An analysis of the Miller model in predicting influenza spread

BRODY H. REID

Department of Mathematics and Statistics, University of Victoria.

The Miller model combines contact networks, probability theory, and infectious-disease dynamics in order to mathematically describe how diseases can spread amongst a population. Following Miller's work closely, we construct a theory of how a population can be modelled by a network—each node in the network represents an individual, and each edge indicates social contact through which the infection can spread. In order to test whether the Miller model is adequate in describing the spread of influenza, we derive the Miller model using the susceptible-infected-recovered (SIR) model. We assume a Poisson degree distribution for the network and fit the model to influenza spread data. The average time for a new individual to become infected was estimated to be 771 days. And the average degree of the network was estimated to be 31. The method of analysis and the model assumptions need to be reconsidered as the results are unrealistic.

Keywords: Networks, Miller model, SIR model, Influenza Spread

1. Introduction

In 2011, Joel C. Miller pioneered a new area of mathematical biology in disease spread on networks [1]. Here, we compare the predictions of Miller's model to data acquired from the Centers for Disease Control and Prevention [2]. The data we examine is the number of hospital and clinic visits in the United States due to influenza-like illnesses from September 2017–2018. This data is appropriately representative of influenza spread, and, therefore, can be compared to the theoretical model derived by Miller.

We follow Miller's work closely and begin with a contact network representing the population. We then assume a degree distribution for the network and define the probability generating function of the distribution—which comes in useful in infectious disease dynamics.

2. Contact networks

Using a contact network, we are able to describe social interaction in the population. This social contact is vital for disease spread dynamics—especially in the case of influenza—since it is what facilitates the transfer of infection from person to person.

Let nodes in the network represent individuals in the population, and let edges represent social contact between individuals. Each node will have degree k, which will vary with different degree distributions. The average degree of the network will be given by the expected value of the distribution, $\langle k \rangle$. Let $P(k)$ be the probability that a randomly chosen node has degree k. 5 We define $\psi(x)$ to be the probability generating function of x:

$$\psi(x) = \sum_{k=0}^{\infty} P(k)x^k. \tag{2.1}$$

The average degree of the network is given by the sum of each degree multiplied by its likelihood of existing; i.e.,

$$\langle k \rangle = \sum_{k=1}^{\infty} kP(k). \tag{2.2}$$

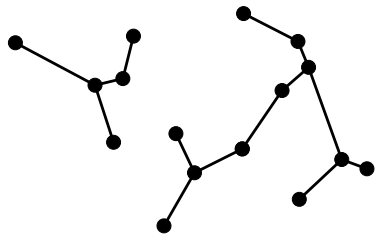

FIG. 1: Example of a network with population $N = 16$ and degree distribution $P(k)$, as given in Ex. 2.1.

It is useful to note that (2.2) is identical to $\psi'(x)|_{x=1}$, where the prime indicates an ordinary derivative. So, the expected value of a degree distribution is equal to the derivative of its generating function evaluated at $x = 1$: $\langle k \rangle = \psi'(1)$.

EXAMPLE 2.1 Say we have a network with a population of 16 and the following degree distribution,

$$P(k) = \begin{cases} 1/2 & \text{if } k = 1, \\ 1/4 & \text{if } k = 2, \\ 1/4 & \text{if } k = 3. \end{cases}$$

An example of a network with this distribution is given in Fig. 1. The probability generating function of this distribution is $\psi(x) = (x^3 + x^2 + 2x)/4$, and the average degree of the network is $\langle k \rangle = 1.75$.

