Modeling Epidemics and Herd Immunity Computationally

Agar-Johnson, M.

Dec. 6, 2018

Introduction

An increasing percentage of the general population is choosing to forego childhood vaccinations, particularly the measles, mumps, rubella (MMR) vaccine. This may be attributable to a number of factors, including fear of vaccine ingredients/adjuvants like thimerosal, falsified medical evidence citing links between vaccination and childhood developmental disorders, and misinformation regarding the necessity and function of vaccines in the population at large[1]. Due to the resurgence of anti-vaccine sentiments, there have been a number of measles outbreaks reported in countries where the disease was previously thought to have been completely eradicated[2]. Individuals with certain immune system conditions, infants, and other members of the population are unable to be given certain vaccinations, but are protected by a concept known as herd immunity. This is the property that if enough individuals in a population are immune to a pathogen, the limited availability of susceptible hosts prevents the spread of the disease throughout the population at large. To assess and visualize the risks of the proliferation of the anti-vaccine movement, an algorithm was developed to simulate and visualize the spread of infectious disease throughout a population, the impact of vaccinations and herd immunity on the spread of disease, and the threshold for the breakdown of herd immunity and realization of an epidemic. These simulations rely on community network theory for the simulation of intra-community interactions, as well as previously established epidemiological research and models to simulate infection dynamics. The result is a stochastic, interactive algorithm that can simulate the spread of a variety of illness types through different populations, and demonstrate the need for high vaccination rates to prevent the spread of highly infectious diseases.

Methods

Community Network Dynamics

Epidemics are spread by contact between individuals in a community network. When attempting to perform disease progression simulations, it is important to develop a stochastic model for community networks that can accurately represent the spread of infection through a population. The most established approach is to generate a probability distribution for the degree of each node in a network. For each node in the network, a degree *k* is sampled from the distribution and the node is connected to *k* random nodes in the network. It has previously been shown that a random graph follows a Poisson degree distribution[3]. The Poisson distribution, however, does not account for key dynamics of disease progression. There are often specific individuals, known as 'super-spreaders', who have contact with many other individuals and are responsible for a significant share of disease spread throughout a population. To accurately model this phenomenon, a scale-free "power law" degree distribution was used. Based on prior epidemiological data, the degree distribution was modeled as:

$$p(k) = Ck^{-\alpha} \exp\left(\frac{-k}{\kappa}\right)$$

With $\alpha = 2, \kappa = 94.2$, and , $C = \frac{N}{10}$, where *N* is the total number of individuals in the population [3].

In order to sample from this distribution, the equation above was re-written as the following:

$$\alpha\kappa(\log(k)) + k - \kappa\left(\log\left(\frac{C}{p(k)}\right)\right) = 0$$

Where $p(k) \sim U(0,1)$. Since the degree *k* always takes an integer value, an approximation of *k* from the above function using Newton's method was sufficient, and was deemed converged when the delta between iterations was ≤ 0.0001. This was more than sufficient and was not costly in terms of computational speed.

To generate the community network, the total number of individuals *N* is taken as an input parameter into the algorithm. The network is initialized as an array objects of length *N* consisting of pointers to Node objects. A Node object is a data structure containing an integer ID, a vulnerability coefficient (discussed in a later section), an infection status indicated by a string, and an array of pointers to other nodes in the network. The algorithm iterates through all *N* Nodes in the network, and for each node N_i, samples a degree *k* from the power-law distribution described above. It then connects N_i to *k* random nodes in the network, and stores their locations as an array of pointers in N_i. Each connection is checked to make sure that there are no auto-connections or repeated connections in the network[4].

Modelling Disease Transmission

For simplicity, the time-course of an epidemic was not considered, and our model is based on the SIR epidemiological model. The population susceptible to infection is marked *S*, infected individuals are marked *I,* and recovered individuals are marked *R*. From some initial population of infected individuals, susceptible individuals are infected with some transmission probability, infect their neighbors, and are then marked as recovered. This model is appropriate and well-established for "simple" disease transmission, that is, transmission through a relatively homogeneous population without secondary transmission through animal vectors or free-living infective pathogens[5].

One of the aims of this project is educational in nature, to visualize the potential consequences of a widespread anti-vaccine movement on a large population. For these purposes, some changes were made to the simple SIR model. We introduced 2 new states – "V" and "D", with "V" represented those who were vaccinated, to distinguish from those who contracted the disease and recovered, and "D" to signify dead, with a mortality rate worked into each pathogen and being drawn from real-life data.

