Pandemic!

(Sethna, "Entropy, Order Parameters, and Complexity", ex. 12.33)

In [ ]:

```
# Sometimes gives interactive new windows
# Must show() after plot, figure() before new plot
# %matplotlib

# Adds static figures to notebook: good for printing
%matplotlib inline

# Interactive windows inside notebook! Must include plt.figure()
# %matplotlib notebook

# Better than from numpy import *, but need np.sin(), np.array(),
import numpy as np
import matplotlib.pyplot as plt
```

Perhaps the most substantive contribution to public health provided by physics is the application of statistical mechanics ideas to model disease propagation. In this exercise, we shall introduce a few categories of epidemiological models, discuss how they can inspire and inform public health strategies (once adapted to real-world data), and then study one model as a continuous phase transition. You should leave this exercise empowered to think about the public health responses and modeling of potential pandemics---Ebola, SARS, and now COVID-19. Perhaps a few of us will contribute to the field.

Pandemics can undergo a phase transition. For diseases like measles, a single contagious child in an environment where nobody is immune will infect between twelve and eighteen people before recovering, depending on details like population density. For influenza, this number is around two to three. We define the basic reproduction number R_0 to be this average number of people infected by a contagious person before recovering, in a fully susceptible community. R_0 thus is 12-18 for measles, 2-3 for influenza. For a new pathogen, where nobody is immune, $R_0 < 1$ will mean that an outbreak will eventually die out, and $R_0 > 1$ means that a large initial outbreak will spread globally until reaching a significant fraction of the entire population. Much effort is spent during a pandemic to lower R_0 into the safe range.

This transition is a continuous phase transition, with fluctuations on all scales near the critical threshold $R_0 = 1$. In this exercise, you will briefly consider three types of epidemic models (compartmental models, network models, and lattice models), compare different social interventions designed to lower R_0, and explore the fluctuations and critical behavior very close to threshold.

Compartmental models use coupled differential equations to model the disease spread between different {\em compartments} of the population. The classic SIR model (see Exercise 6.25) involves three coupled compartments,

$$\frac{dS}{dt} = -\beta I S, \qquad \frac{dI}{dt} = \beta I S - \gamma I, \qquad \frac{dR}{dt} = \gamma I,$$

where $S(t)$, $I(t)$, and $R(t)$ are the proportions of the population that are susceptible, infected, and recovered. The parameter β measures the rate of infection spreading contact between people and γ is the rate at which people recover.

Network models treat people as nodes, connected to their contacts with edges. They assume a transmissibility T, the average probability that a victim will infect each of their contacts. For low T the epidemics die out; there will be a critical T_c above which a large outbreak will continue to grow exponentially. There are a variety of networks studied: fully connected networks (where SIR models become valid), loopless branching tree networks where everyone has k neighbors, real-world networks gleaned from data on households and school attendance, and scale-free networks with a power-law distribution $p(k) \propto k^{-\alpha}$ for the probability that a person has k contacts (has degree k). (Scale-free networks have been found to approximate the pattern of interactions between proteins in cells and nodes on the Internet, and serve as our model for populations with wide variation in the number of social contacts with potential for disease transmission.)

Lattice models---networks in two dimensions where only near neighbors are contacts---are sometimes used in agricultural settings, where the plant victims are immobile and the disease is spread only by direct proximity.

(a) *In the SIR model, if a person is first infected at t_0, what is the probability they are still infected, and contagious, at time t, given the recovery rate γ? If they infect people at a rate βS, how many will they infect before they recover, in a fully susceptible large*

Your answer here.

Network models usually ignore the long-range correlations between nodes: except for real-world networks, the contacts are usually picked at random so there are very few short loops. In that limit, Meyers et al. [133] express R_0 in terms of the moments $\langle k^n \rangle = \sum k^n p(k)$ of the degree distribution, which they solve for using generating functions (see Exercise 2.23):

$$R_0 = T\left(\frac{\langle k^2 \rangle}{\langle k \rangle} - 1\right).$$

People like nurses and extroverts with a lot of contacts can act as super-spreaders, infecting large numbers of colleagues. Scale-free networks explore what happens with a range of contacts: the smaller the exponent α, the larger the range of variation.

