\tilde{\ell}_{n+1}$. The time for the bins to fill with sufficiently high probability once they have been seeded is $T_B \leq U_n \tilde{\tau}_n + W_n/\gamma_n$. With $T_I - T_S$ as above, we have a time for the source to fill

$$T_S \geq T - \tilde{\tau}_n \left[U_n + \frac{W_n \sqrt{\tilde{D}_n}}{\Phi_n}\right] - \frac{W_n}{\gamma_n}. \qquad \textbf{[S44]}$$

Plugging in the probability that the source is filled in this time gives a bound on $\Lambda(T, \tilde{\ell}_{n+1}, \Phi_{n+1})$ of the same form but with, dividing out $\tilde{\tau}_{n+1} = 2\tilde{\tau}_n$,

$$U_{n+1} \leq U_n + \frac{W_n \sqrt{\tilde{D}_n}}{2\Phi_n} + \frac{W_n}{2\gamma\tilde{\tau}_n}. \qquad \textbf{[S45]}$$

As $\tilde{\tau}_n$ increases rapidly and W_n only slowly, the last term contributes only at small scales.

For the long-range regime, $\tilde{D}_n \sim e^{-\eta \log 2 n}$ so the second term in **[S45]** is also small except at small scales and we conclude that U is bounded above by a μ-dependent constant. Thus, the lower bound for $\ell(t)$ and the upper bound for $\tau(r)$ have exactly the same form as the opposite bounds, except with the scale of t—i.e., $B_\mu^{-1/\eta}$—different.

For the marginal case, $\tilde{D}_n \approx 2/n$ so that U_n changes slowly at large scales. Integrating up, we see that

$$U_n < C\sqrt{n}\log n \sim \sqrt{\log \tilde{\tau}_n}\,\log\log\tilde{\tau}_n \qquad \textbf{[S46]}$$

with a coefficient independent of n (but depending on Φ and α). We can now solve for the timescale above which mostly filled is likely, $\tau_>(r) = U(\tilde{\tau}(r))\tilde{\tau}(r)$, to find a lower bound, $\ell_<(t)$, on $\ell(t)$,

$$\log(\ell(t)) \geq \log(\ell_<(t)) = \frac{\log(t)}{4\log 2}[\log t - 2\log\log t - \mathcal{O}(\log\log\log t)], \qquad \textbf{[S47]}$$

which is very close to the upper bound derived above,

$$\log(\ell(t)) \leq \log(\ell_>(t)) = \frac{\log(t)}{4\log 2}[\log t - \log\log t - \mathcal{O}(1)], \qquad \textbf{[S48]}$$

differing only in the coefficient of the correction term.

One of the advantages of this iterative approach is that the crossover regime can be handled similarly by integrating up **[S45]**. The lower bound will be similar to the upper bound throughout this crossover regime and into the asymptotic regimes for the marginal and long-range cases.

2. Fluctuations and intermediate-range regime. The reason that the fluctuation effects are relatively small for the marginal and long-range regimes is that at each successive timescale, more and more roughly independent long jumps are involved in filling up to the next length scale; i.e., K_n continues to grow. For the long-range regime, it grows so rapidly that almost all of the fluctuations come from early times: This is like what occurs for the fully mixed model. For the marginal case, the fluctuations are dominated by the smallest scales but the cumulative effects of them over the longer scales do make a difference as found in obtaining the lower bounds.

For the intermediate-range case, the ratio of length scales for each factor of 2 in the timescale saturates (when out of the crossover regime) at $K \approx 2^\beta$. This means that whatever fraction, ϕ_n, of the bins are to be filled at each scale, the probability that this occurs either decreases with scale if the product of the ϕ_n does not go to zero or saturates to a constant if the ϕ_n s do also, in which case the overall filling fraction Φ_n tends to zero as a power of time. At each scale there are a comparable number of long jumps that are needed, and thus we should expect that fluctuation effects will be scale invariant and not decrease with scale.