Showing potential deaths from widespread epidemics was thought to carry a more realistic, but also severe and noteworthy, set of consequences from an under-vaccinated population.

The model takes as an input a ".PATHOGEN" file, which is a simple plaintext file containing 3 pieces of information: the plaintext name of the pathogen, the "base reproductive ratio" R_o, and a mortality rate as a decimal number between 0 and 1, inclusive. To run the program, the user inputs a filepath to a .PATHOGEN file or the string "CUSTOM", which then prompts user inputs to create a new .PATHOGEN file and automatically loads it into the algorithm.

The actual infection proceeds in the following way. At each iteration through the network, all infected nodes infect their susceptible neighbors with a probability proportional to the disease transmissibility *T.* The transmissibility is a function of both R_o of the pathogen, and the mean degree of the community network, and is derived from the following equation:

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

Where $\langle k \rangle$ and $\langle k^2 \rangle$ are the mean degree and mean squared-degree of the network respectively. The final probability of infection for a susceptible neighbor node N_i is given by:

$$p(I_{N_i}) = T \times v_{N_i}$$

Where v_{N_i} is the vulnerability coefficient of N_i . The vulnerability coefficient of each node is a decimal number sampled from the gaussian distribution $G(1, 0.556)$. This distribution was determined from statistics about immunocompromised and disease-susceptible individuals in the U.S. general population[6]. After infected its neighbors, each infected node was set to either "R" for recovered or "D" for dead according to $p(D) = m_P \times v_{N_i}$, where m_P is the mortality rate of the pathogen. The algorithm continued to iterate over the nodes in the network until no infected nodes remained.

Vaccination and Network Frailty

Vaccination of a population was accomplished simply by taking a user input for the percentage of the population to be vaccinated. These nodes were assigned a status of "V" and were treated as immune from infection. In the context of herd immunity and disease networks, there is an important concept of network "frailty". Following an epidemic, the individuals susceptible to the disease are less connected, due to either deaths or a higher percentage of individuals immune following infection. The 'frailty' of a network can be interpreted as its susceptibility to future epidemic and infection. "Herd immunity", the property by which vaccinated individuals protect those who must remain unvaccinated from disease exposure, can also be thought of in terms of network frailty. In "power-law" networks, where infection and spread of disease may be centered around a specific group of individuals, the vaccination of random individuals will have a smaller impact on network frailty than the passing of an epidemic, as an epidemic is likely to target the "super-spreaders" and those around them, making them immune to future infection. Because vaccination does not take these network dynamics into account, it is important to vaccinate a very large percentage of the population in order to reduce the network frailty sufficiently to protect the total population from epidemic.

Mathematically, we model network frailty using two parameters, frailty and interference. Frailty "quantifies the extent to which the high-degree individuals are preferentially infected during an epidemic"[7] and is described by the following equation:

$$\phi = \frac{\langle k \rangle - \langle k_r \rangle}{\langle k \rangle}$$

Where $\langle k \rangle$ again represents the mean degree of the network, and $\langle k_r \rangle$ represents the mean original degree of the residual network. The residual network is the network of connected, susceptible nodes

following the infection. The mean original degree of the residual network is the mean degree of all nodes in the residual network prior to the start of the infection.

The interference is "the extent to which the epidemic has cut-off connectivity among the remaining susceptible population"[7], and is described by the following equation:

$$\theta = \frac{\langle k_r \rangle - \langle k_r \rangle_r}{\langle k \rangle}$$

Where $\langle k_r \rangle$ again represents the original mean degree of the nodes in the residual network, and $\langle k_r \rangle_r$ represents the residual degree of the nodes in the residual network following the outbreak. To make the difference between $\langle k_r \rangle$ and $\langle k_r \rangle_r$ more clear, $\langle k_r \rangle$ is the average number of connections every residual node has, whereas $\langle k_r \rangle_r$ is the average number of connections every residual node has to another residual node.

The algorithm does not perform any further actions with these two network frailty parameters, but they are calculated and provided in the epidemic summary as an indicator of the effects of vaccination and herd immunity on overall population susceptibility to disease outbreaks and epidemics.

Results

All results were obtained by running the simulation on a sample population of 50,000 with measles, starting with 5 infected individuals (patients zero). First, to visualize the breakdown of herd immunity, the results of the epidemics are visualized below (Figure 1.). Red squares represent dead individuals, blue squares are those who were infected but recovered, and white squares are individuals who are either vaccinated or have not contracted the disease.

