Size of Outbreaks Near the Epidemic Threshold

E. Ben-Naim*
Theoretical Division and Center for Nonlinear Studies,
Los Alamos National Laboratory, Los Alamos, NM, 87545

P. L. Krapivsky†
Center for BioDynamics and Department of Physics, Boston University, Boston, MA, 02215

The spread of infectious diseases near the epidemic threshold is investigated. Scaling laws for the size and the duration of outbreaks originating from a single infected individual in a large susceptible population are obtained. The maximal size of an outbreak n_* scales as $N^{2/3}$ with N the population size. This scaling law implies that the average outbreak size $\langle n\rangle$ scales as $N^{1/3}$. Moreover, the maximal and the average duration of an outbreak grow as $t_* \sim N^{1/3}$ and $\langle t\rangle \sim \ln N$, respectively.

PACS numbers: 02.50.-r, 05.40.-a, 87.23.Cc, 87.19.Xx

Infection processes typically involve a threshold [1–9]. Below the epidemic threshold, outbreaks quickly die out, while above the threshold, outbreaks may take off. We study epidemic outbreaks near the threshold. Such outbreaks arise naturally. On the one hand, human efforts at disease prevention reduce the infection rate thereby crossing the epidemic threshold [2]. On the other hand, evolution may increase the infection rate of diseases hovering just below the threshold, enhancing the likelihood of near-threshold outbreaks [10]. Typically, detection, modeling, and eradication of infectious diseases are subtle for outbreaks near the epidemic threshold.

The total number of infected individuals is a basic measure of the severity of an epidemic outbreak. We study outbreaks originating from a single infected individual in a large susceptible population. Our main result is that near the epidemic threshold, the maximal outbreak size n_* grows as a power-law of the population size N,

$$n_* \sim N^{2/3}\,. \tag{1}$$

In contrast, below the epidemic threshold, endemic outbreaks involve a small number of infected individuals, while above the epidemic threshold, pandemic outbreaks involve a fraction of the population $n_* \sim N$. Therefore, outbreaks near the epidemic threshold have a distinct intermediate size between a pandemic and an endemic outbreak [1]. Loosely speaking, epidemics come in three sizes: large, medium, and small.

The scaling law (1) has several important implications concerning the statistics of both the size and the duration of the outbreaks. It implies that the average size of outbreaks $\langle n\rangle$ and the maximal duration of outbreaks t_* both scale as $\langle n\rangle \sim t_* \sim N^{1/3}$ near the epidemic threshold. Furthermore, the average duration of the outbreaks $\langle t\rangle$ scales logarithmically, $\langle t\rangle \sim \ln N$. These behaviors hold in a sizable range of infection rates, namely in a window of the order $\mathcal{O}(N^{-1/3})$ around the epidemic threshold.

These scaling laws are demonstrated for the classic Susceptible-Infected-Recovered (SIR) infection process [1–3]. In this model, the population consists of s susceptible, i infected, and r recovered individuals with $N = s + i + r$. These sub-populations change due to two competing processes: infection and recovery. The disease is transmitted from an infected individual to a susceptible one with rate α/N, where α is the infection rate:

$$(s, i, r) \xrightarrow{\alpha si/N} (s-1, i+1, r). \tag{2}$$

Infected individuals recover with a unit rate:

$$(s, i, r) \xrightarrow{i} (s, i-1, r+1). \tag{3}$$

The infection process starts with a single infected individual, $(s, i, r) = (N-1, 1, 0)$, and it ends with none $(s, i, r) = (N-n, 0, n)$.

The total size of the outbreak n and the duration of the outbreak t are the outcomes of a stochastic process. We study statistical properties of these random variables, particularly, their average and maximal size, as a function of the population size (We implicitly consider an average over infinitely many realizations of the infection process.)

In the infinite population limit, the epidemic threshold is $\alpha = 1$. Since infection occurs with probability $\frac{\alpha}{1+\alpha}$ and recovery with probability $\frac{1}{1+\alpha}$, the average outbreak size satisfies $\langle n\rangle = \frac{1}{1+\alpha} + 2\frac{\alpha}{1+\alpha}\langle n\rangle$. Thus, below the threshold ($\alpha < 1$), a finite number of individuals is infected, $\langle n\rangle = (1-\alpha)^{-1}$. Above the threshold ($\alpha > 1$), there is a pandemic outbreak with a finite fraction of the population infected: $\langle n\rangle = rN$ [1, 11]. At the threshold ($\alpha = 1$), the probability that the outbreak size equals n, G_n, is found recursively: $G_n = \frac{1}{2}\sum_{m=1}^{n-1} G_m G_{n-m}$ starting with $G_1 = 1/2$. This recursion reflects the fact that the first infection event results in two independent infection processes [12]. The generating function underlying this standard branching process is $\sum_{n\geq 1} G_n z^n = 1 - \sqrt{1-z}$, from which the size distribution is a power-law,