To get useful lower bounds on $\ell(t)$ via an upper bound on $\tau(r)$, $\tau_>(r)$, the easiest way is to require that the source be completely full and that jumps from this completely fill all of the bins at the next scale: This avoids the problems with the ϕ_n. As the probability that all of the bins are filled is (readily) bounded only by $(1 - e^{-W} - e^{-\Lambda_B})^K \approx 1 - 2Ke^{-W}$ if we again chose $\Lambda_B \geq W$, W must be larger by $\log K \approx \beta \log 2$ than for the partially filled analysis above. Carrying through to a similar analysis gives for large β a coefficient $\gamma_n \approx 2\sqrt{\beta}/(n\tilde{\tau}_n)$, which means that the filling probability decays for large times as roughly the inverse of the typical time—natural as the needed long jumps that occur at rate $\sim 1/\tilde{\tau}_n$ have a distribution of when they occur on the same timescale. Note the contrast to the rapid decay of the not mostly filled probability on a timescale of order unity from the partially filled bound derived above. The time beyond which the full filling is likely is bounded only by, in this analysis, $\tilde{\tau}_n U_n \sim \tilde{\tau}_n \sqrt{\beta} n^2$. This gives a lower bound on $\ell(t)$ proportional to $t^\beta/\log^{2\beta} t$. Although on a logarithmic scale the additional factor is smaller, we wish to do better.

The bound can be improved by considering a source that is somewhat smaller—by a factor of 2 is sufficient—than $\tilde{\ell}_n$, which

increases the probability that it is filled, but means that more extra time, $T_I - T_S$, is needed to produce a mean number of jumps to the farthest bin of at least W. Using that U and $\gamma\tilde{\tau}$ vary slowly with scale, one can expand $\Lambda(r,t)$ around $r=\tilde{\ell}_n$ and analyze the changes on the bounds at the next scale. This improves the bounds to $\gamma \sim 1/(\sqrt{\beta}\tilde{\tau})$ and $U \sim \beta^{5/2}$ with $\beta = 1/(\mu-1)$. The resulting lower bound on $\ell(t)$ is

$$\ell(t) > \widetilde{A_\mu}\beta^{-\frac{5\beta}{2}}t^{\beta} \qquad \textbf{[S49]}$$

with the coefficient $\widetilde{A_\mu} \sim 4^{\beta^2}$ that from the deterministic iteration, which is the same, up to an order unity prefactor, as the upper bound. It is not clear where between the lower and upper bounds on the coefficient will be the typical behavior or how broad the fluctuations will be—even on a log scale.

The behavior as $\mu \nearrow 2$ we have not analyzed explicitly, instead focusing on the rapidly growing regime for $\mu \searrow 1$, but the bounds will be of similar form although more care is needed to get upper and lower bounds reasonably close to one another due to the important jumps being only a modest fraction of the size of the already occupied region.

For the marginal case, $\mu = 1$, one can find an upper bound on the time at which the region out to r is likely to be fully filled by similar methods to that for the intermediate-range power-law regime. This yields a bound $\ell_<^{\Phi=1}(t)$ of the same form as that above for partial filling ($\Phi < 1$), except with the coefficient of the $\log\log t$ term in [**S47**] equal to 6 instead of 2. The convergence of the probability of being fully filled is, however, much slower for this bound on compete filling than for the bound on being mostly filled. Whereas the latter converges for $t > \tau_>^{\Phi}(r) \sim \tilde{\tau}(r)\sqrt{\log\tilde{\tau}(r)}\log\log\tilde{\tau}(r)$ with a rate of order unity—dominated by the small scales—the former converges as the time increases above $\tau_>^{\Phi=1}(r) \sim \tilde{\tau}(r)\log^{5/2}\tilde{\tau}(r)$ on a timescale, $1/\gamma$, of order $\tilde{\tau}(r)\sqrt{\log\tilde{\tau}(r)}$—faster than $\tau_>^{\Phi=1}(r)$ but not much so.

Note that the convergence of the probability for being mostly filled to more than a fixed filling fraction, Φ, is a hybrid property: The probability of a fixed site being filled by time t, $q(r,t)$, is bounded below by (roughly) the product of Φ and the probability that the region out to r is filled to above Φ. To get the convergence of this to unity, Φ needs to be adjusted and the thus-far ignored $\log(1-\Phi)$ factors kept track of. This also necessitates treating intermediate scales differently as the number of bins that do not need to be filled, $(1-f_n)K_n$, is not large. However, a different approach would provide a better bound: Focusing on a specific site being filled with high probability can be done by a method more analogous to the funnel picture in the main text. For the site to be filled, it needs to be in a small-scale bin that is mostly but not necessarily fully filled with high probability, which needs itself to be in a larger bin similarly, etc. However, these can be filled from source regions that are not fully filled: Being partly filled with high enough probability is sufficient. We have not carried out such analysis in detail in part because the actual mechanism by which sites that are empty for an anomalously long time will be filled is more complicated as it will involve filling from nearby regions on a hierarchy of scales that were filled at more typical times.