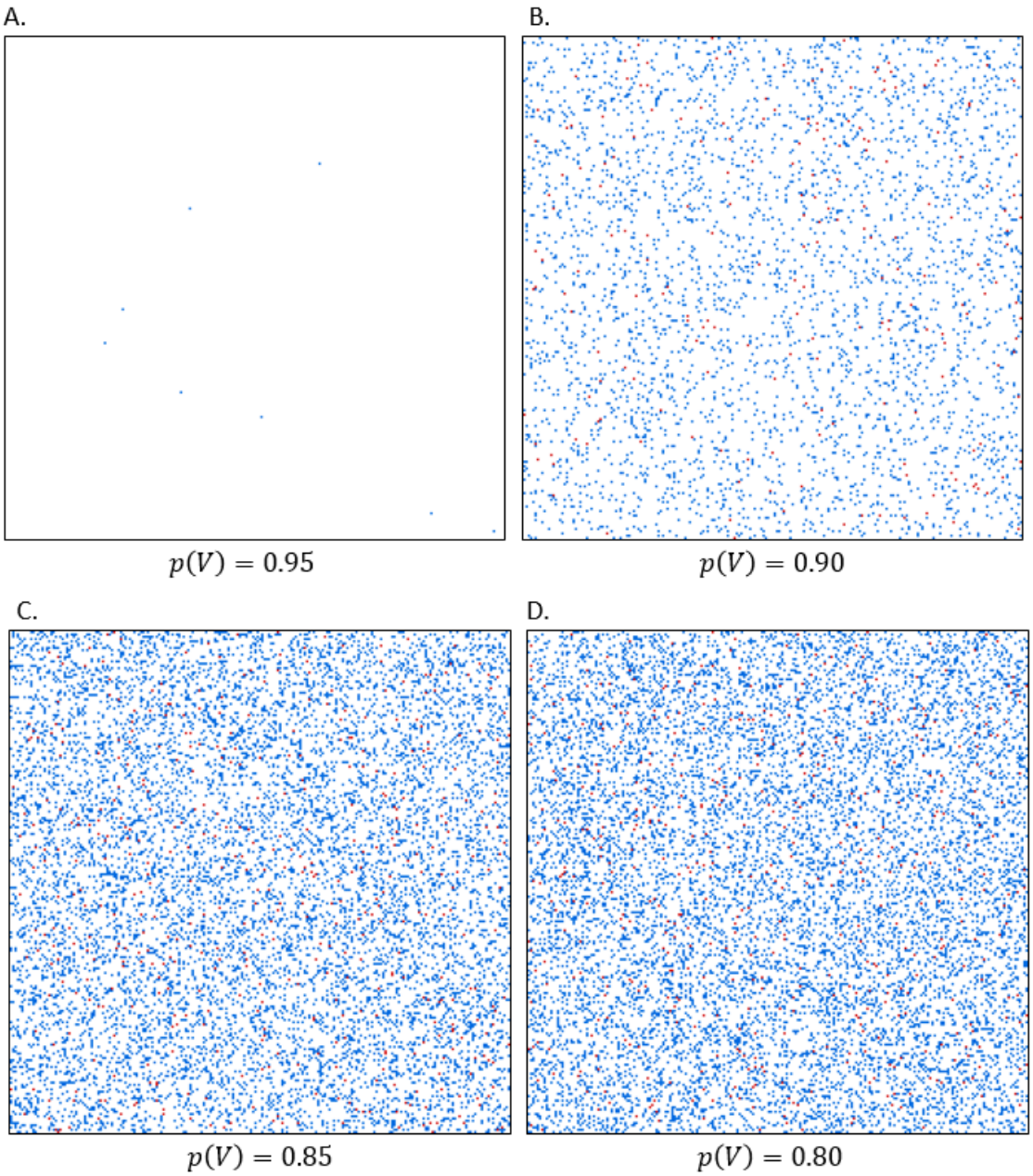

Figure 1. Final population profile post-measles infection with vaccination rates of 0.95 (A), .90 (B), .85 (C), and .80 (D). Drawn from an initial population of 50,000; blue squares represented infected but recovered individuals, red squares dead individuals, and white squares noninfected. Measles was initialized with an R_o of 13.5 and a mortality rate of 0.03.

It is quite clear from these visualizations alone that the properties of the infection change dramatically between .95 and .80, and even from .95 to .90. With a 95% vaccination rate, the infection spread to only 3 individuals and killed 0 people. At 90%, this jumped to 202 deaths and 3301 additional infections. By

75%, there were 700 deaths and 11643 additional infections. These results are interpreted as the breakdown of herd immunity, which is consistent with the ~90-95% requirement for herd immunity determined by the Oxford Pediatric Group[8].