$$G_n \sim n^{-3/2}, \tag{4}$$

*Electronic address: ebn@lanl.gov
†Electronic address: paulk@bu.edu

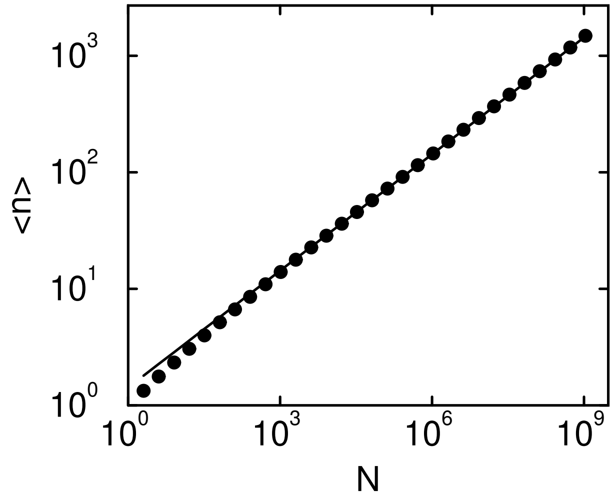

FIG. 1: The average outbreak size versus the population size for the SIR infection process at the epidemic threshold ($\alpha = 1$). Shown are Monte Carlo simulation results representing an average over 10^9 independent realizations of the infection process (circles). A line of slope 1/3 is also shown as a reference. A least-square-fit to $\langle n\rangle \sim N^{\gamma}$ in the range $10^3 < N < 10^9$ yields $\gamma = 0.334 \pm 0.001$.

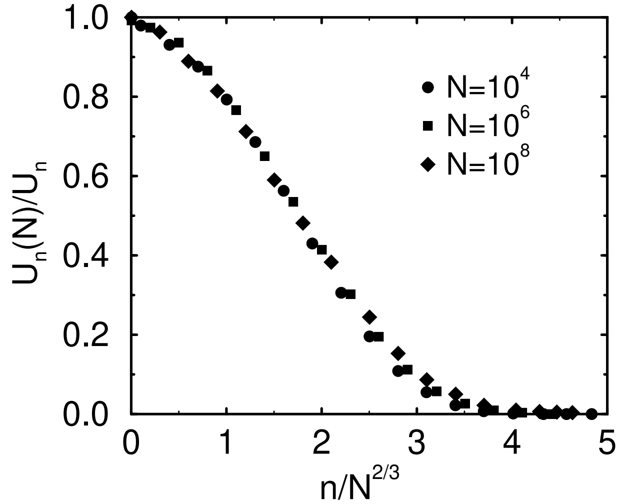

FIG. 2: The normalized cumulative distribution $U_n(N)/U_n(\infty)$ versus the normalized outbreak size $n/N^{2/3}$. The data corresponds to an average over 10^6 independent realizations.

for sufficiently large outbreaks $n \gg 1$.

For a finite, yet large population, the outbreak size distribution (4) holds, but only up to the maximal outbreak size: $1 \ll n \ll n_*$. Outbreaks beyond the maximal size are practically impossible. Therefore, the average outbreak size grows according to $\langle n\rangle = \sum_{n=1}^{n_*} nG_n \sim n_*^{1/2}$. Naively assuming that a finite fraction of the population may become infected, $n_* \sim N$, would lead to $\langle n\rangle \sim N^{1/2}$. While consistent with the generic statistical uncertainties, this law is in fact *erroneous*. Instead, the outbreak size is much smaller because the epidemic outbreak weakens as more individuals become infected, and it finally dies out when the number of infected individuals becomes of the order n_*. When there are $N - n_*$ susceptible individuals, the total infection rate $\alpha(N-n_*)i/N$ shows that the infection rate is effectively reduced, $\alpha_{\rm eff} = 1 - n_*/N$. Therefore, the epidemic becomes essentially endemic. (This is clearly a finite population effect: the susceptible population "reservoir" is never affected in the infinite population limit.) Equating the outbreak size in the endemic phase $\langle n\rangle \sim (1-\alpha_{\rm eff})^{-1} \sim N/n_*$ with that estimated from the size distribution, $\langle n\rangle \sim n_*^{1/2}$, gives the scaling law (1) governing the maximal outbreak size. Hence, in the worse case scenario, only a fraction of the order of $N^{-1/3}$ of the entire population can ever be infected.