The analyses here can be immediately extended to give lower bounds on the average density profile at long distances: These will be of the same form as the lower bounds, thus demonstrating that the predicted $\ell(t)^{-(1+\mu)}/r$ is essentially correct.

3. Comparisons with results of Chatterjee and Dey. As noted in the main text, when this work was essentially complete, a preprint by Chatterjee and Dey (CD) appeared, which derives and proves some results closely related to ours in the context of long-range first passage percolation, which is essentially equivalent to the lattice dispersal model, with the jump kernel $G(r) \sim r^{-\alpha}$ equivalent to the $1/r^{-(d+\mu)}$ that we use (2). Although some of the quantities CD focus on are different, the leading asymptotic scaling behaviors they obtain are essentially the same, and their proofs apply in all dimensions. Our $q(r,t)$ corresponds to the probability that the first passage time $T^F(r)$ is less than t and their diameter, $D(t)$—the maximum distance between any pair of occupied points at time t—is, with high probability that decays as a power of ℓ/D—not many times $\ell(t)$, as we both obtain.

CD give heuristic arguments for the extent of the linear regime, the behavior in the power-law regime, and where this breaks down (at $\mu = d$), which are related to ours. Some of their bounds also make use of inequalities related to the simplified form of our self-consistency condition, Eq. **4**.

However, CD's results are suboptimal. In particular, for the coefficient, C, of $\log\ell(t)/\log^2 t$ in the marginal case, they obtain only upper and lower bounds instead of our exact result $C = 1/4d\log 2$. Indeed, in one dimension we obtain rather tight upper and lower bounds on the errors,

$$1 - c_-\frac{\log\log t}{\log t} < \frac{4\log 2\log\ell(t)}{\log^2 t} < 1 - \frac{\log\log t}{\log t} \qquad \textbf{[S50]}$$

with high probability—in senses that can be made precise from our analysis—with the coefficient c_- either 2 or 6, depending on the definition of $\ell(t)$ used. In the intermediate-range power-law growth regime, CD's theorems do not appear to exclude $\log t$ prefactors in $\ell(t)$, although their analysis might well do so. However, the main difference is our analysis of the whole crossover regime for μ near d, including the divergences and vanishings of coefficients of the asymptotic forms, which they do not consider. These are crucial for comparisons with simulations because of the very long length scales of the crossovers.

To turn our upper and lower bounds into formal proofs in one dimension requires primarily filling in some details associated with the small-scale regime. For higher dimensions substantial additional work may be needed, although we believe the strategies we developed here should work without major modifications.

1. Mollison D (1972) The rate of spatial propagation of simple epidemics. *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, eds Le Cam LM, Neyman J, Scott EL (Univ of California Press, Berkeley, CA), Vol 3, pp 579–614.

2. Chatterjee S, Dey PS (2013) Multiple phase transitions in long-range first-passage percolation on square lattices. *arXiv*:1309.5757.

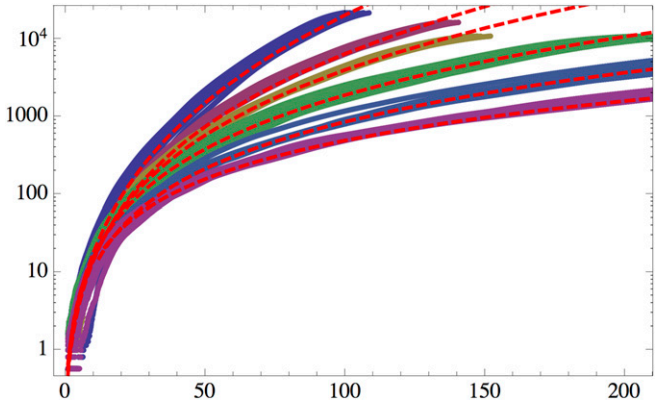

Fig. S1. Summary of the spreading dynamics in two spatial dimensions. The effective radius $\sqrt{M(t)/\pi}$ of the region occupied by the mutant population is plotted as a function of time t, for various long-range jump kernels. Each colored cloud represents data obtained from 10 runs for a given jump kernel with tail exponent μ as indicated. Red dashed lines represent predictions, obtained from Eq. **8** (main text) with fitted crossover scales. The jump exponents are $\mu \in \{1.6,1.8,1.9,2.1,2.3,2.5\}$.