To have a clearer quantitative vision of these results, the residual fraction, defined as the percentage of susceptible, non-vaccinated individuals who were not infected, was calculated.

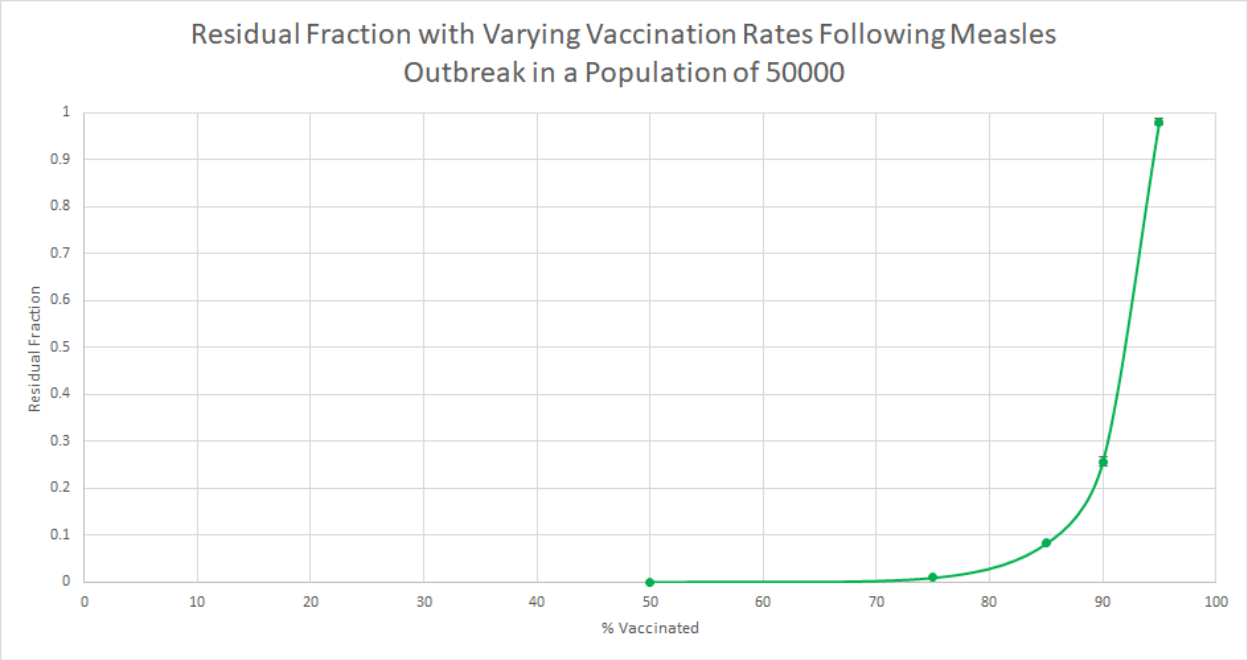

Figure 2. The mean residual fraction at different vaccination rates (n = 3 for each datapoint) was simulated. The residual fraction represents the percentage of non-vaccinated, susceptible individuals that did not contract the illness.

The residual fraction displays non-linear behavior and changes dramatically between 80-95% vaccination rates. If herd immunity were not a factor, we would expect linear behavior. Not only do the simulations replicate herd immunity, they also show that a very high vaccination rate is needed in order to achieve these protective effects. Even with 75% immunization, almost all susceptible individuals within the community network eventually developed the disease.

We can further observe these effects by looking at the changes in network frailty as a function of vaccination rates.

