Epidemiology Inspired Rumor Modeling

Zijing Wu

1 Introduction

Imagine such a scenario. There is no pandemic. And every Georgia Tech students take their summer class on campus as usual. One afternoon, Miles, a student who had some problems with his latest homework set, approached Professor Mayer's office. He heard Professor Mayer was talking with someone, so he patiently waited by the door while also curiously paid attention to the talk. Suddenly, he thought he heard the other person said 'no final.' This really spurred a great excitement in his mind. 'There will be no final,' he thought, 'I got to tell everyone this news!' He left with his problems unsolved.

The rumor that there would be no final for MATH 2552 was spread out. And our task is to model the spread of rumor in this scenario. Actually, the spread of rumor behaves similarly to the spread of flu. So we could find our solution by first inspecting models of epidemic, in which we are concerned with the numbers of infection and recoveries and their corresponding rate of change. On modeling the epidemic, we have the 'SIR' model, which modeling the rate of changes using differential equation with three variables the sum of which is the concerned population:

$$\begin{aligned} S(t) &: \text{the susceptible, people who have not contracted the disease} \\ I(t) &: \text{the infected, people who have contracted the disease and not recovered} \\ R(t) &: \text{the recovered, people who have recovered from infection} \end{aligned}$$

The population P of concern is the sum of the three variables as $P = S(t) + I(t) + R(t)$. Let's set the time unit to be 'day'. Let's assume that each infected subjects meet a b proportion of the susceptible every day and would certainly infect the latter. Then we have

$$\frac{dS}{dt} = -bS(t)I(t). \tag{1}$$

It indicates the rate of change of the susceptible population in respect to time. We expect the amount of decrease in $S(t)$ to be the same to the amount of increase in $I(t)$. And let's assume a k proportion of the infected will recover every day, which means $R(t)$ increases in proportional to $I(t)$. And as we expect the amount of increase in $R(t)$ to be the same to the amount of increase in $I(t)$, we have

$$\frac{dI}{dt} = bS(t)I(t) - kI(t). \tag{2}$$

$$\frac{dR}{dt} = kI(t). \tag{3}$$

2 Methods

The rumor spread model mimics the spread of infection within a population. The variables 'S' (The susceptible), 'I' (The infected), and 'R' (The recovered), all have an analogical role in the SIR model in epidemic modeling. Their rates of change are dependent on each other. Thus, we can model the spread of rumor within a population in a similar way [1].

Let's think about the scenario described at the beginning. So, the disease spread is the rumor that MATH 2552 is not going to have final. The population of concern is the students taking this class in summer 2020, which has a total of 163. For the variable $R(t)$, instead of applying the original concept of 'recovered,' let's introduce a new population, which consists of the 'rational,' who could 'counter-infect' the 'infected' by telling them the truth. let's set the initial number of rational people to be 4, Professor. Mayer plus our 3 TAs of the course. And we have $I(0) = 1$, which stands for Miles, the rumor source in our scenario and $S(0) = P = 163$, the total number of students enrolled in the class. $S(0)$ should be $163 - 1 = 162$, but the simplification should be reasonable. Thus, we have modeled the spread of rumor with the IVP:

$$\frac{dS}{dt} = -bSI(t) \quad S(0) = 163 \tag{4}$$

$$\frac{dI}{dt} = bS(t)I(t) - kI(t) \quad I(0) = 1 \tag{5}$$

$$\frac{dR}{dt} = kI(t) \quad R(0) = 4 \tag{6}$$

Here, we have a first order homogeneous nonlinear ODE system. By inspection, with the techniques learned in this class, we are not able to find the analytical solutions to the system. To find the solutions, we need to use numerical methods.

Euler method is a numerical method sufficient to solve for the solutions. The intuition behind Euler method is to approximate the values of the solution functions at the next step t' using values and the gradients of the solution functions at the current t. We can directly obtain the gradients by substituting the values of the solution functions at time t. As we know the function values when $t = 0$, which are the initial conditions, we can obtain the function values of the subsequent steps by propagation. Let each step be δt which is reasonably small.