3. The SIR model

We assign each node in the network one of three classes: susceptible, infected, or recovered. If a node is classified as a susceptible, then it does not have the infection, but is able to contract it from an infected. If a node is classified as an infected, then it can transmit the infection to susceptible nodes at a rate β, or it can recover from the infection and become a recovered node at a rate γ. In a constant population of N individuals, the fraction of susceptible individuals is given by $S(t)$, the fraction of infected individuals is given by $I(t)$, and the fraction of recovered individuals by $R(t)$. The dynamics of these classes of nodes are given by [3]

$$\begin{aligned} \dot{S} &= -\beta SI, \\ \dot{I} &= \beta SI - \gamma I, \\ \dot{R} &= \gamma I, \end{aligned} \tag{3.1}$$

where a dot represents the time derivative, β is the transmission rate of the infection, and γ is the recovery rate of infected individuals. For influenza, β is estimated to be 0.06 new infections per day, and γ is estimated to be 0.2 recoveries per day [4]. A flow chart of system (3.1) is given in Fig. 2.

4. The Miller model

Miller defines two important quantities, each with quite subtle definitions: $\theta(t)$ is the probability that a randomly chosen node has *not* transmitted the infection at time t; and $\varphi(t)$ is the probability that a randomly chosen base node of an edge is infected, but the edge has not transmitted the infection at time t [1]. So the rate of change of the probability that a random edge has not transmitted the infection is equal to the rate of

FIG. 2: SIR flow chart where β is the transmission rate of the infection and γ is the recovery rate. We make the assumption that once an individual recovers from the infection, they remain classified as recovered and do not become susceptible once again.

change at which the infection crosses edges:

$$\dot{\theta} = -\beta\varphi. \tag{4.1}$$

To derive an equation for $\dot{\varphi}$, consider what happens when an infected node transmits the infection to a susceptible node. The rate at which neighbours of an infected node become infected is exactly the rate at which the neighbours stop being susceptible. So if $h(t)$ is the probability that a neighbour of an infected node is susceptible, then $\dot{\varphi} = -(\beta+\gamma)\varphi - dh/dt$.

In order to determine $h(t)$, consider the probability that a neighbour of an infected node has degree k. This is just the ratio of the degree of the node to the average degree of the network multiplied by the probability of the node having that degree [1]: $kP(k)/\langle k\rangle$. And since the neighbouring node cannot infect itself, we multiply by θ^{k-1} and sum over all possible degrees. So

$$h(t) = \frac{1}{\langle k\rangle}\sum_{k=0}^{\infty} kP(k)\theta^{k-1} = \frac{1}{\langle k\rangle}\,\psi'(x)\big|_{x=\theta},$$

which is simply $\psi'(\theta)/\psi'(1)$. And so the rate of change of the probability that a neighbour of an infected node is susceptible is

$$\frac{dh}{dt} = \frac{1}{\psi'(1)}\frac{d}{dt}\left[\psi'(\theta)\right] = \frac{1}{\psi'(1)}\left[\psi''(\theta)\dot{\theta}\right] = -\beta\varphi\frac{\psi''(\theta)}{\psi'(1)}.$$

Including dh/dt in our equation for $\dot{\varphi}$, we get a system of coupled ordinary differential equations:

$$\begin{aligned}\dot{\theta} &= -\beta\varphi,\\ \dot{\varphi} &= -(\beta+\gamma)\varphi + \beta\varphi\frac{\psi''(\theta)}{\psi'(1)}.\end{aligned} \tag{4.2}$$

Since no analytic solution for $\theta(t)$ exists, system (4.2) was solved numerically in *Matlab* (see Section 5).

To find the rate at which susceptibles are becoming infected, consider a random, susceptible node with degree k. The probability that this node is still susceptible at time t is $\theta(t)^k$. So the proportion of susceptibles is given by

$$S(t) = \sum_{k=0}^{\infty} P(k)\theta(t)^k = \psi(\theta(t)). \tag{4.3}$$

And the rate at which susceptibles are becoming infected is

$$-\dot{S} = -\frac{d}{dt}\left[\psi'(\theta(t))\right] = \beta\varphi\psi'(\theta(t)). \tag{4.4}$$

However, we want this rate in terms of individuals in the population, so we multiply (4.4) by N and add I_0:

$$-\dot{S} = I_0 + N\beta\varphi\psi'(\theta(t)). \tag{4.5}$$

4.1 *Poisson degree distribution*

If we assume that the degree distribution of the network is Poisson with parameter λ, then $P(k) = \lambda^k e^{-\lambda}/k!$ and the probability generating function is

$$\begin{aligned}\psi(x) &= \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} x^k \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda x)^k}{k!} \\ &= e^{-\lambda} e^{\lambda x} \\ &= e^{\lambda(x-1)}.\end{aligned}$$

And the average degree of the network is $\psi'(1) = \lambda$.