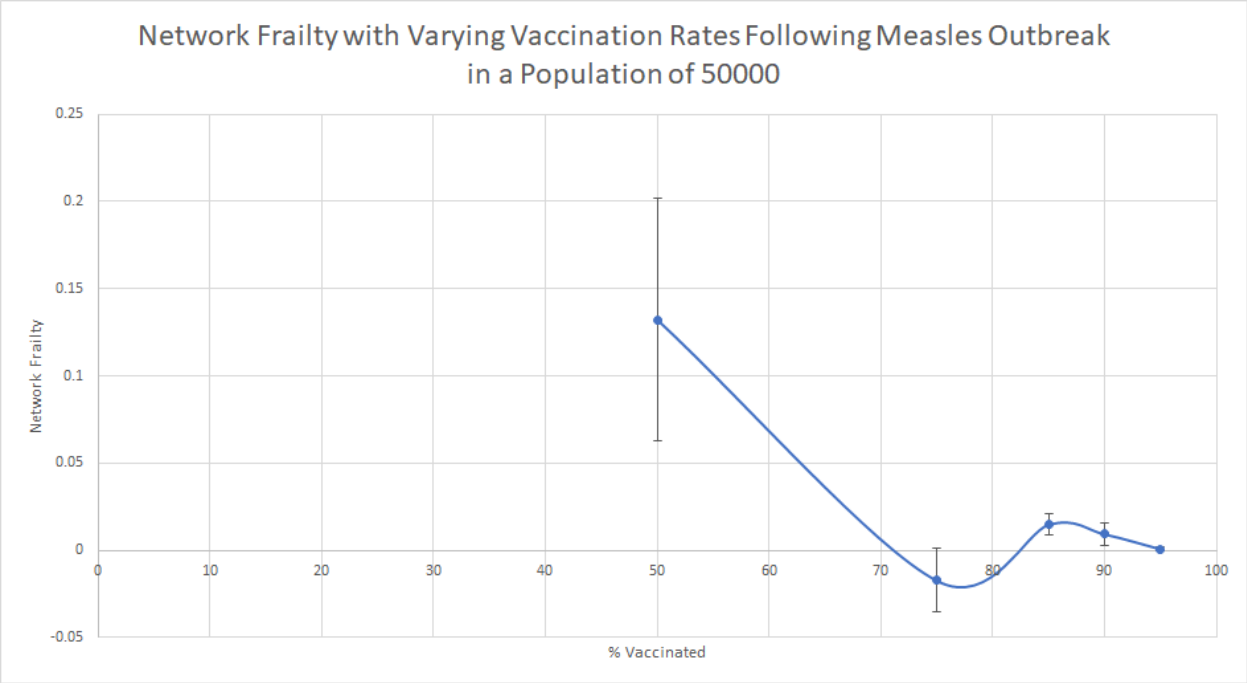

Figure 3. The mean network frailty at different vaccination rates (n = 3 for each datapoint) was simulated. The network frailty is a visualization of the extent to which high-degree "super-spreaders" were targeted during the epidemic.

There is a large degree of variation within these results, but an overall trend of increasing frailty as vaccination rates decrease was observed. These data provide further evidence that an underlying property of the network itself is responsible for the increased ability of the pathogen to proliferate, and show how even with high rates of vaccination, vulnerabilities in total population immunity can arise.

Conclusion

While many in the scientific community and the general population are aware of the dangers of anti-vaccine sentiments and lower overall vaccination rates, it may not be immediately obvious how small a percentage needs to abstain from vaccination for the threat of large-scale epidemics to become a reality. There will always be individuals who are unable to be vaccinated, and those individuals are often the ones most at risk of serious complications from infectious disease. These simulations supported the notions of herd immunity, but also support its fragility. It is the intention of this study to provide a clear picture of how vital vaccine compliance is, especially in a world trending toward increased connectivity and population density.

1. Hussain, A., et al., *The Anti-vaccination Movement: A Regression in Modern Medicine.* Cureus, 2018. **10**(7): p. e2919.
2. (WHO), W.H.O., *Europe observes a 4-fold increase in measles cases in 2017 compared to previous year.* WHO Press Release, 2018.
3. Meyers, L.A., et al., *Network theory and SARS: predicting outbreak diversity.* J Theor Biol, 2005. **232**(1): p. 71-81.
4. Maslov, S., K. Sneppen, and A. Zaliznyak, *Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet.* Physica A: Statistical Mechanics and its Applications, 2004. **333**: p. 529-540.
5. McCallum, H., N. Barlow, and J. Hone, *How should pathogen transmission be modelled?* Trends Ecol Evol, 2001. **16**(6): p. 295-300.
6. Kemper, A.R., M.M. Davis, and G.L. Freed, *Expected adverse events in a mass smallpox vaccination campaign.* Eff Clin Pract, 2002. **5**(2): p. 84-90.
7. Ferrari, M.J., et al., *Network frailty and the geometry of herd immunity.* Proc Biol Sci, 2006. **273**(1602): p. 2743-8.
8. Sadarangani, M., *Herd Immunity: How does it work?* Oxford Vaccine Group, Dept. of Pediatrics, 2016. https://www.ovg.ox.ac.uk/news/herd-immunity-how-does-it-work