For example, let $\delta t = 1$ day, $b = 0.1$,and $k = 0.2$, then the function values at $t = 1$ can be approximated as follow:

$$S(1) = S(0) + S'(0) \cdot \delta t = 163 + (-16.3) = 146.7 \tag{7}$$
$$I(1) = I(0) + I'(0) \cdot \delta t = 1 + 16.1 = 17.1 \tag{8}$$
$$R(1) = R(0) + R'(0) \cdot \delta t = 4 + 0.2 = 4.2. \tag{9}$$
$$\tag{10}$$

As the solution values represent number of people, we will round off the decimals as we want to get meaningful information about the distribution of the S, I, R at a given day. Other than that, we would continue to iterate the process with the obtained results to find approximated curves for the solutions.

In this case, to solve for the solutions, we will use the odeint function in Scipy, an open source python library [2]. The function odeint function integrate the differential equation system using numerical approach, the methodology of which is similar to Euler method's as described above. We will test with certain initial value conditions and plot the solution curves.

3 Results

3.1 Model

We first test the model by finding the solution curves with the initial conditions introduced. Let's suppose on average, every day, a student in the course closely interact with 2 students who are also in the course, and let's suppose the 'rationals' will encounter 20 percent of the misled students and tell them truth. Then we can construct the following model:

$$\frac{dS}{dt} = -\frac{2}{163} SI(t) \quad S(0) = 163 \tag{11}$$
$$\frac{dI}{dt} = \frac{2}{163} S(t)I(t) - 0.2I(t) \quad I(0) = 1 \tag{12}$$
$$\frac{dR}{dt} = 0.2I(t) \quad R(0) = 4. \tag{13}$$

Since we cannot find an analytical expression for the solution, we can only find the approximated values of the solution functions at a given time t.

Let's plot the the solution curves and represent the values of the solution in a graph and table below.

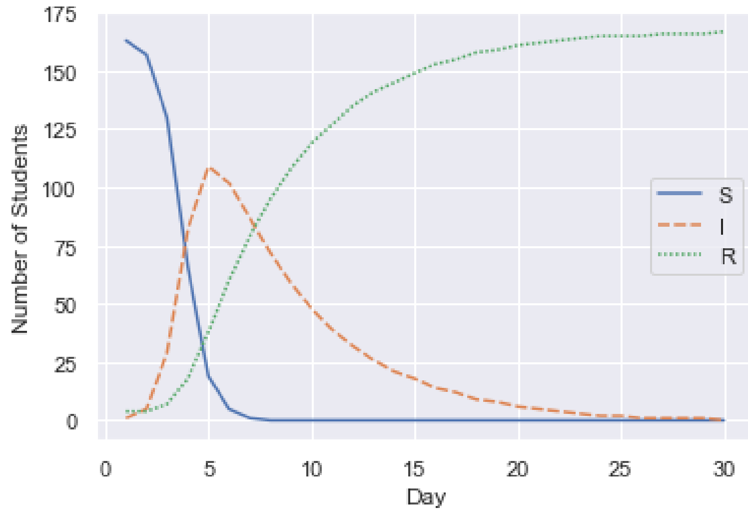

Figure 1: The three curves represent the change of number of the susceptible, infected, and rational subjects with $b = 2/163$, $k = 0.2$, and initial conditions mentioned in Eq.(11), Eq.(12), and Eq.(13).

Day	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29
S	163	130	19	1	0	0	0	0	0	0	0	0	0	0	0
I	1	29	109	87	59	39	26	18	12	8	5	3	2	1	1
R	4	7	38	79	108	127	141	149	155	159	162	164	165	166	166

Figure 2: The table shows us the values of the curves in the 30 day range. Though only the results of odd number days are shown for ideal visual representation, the trend of changing of the number should be clear.

We can see that with $b = 2/163$, $k = 0.2$, and our first set of initial value conditions, the rumor spread rapidly. In five days, almost every one has been spread with the rumor. The number of students 'infected' reaches its peak of 109 on day 4. Notice that the highest infection number is about two thirds of the population. It means suppose $k = 0.2$, which means the rational subjects will clear the rumor for 0.2 of the infected every day, then it is quite effective and would stop the growth of infection at an early stage. And as we reach the end of the 30 days, the rumor is cleared.