As a byproduct we obtain the scaling law for the average outbreak size

$$\langle n\rangle \sim N^{1/3}\,. \tag{5}$$

Large scale Monte Carlo simulations confirm this behavior (Fig. 1). The simulations are a straightforward realization of the infection process. When there are s susceptible individuals, with probability $1/(1+\alpha s/N)$ a recovery event occurs, and otherwise, an infection event occurs. The simulation results represent an average over a remarkably large number of independent realizations.

Statistical properties of the outbreak size are self-similar as they follow a universal, population-size independent law. Once the outbreak size distribution and the outbreak size are properly normalized by the infinite population distribution and the maximal outbreak size respectively, a universal behavior emerges: $G_n(N)/G_n(\infty) \to \mathcal{G}(n/N^{2/3})$. This universality, reminiscent of finite-size scaling in critical phenomena [13], was confirmed numerically by studying the cumulative distribution $U_n(N) = \sum_{m\geq n} G_n(N)$ (Fig. 2). This provides further verification of the scaling law (1).

The scaling laws characterizing the outbreak size hold not only at the threshold but also in a window around the threshold. Equating the average outbreak size (5) with the behavior in the endemic phase, $\langle n\rangle = (1-\alpha)^{-1}$, we find that the threshold window (i.e., the range of infection rates for which the intermediate behavior holds) diminishes with the population size as

$$|1-\alpha| \sim N^{-1/3}. \tag{6}$$

This parameter range can be sizable for moderate populations — for example, when $N = 10^3$, the threshold window is roughly $0.9 < \alpha < 1.1$ and the maximal outbreak size is smaller than the population size by a factor of 10.

The behavior of $\langle n\rangle$ near the epidemic threshold provides another manifestation of the scaling law (6). Indeed, plotting the average outbreak size versus the infection rate normalized according to (5) and (6), respectively, shows a universal behavior: $\langle n\rangle/N^{1/3} \to \mathcal{Q}\left[(1-\alpha)N^{1/3}\right]$ (Fig. 3).

The threshold window is larger than the canonical

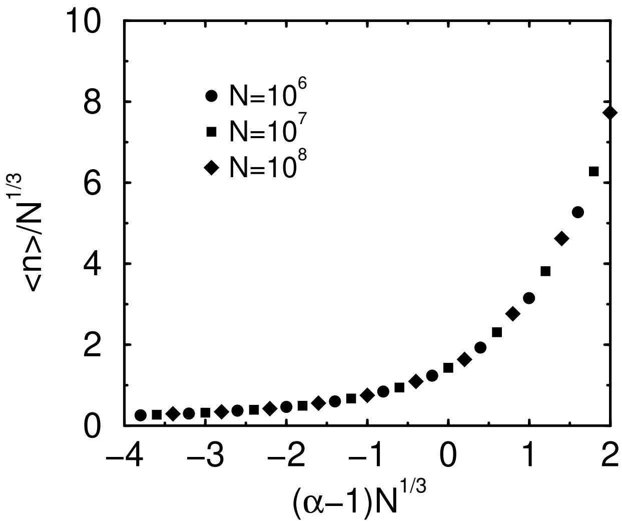

FIG. 3: The near threshold behavior. Shown is the normalized outbreak size $\langle n\rangle/N^{1/3}$ versus the normalized distance from the threshold $(\alpha-1)N^{1/3}$. The data corresponds to an average over 10^6 independent realizations.

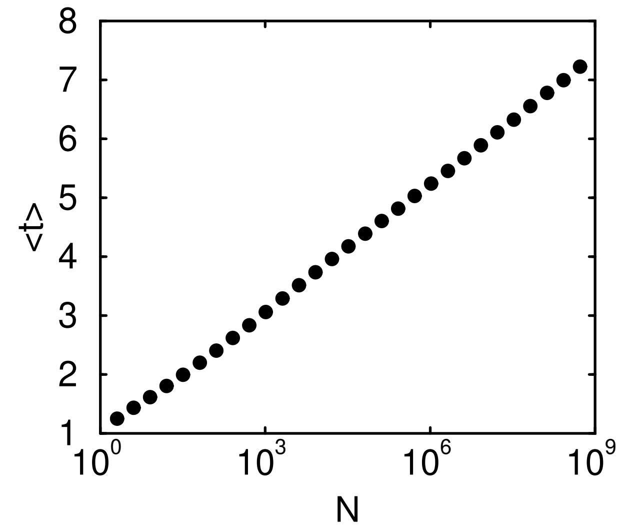

FIG. 4: The average outbreak duration at the epidemic threshold versus the population size. Simulation results, obtained from an average over 10^9 realizations are consistent with the theoretical prediction (10). A best fit to $\langle t\rangle = \beta \ln N$ yields $\beta = 0.32 \pm 0.01$.