(b) *What is the critical transmissibility T_c predicted by the network model in the equation above? Show that, for a scale-free network with $\alpha \leq 3$ the critical transmissibility $T_c = 0$; no matter how unlikely a contact will cause disease spread, there are rare individuals with so many contacts that they (on average) will cause exponential growth of the pathogen. If our population had $\alpha = 3$, what percentage of the people would we need to vaccinate to immunize everyone with more than 100 contacts? What would the resulting T_c, the maximum safe transmissibility, be? (If you find that the first percentage is small, use that fact to simplify your calculation of T*

Your answer here.

An important limitation of these network results is that they assume the population is structureless: apart from the degree distribution, the network is completely random. This is not the case in a 2D square lattice, for example. It has degree distribution $p_k = \delta_{k,4}$, but connections between nodes are defined by the lattice, and not randomly assigned. As you might expect, disease spread is closely related to percolation. In the mean-field theory, percolation predicts that the epidemic size distribution exponent is $\tau = 3/2 = 1.5$; you will explore this in parts (e) and (f). In 2D, the lattice structure changes the universality class, the epidemic sizes are given by the cluster-size distribution exponent $\tau = 187/91 \approx 2.055$.

Besides exhibiting different power-law scaling, the value of the critical transmissibility can be quite different in structured populations.

(c) *What is T_c for a branching tree with $k = 4$ branches at each node (so $p(k) = \delta_{k,4}$)? (Hint: You can derive this directly, or use your answer to the first question in part (b).) Compare that to the critical transmissibility for a 2D square lattice, $T_c = 0.5384$. Which is more resistant to disease spread?*

Your answer here.

One might imagine that a lattice model would mimic the effect of travel restrictions to prevent disease spread. Travel restrictions reduce the contact numbers, but do not change the qualitative behavior. This is due to the small world phenomenon: a surprisingly small number of long-range contacts can change the qualitative behavior of a network (see Exercise 1.7). Only a few long-distance travelers are needed to make our world well connected.

Finally, let us numerically explore the fluctuations and scaling behavior exhibited by epidemics at their critical points. We shall assume (correctly) that our population is well connected. We shall also assume that our population does not have system-scale heterogeneities: we ignore cities, subpopulations of vulnerable and crowded people, and events like the Olympics. Given these assumptions, one can argue that the qualitative behavior near enough to the critical point $R_0 = 1$ is universal, and controlled not by the details of the network or SIR model but only by the distance $1 - R_0$ to the critical point.

Let us organize our victims in "generations" of infected people, with I_{n+1} the number of victims infected by the I_n people in generation n; we shall view the generation as roughly corresponding to the time evolution of the pandemic. The mean $\langle I_{n+1} \rangle = R_0\ I_n$, but it will fluctuate about that value with a Poisson distribution, so I_{n+1} is a random integer chosen from a Poisson distribution with mean $R_0\ I_n$.

(d) *Write a routine pandemicInstance, that returns the evolution vector* $[I_0, I_1 \ldots I_n \ldots]$ *and the total size* $S = \sum_n I_n$*. Iterate your routine with with* $R_0 = 0.9999$ *and* $I_0 = 1$ *in a loop until you find an epidemic with size* $S \geq 10^5$*. Plot the trajectory of this epidemic,* I_n *vs.* n*. Does your epidemic nearly halt during the time interval? Do the pieces of the epidemic before and after this near halt appear statistically similar to the entire epidemic?*

In [ ]:

```
def pandemicInstance(R0=1., i0 = 1):
    I = i0
    Itraj = [I]
    size = i0
    while I!=0:
        I = np.random.poisson(R0*I)
        size += ...
        Itraj.append(...)
    return size, Itraj

size, traj = pandemicInstance(...)
while size < 1000000:
    size, traj = pandemicInstance(0.9999)
plt.plot(I)
plt.title("size is "+ str(size))
plt.xlabel("Time in shells")
plt.ylabel("Infected in this shell")
plt.show()
```

Your answer here.

One might presume that these large fluctuations could pose a real challenge to guessing whether social policies designed to suppress a growing pandemic are working. We must note, however, that the fluctuations are important only near $R_0 = 1$, or when the infected population becomes small.

At $R_0 = 1$, the size of the epidemic S has a power-law probability density $P(S) \propto S^{-\tau}$ for large avalanches S.