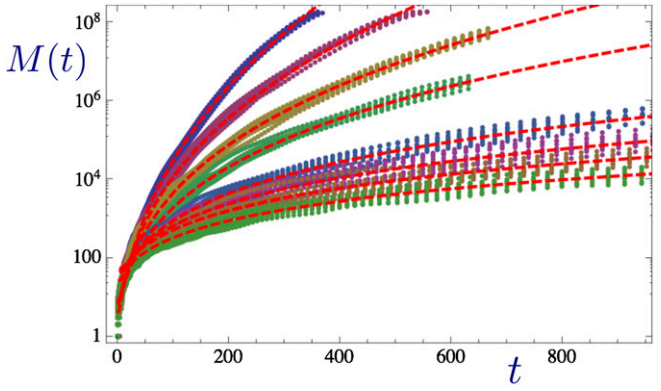

Fig. S2. Summary of the spreading dynamics in one spatial dimension with short-range as well as long-range dispersal. The total number of occupied sites $M(t)$ (the "mass") is shown as a function of time. For these simulations each cluster expands at a linear speed even in the absence of long-range jumps. In each time step, a long-range jump is performed only with probability $\tilde{\epsilon} = 0.1$. For the short-range part, a pair of neighboring sites is chosen at random. If this pair happens to fall on a boundary of a mutant cluster, i.e., the identity of both sites is mixed, then the wild-type site is switched to a mutant site. This leads to expansion of mutant clusters at average speed of $v_0 = 2$ sites per generation. The jump exponents are $\mu \in \{0.6,0.7,0.8,0.9,1.0,1.1,1.2,1.3,1.4\}$ from the fastest- to the slowest-growing case. The theoretical predictions indicated by the red dashed lines fit the data after choosing appropriate crossover timescales and length scales between linear growth and superlinear growth.

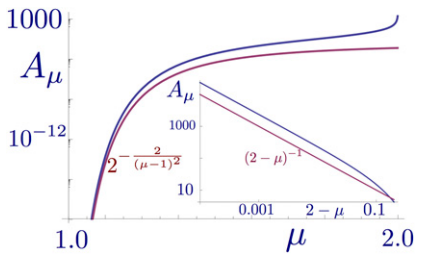

Fig. S3. The prefactor A_μ of the predicted power law growth in Eq. **3** (main text) in one dimension: $\ell(t) \approx A_\mu t^\beta$ with $\beta = 1/(\mu - 1)$ for $1 < \mu < 2$. The blue curve is obtained numerically from solving Eq. **1** (main text) with the power-law ansatz; the red curve represents an analytic approximation derived in *Crossovers and Beyond Asymptopia* (main text). Note the sharp (nonanalytic) drop of the prefactor as μ approaches 1. The reason is very slow crossover to the power law from an intermediate asymptotic regime controlled by the dynamics of the marginal case. As μ approaches 2, the prefactor diverges as $A_\mu \sim (2-\mu)^{-1}$, indicative of another slow crossover at $\mu = 2$ (Fig. S4).

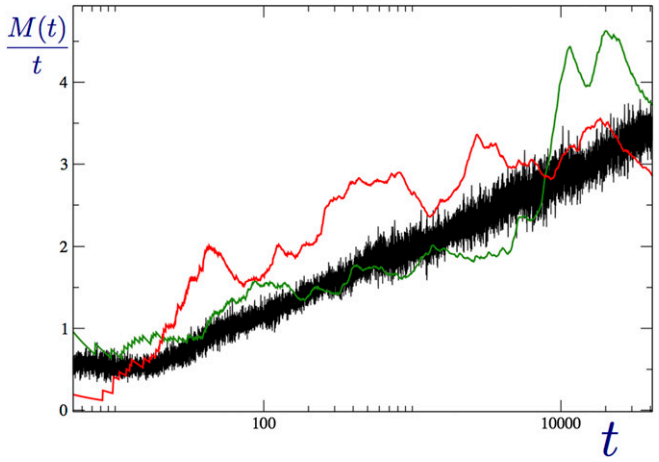

Fig. S4. Dynamics of growth in one dimension at the marginal point between superlinear and linear growth, $\mu = 2$. The number of mutant sites, $M(t)$, scaled by time, $M(t)/t$, is plotted as a function of time, averaged over 10 realizations (black) and for two individual realizations (red and green). Whereas the averaged data suggest $M(t) \sim t \ln(t)$, the individual realizations indicate strong fluctuations caused by occasional rare jumps, which are of order t. These leaps forward (1) are driving the logarithmic increase of the spreading velocity.