$N^{-1/2}$ estimate arising either from the standard large-population analysis [14, 15] or from the widely-used deterministic SIR ordinary differential equations [16], describing the evolution of the average susceptible and infected populations [17]. Moreover, the related SI (sometimes also termed SIS) model, where a recovered individual immediately becomes susceptible, is characterized by the simpler behavior $n_* \sim N$ and $\langle n\rangle \sim N^{1/2}$; finite size effects are not as pronounced because there is no depletion of the susceptible reservoir.

The scaling laws for the outbreak size have direct implications concerning the dynamics and in particular, the duration of infection processes near the epidemic threshold. To obtain these scaling laws, we again consider first the infinite population limit. At the epidemic threshold, $\alpha = 1$, infection and recovery occur with equal probabilities and therefore, the average number of infected individuals is conserved, $I(t) = 1$. The probability $P_i(t)$ that there are i infected individuals at time t satisfies

$$\frac{d}{dt}P_i = (i+1)P_{i+1} + (i-1)P_{i-1} - 2iP_i \tag{7}$$

together with the initial condition $P_i(0) = \delta_{i,1}$. The distribution is geometric, $P_i(t) = t^{i-1}(1+t)^{-(i+1)}$ [17, 18] for $i \geq 1$, and $P_0(t) = t(1+t)^{-1}$ for $i = 0$. Therefore, the survival probability of the outbreak, i.e., the probability that the outbreak is still active at time t is simply

$$P(t) = (1+t)^{-1} \tag{8}$$

since $P(t) = 1 - P_0(t)$. Restricting attention to active outbreaks, the average number of infected individuals grows linearly with time $\langle i\rangle = I(t)/P(t) = 1 + t$. Consequently, the typical number of recovered individuals $r \sim \int_0^t dt'\,(1+t')$ grows quadratically with time: $r \sim t^2$.

For finite populations, the probability that the outbreak is still alive at time t decays as $P(t,N) \sim t^{-1}$ up to the maximal time scale $t \ll t_*$. The survival probability is sharply suppressed for times larger than the maximal time. The maximal duration of outbreaks is estimated by equating the time dependent outbreak size $n \sim r \sim t^2$ with the maximal outbreak size $n_* \sim N^{2/3}$. Therefore,

$$t_* \sim N^{1/3}. \tag{9}$$

The maximal duration of outbreaks greatly exceeds both the typical duration that is of the order of one and the average duration of an outbreak $\langle t\rangle$ which exhibits an interesting logarithmic growth. To derive the logarithmic law, we first note that, by definition, the average duration of an outbreak is $\langle t\rangle = \int_0^\infty dt\, t\, \frac{d}{dt}P(t,N)$. Using the infinite population result (8) and integrating up to t_* that plays the role of a cutoff, we get

$$\langle t\rangle \simeq \frac{1}{3}\ln N. \tag{10}$$

Numerical simulations confirm this behavior (Fig. 4). The probability distribution for the duration of outbreaks also follows a population-size independent law: $P(t,N)/P(t) \to \mathcal{P}\left(t/N^{1/3}\right)$ as shown in Fig. 5. However, the convergence to this law is not uniform: it is slow for short durations but fast at large durations.

In summary, we found that outbreaks in the vicinity of the epidemic threshold have a distinct size, characterized by a distinct power-law dependence of the population size. This behavior describes a range of infection processes in the vicinity of the epidemic threshold. The size of this threshold window is larger than expected from the traditional large system size analysis techniques or from the deterministic description. We conclude that statistical fluctuations and finite population effects are most pronounced and may be quite subtle near the epidemic threshold.