(e) *Write a routine pandemicEnsemble that does not store the trajectory, but instead runs* N *epidemics at a given value of* R_0*, and returns a list of their sizes. Plot a histogram of the sizes of* 10^4 *epidemics with* $R_0 = 0.99$*, with, say, 100 bins.*

In [ ]:

```
def pandemicEnsemble(N, R0=1., i0 = 1):
    sizes = []
    for n in range(N):
        I = ...
        size = i0
        while I!=0:
            I = np.random.poisson(R0*I)
            size += ...
        sizes += [...]
    return sizes

sizes = pandemicEnsemble(10**5, R0=...)
plt.hist(sizes, bins=100)
plt.title("Epidemic size histogram")
plt.xlabel("Size")
plt.ylabel("Counts")
plt.show()
```

Regular histograms here are not useful; our distribution has a long but important tail of large events. Most epidemics subside quickly at this value of R_0, but a few last for hundreds of generations and infect tens of thousands of people. We need to convert to logarithmic binning.

(f) *Change the bins used in your histogram to increase logarithmically, and be sure to normalize so that the counts are divided by the "bin width" (the number of integers in that bin) and the number of epidemics being counted. Present the distribution of sizes for 10^4 epidemics at $R_0 = 0.99$ on log--log plots. On the same plot, show the power-law prediction $\tau = 3/2$ at the critical point.*

In [ ]:

```
def intlogspace(start, stop, num=50, endpoint=True, base=10.0):
    realBins = np.logspace(start, stop, num, endpoint, base)
    bins = np.unique(realBins.astype(int))
    return bins
```

In [ ]:

```
sizes = pandemicEnsemble(10000, R0=0.99)

def logbinnedHist(vals, nBins=30, exponent=3/2):
    maxval = max(vals)
    bins = intlogspace(0,int(np.log10(maxval))+1,nBins)
    widths = (bins[1:]-bins[:-1])
    counts, edges = np.histogram(vals, bins=bins)
    hist_norm = counts/(widths*len(vals))
    plt.plot(bins, (exponent-1) * bins**(-exponent), color='green
    plt.bar(bins[:-1], hist_norm, widths)
    plt.xscale('log')
    plt.yscale('log')
    plt.legend()

logbinnedHist(sizes)
plt.title("Epidemic size probability distribution")
plt.xlabel("Size S")
plt.ylabel("P(S)")
plt.show()
```

In Exercise 12.28 we derived the universal scaling form for the avalanche size distribution in the random-field Ising model. This calculation also applies to our pandemic model. It predicts that the probability $P(S)$ of an epidemic of size S for small distance $r = (1 - R0)$ below the critical point is given by

$$P(S) = C S^{-3/2} e^{-Sr^2/2},$$

where the nonuniversal constant C is around 0.4 to 0.5. Note that this is the predicted power law $\tau = 3/2$, but cut off above a typical size that grows quadratically in $1/r$.

(g) *Multiply your data and the scaling prediction by $S^{3/2}$ to make them near constant for small sizes (to make it easier to study the cutoff). Plot both on a log-log plot. Does the universal scaling function describe your simulated epidemic ensemble?*

In [ ]:

```
def logbinnedHistScaling(vals,R0=0.9,C=0.45):
    num = len(vals)
    maxval = max(vals)
    bins = intlogspace(0,int(np.log10(maxval))+1,30)
    widths = (bins[1:]-bins[:-1])
    centers = (bins[1:]+bins[:-1])/2
    counts, edges = np.histogram(vals, bins=bins)
    counts_rescaled = ...*counts/(num*widths)
    plt.plot(centers, C*np.exp(-centers*(1-R0)**2/2), color='gree
    plt.plot(centers, ..., label = "Rescaled data")
    minY = max(min(counts_rescaled),10**(-4))
    plt.ylim(minY,10**0)
    plt.title("Epidemic size distribution scaling plot")
    plt.xlabel("Size S")
    plt.ylabel(r"S**(3/2) P(S)")
    plt.xscale('log')
    plt.yscale('log')

sizes = pandemicEnsemble(100000, R0=0.9)
logbinnedHistScaling(sizes,R0=0.9)
```

The tools we learn in statistical mechanics---generating functions, universality, power laws, and scaling functions---make tangible predictions for practical models of disease propagation. They work best in the region of greatest societal importance $R_0 \approx 1$, where costly efforts to contain the pandemic are minimized while avoiding uncontrolled growth.