System (4.2) becomes

$$\begin{aligned}\dot{\theta} &= -\beta\varphi, \\ \dot{\varphi} &= -(\beta+\gamma)\varphi + \beta\varphi\lambda e^{\lambda(\theta-1)}.\end{aligned} \tag{4.6}$$

4.2 *Negative binomial distribution*

It is, perhaps, more realistic to assume that a network of individuals has a negative binomial degree distribution. Here, the parameters p is the probability that the infection successfully spreads from one node to another, and n is the number of unsuccessful contacts—where the infection was not spread—until a successful contact—where the infection *was* spread.

The probability of a random node having degree k is given by

$$P(k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}.$$

And the probability generating function is [5]

$$\psi(x) = \left(\frac{1-p}{1-px}\right)^n.$$

Differentiating this generating function once with respect to x allows us to complete (4.5). Evaluating the first derivative at $x = 1$, combined with differentiating the generating function twice, allows us to complete the equation for $\dot{\varphi}$ in (4.2). (These calculations *were* performed; however, they yielded unrealistic results and so were not included—see section 6.)

5. Analysis

Via *Matlab*, vectors of values for $\theta(t)$ and $\varphi(t)$ were calculated by solving system (4.2) using the ODE solver, `ode45`. Table 1 contains the values used in the calculation.

A function, $f(t, x_1, x_2)$, of system (4.2) was assembled and evaluated for $t \in [0, 400]$ with initial conditions $\theta(0) = 1$ and $\varphi(0) = I_0/N$. Here, $x_1 = \theta$ and $x_2 = \varphi$. The initial condition for θ is reasonable since at time

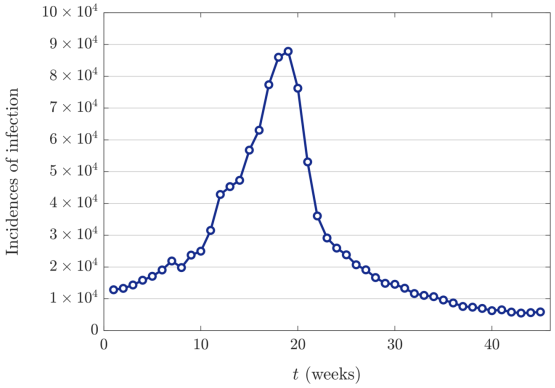

FIG. 3: Number of hospital and clinic visits for influenza-like illnesses—this includes influenza and other illness that cause symptoms typical to those of influenza. Data is from September 30, 2017 to September 29, 2018 [2].

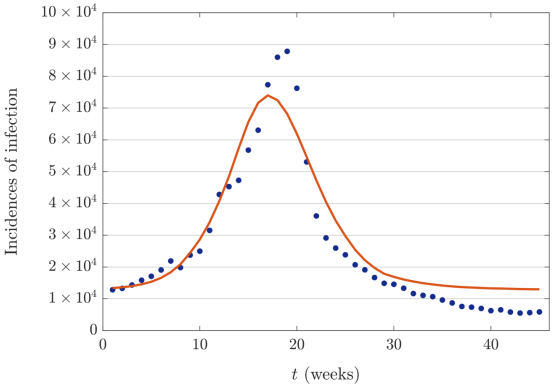

FIG. 4: The red line is (4.5) using the estimated values $\beta = 0.0013$ and $\lambda = 31$. The points are the number of influenza infections starting September 30, 2017.