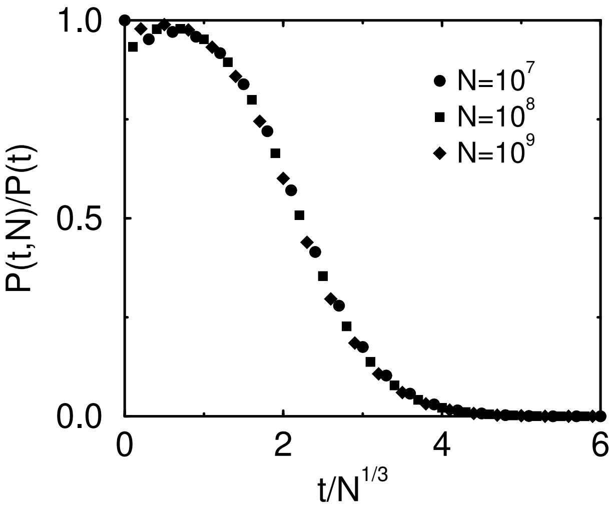

FIG. 5: The survival probability at the epidemic threshold. Shown is the normalized survival probability $P(t,N)/P(t)$ versus the normalized duration time $t/N^{1/3}$. The data corresponds to an average over 10^8 realizations.

The scaling laws have concrete implications regarding the computational complexity of near-threshold infection processes. Typically, one has to compute $P_{i,r}$, the probability that there are i infected individuals and r removed individuals from the master equations. Although there are N^2 such coupled ordinary differential equations, the scaling laws $i \sim N^{1/3}$ and $r \sim N^{2/3}$ imply that the number of relevant equations is much smaller and scales only linearly with the population size.

Several questions arise, e.g., what is the shape of the scaling functions $\mathcal{G}\left(n/N^{2/3}\right)$ and $\mathcal{P}\left(t/N^{1/3}\right)$ characterizing the size and duration of outbreaks near the epidemic threshold? Numerically, we observe the both distributions have stretched exponential tails, and that $\mathcal{P}(w) \sim \exp(-w^\delta)$ with $\delta \approx 2$. Analytical determination of these functions is very challenging as it requires treatment of the full master equations describing the stochastic infection process [1], that is, the distribution $P_{i,r}(t,N)$ is needed [17].

Further related problems include the corresponding near-threshold scaling laws for spatial epidemic models, where the geometry and the spatial structure of the infected domain play a role [19–22], and infection processes on networks [23, 24]. We anticipate that finite size effects should be relevant in these systems as well.

Acknowledgments

We are grateful to Aric Hagberg for initial collaboration on this work. We also thank Gary Doolen, Hans Frauenfelder, and Sergei Rudchenko for careful reading of the manuscript. This research was supported in part by DOE(W-7405-ENG-36).

[1] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases* (Charles Griffin, London, 1975).
[2] R. Anderson and R. May, *Infectious Diseases: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
[3] H. W. Hethcote, SIAM Rev. **42**, 599 (2000).
[4] G. H. Weiss and M. Dishon, Math. Biosci. **11**, 261 (1971).
[5] H. E. Daniels, Biometrika **59**, 211 (1972).
[6] C. Lefevre and C. Picard, Adv. Appl. Prob. **22**, 25 (1990).
[7] F. Ball and D. Clancy, Adv. Appl. Prob. **25**, 721 (1993).
[8] H. Andersson and B. Djehiche, J. Appl. Prob. **35**, 662 (1998).
[9] J. Grasman, Math. Biosci. **152**, 13 (1998).
[10] J. Bull and D. Dykhuizen, Nature **426**, 609 (2003); R. Antia, R. R. Regoes, J. C. Koella, and C. T. Bergstrom, Nature **426**, 658 (2003).
[11] The fraction r satisfies $r + \exp(-\alpha r) = 1$.
[12] T. E. Harris, *The Theory of Branching Processes* (Dover, New York, 1989).
[13] V. Privman and M. E. Fisher, Phys. Rev. B **30**, 322 (1984).
[14] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North Holland, Amsterdam, 2003).
[15] C. R. Doering, K. V. Sargysyan, and L. M. Sander, *q-bio*/0401016.
[16] J. D. Murray, *Mathematical Biology* (Springer-Verlag, New York, 1989).
[17] E. Ben-Naim and P. L. Krapivsky, in preparation.
[18] E. Ben-Naim and P. L. Krapivsky, Phys. Rev. E **68**, 031104 (2003).
[19] D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor & Francis, Briston PA, 1994).
[20] P. Grassberger, Math. Biosci. **63**, 157 (1983).
[21] H. Andersson and B. Djehiche, J. Appl. Prob. **34**, 698 (1998).
[22] C. P. Warren, L. M. Sander, and I. M. Sokolov, Phys. Rev. E **66**, 056105 (2002).
[23] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[24] M. E. J. Newman, Phys. Rev. E **66**, 016128 (2002).