$t = 0$ we would expect the probability that the infection has *not* spread to be 1—there has been no time for any edges to have transmitted the infection.

```
f = @(t,x)[-beta*x(2); -(beta + gamma)*x(2) + beta*x(2)*exp(5(x(1) - 1))];
[t,xa] = ode45(f, [0 400], [1 3.9239e-05]);
```

Column 1 of `xa` contains 53 values of $\theta(t)$ and column 2 contains 53 values of $\varphi(t)$. So we export these vectors into *R* for model fitting.

```
theta = xa(:,1);
phi = xa(:,2);
```

In *R*, we use (4.5) to find the theoretical curve of the incidences of infection based on hospital and clinic visits from influenza-like illnesses. This data was taken from the Centers for Disease Control and Prevention website [2]. These experimental values were extracted and placed in a vector called `incidences_data`.

```
model <- nlsLM(incidences_data ~ 12839 + 327200000*beta*lambda*phi*exp(lambda(theta
- 1)), start = list(beta = 0.06, lambda = 5))
```

A summary of `model` yielded estimated values of $\beta = 0.0013$ with a standard error of 7×10^5 and $\lambda = 31$ with a standard error of 3. Using these values, the final model was plotted over the data in Fig. 4.

t	$\in [0, 400]$
β	0.06 new infections/day
γ	0.2 individuals recover/day
λ	5 contacts
N	327.2 million individuals
I_0	12,839 individuals
$\theta(0)$	1
$\varphi(0)$	$I_0/N = 3.9 \times 10^{-5}$

Table 1: Values used in the numerical calculations assuming a Poisson distribution.

6. Discussion

The results of the section 5 imply that it takes an average of 771 days for an infected to transmit the infection to a susceptible. Obviously, this is not very reasonable given that within 20 weeks of influenza spread there is a maximum (see Fig. 3).

There are many possible sources of error—one of the most obvious being that a Poisson degree distribution is not the most realistic assumption [6]. Perhaps a negative binomial distribution or scale-free network would have been better choice [7]. (An immense amount of time and energy was spent attempting to fit a negative binomial distribution to the data; however, these attempts provided no better results.) Another possible source of error could have been the assumption that recovered individuals do not become recategorized as susceptibles. A more realistic assumption would have been that individuals who have recovered from one strain of influenza become susceptible again since they are at risk for becoming infected by another strain.

Upon review, it appears obvious that taking averages of the data over many years would have reduced large variances in certain years and may have provided more accurate results. Despite the latter two potential sources of error, it seems evident that the degree distribution had the most effect on the results of the analysis. Given more time and guidance, it would have been possible to create a more accurate model of influenza spread using, say, a negative binomial degree distribution. However, any attempts at performing the analysis with other degree distributions led to inaccurate and—similarly—unrealistic results.

References

[1] Joel C Miller. A note on a paper by erik volz: Sir dynamics in random networks. *Journal of mathematical biology*, 62(3):349–358, 2011.

[2] Centers for Disease Control and Prevention. Weekly u.s. influenza surveillance report. `https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html`, March 2019. Accessed on 2019-03-27. Season selection: 2017–2018.

[3] Eric W. Weisstein. Kermack-mckendrick model. `http://mathworld.wolfram.com/Kermack-McKendrickModel.html`. From MathWorld–A Wolfram Web Resource.

[4] Junling Ma, P van den Driessche, and Frederick H Willeboordse. The importance of contact network topology for the success of vaccination strategies. *Journal of theoretical biology*, 325:12–21, 2013.

[5] Proof Wiki. Probability generating function of negative binomial distribution. `https://proofwiki.org/wiki/Probability_Generating_Function_of_Negative_Binomial_Distribution/Second_Form`, May 2013.

[6] Roderick Edwards, April 5, 2019. Personal conversation.

[7] Junling Ma, April 1, 2019. Personal conversation.