3.2 Model Variations

So far, we have successfully constructed a model to model the spread of the rumor. But the model is based on some ideal assumptions; for example, we assume that only the instructors can be the rational people who could counter-infect the 'infected,' and the model indicates that there is no variety among the students. Thus, we are doing to derive 2 variations from our first model as an attempt to make it more realistic.

3.2.1 Variation 1

As a 'rational' student can also 'recover' an 'infected' student, though should be at a relatively lower rate, we are going to take this fact into account. Then we have

$$\frac{dS}{dt} = -bS(t)I(t) \quad S(0) = 163 \tag{14}$$

$$\frac{dI}{dt} = bS(t)I(t) - k_i I(t) - k_s I(t)R(t) \quad I(0) = 1 \tag{15}$$

$$\frac{dR}{dt} = k_i I(t) + k_s I(t)R(t) \quad R(0) = 4, \tag{16}$$

where k_i is the proportion of the "infected" our four instructors would interact with each day and k_s is the proportion of the 'infected' a rational student would meet and clarify the rumor every day. Then we set the value for the constants so we get

$$\frac{dS}{dt} = -\frac{2}{163}S(t)I(t) \quad S(0) = 163 \tag{17}$$

$$\frac{dI}{dt} = \frac{2}{163}S(t)I(t) - 0.2I(t) - \frac{2}{163}I(t)R(t) \quad I(0) = 1 \tag{18}$$

$$\frac{dR}{dt} = 0.2I(t) + \frac{2}{163}I(t)R(t) \quad R(0) = 4. \tag{19}$$

Let's plot the solution curves and table below.

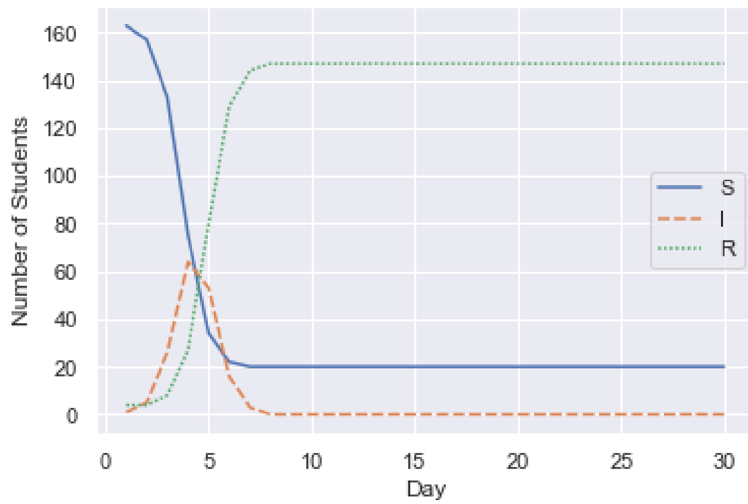

Figure 3: The three curves represent the change of number of the susceptible, infected, and rational subjects with $b = 2/163$, $k_i = 0.2$, $k_s = 2/163$, and initial conditions mentioned in Eq.(17), Eq.(18), and Eq.(19).

Day	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29
S	163	133	34	20	20	20	20	20	20	20	20	20	20	20	20
I	1	26	53	3	0	0	0	0	0	0	0	0	0	0	0
R	4	8	80	144	147	147	147	147	147	147	147	147	147	147	147

Figure 4: We can see that the introduction of k_s strengthened the overall recover rate of the whole population. The infection was wiped out before all students would have been infected.

3.2.2 Variation 2

Our second model variation is based on the first variation. Here, let's try to address another assumption we made before, which is that we assumed all students act in the same way. But in reality, it is not the truth. While the majority of the students take classes normally, some students might not go to the class and interact with the instructors due to a variety of reasons. We could account this in our model by setting new population groups for these students:

$$\frac{dS1}{dt} = -\frac{2}{163}S1(t)(I1(t) + I2(t)) \quad S1(0) = 163 * 0.9 \simeq 147 \tag{20}$$

$$\frac{dS2}{dt} = -\frac{2}{163}S2(t)(I1(t) + I2(t)) \quad S2(0) = 163 * 0.1 \simeq 16 \tag{21}$$

$$\frac{dI1}{dt} = \frac{2}{163}S1(t)(I1(t) + I2(t)) - 0.2I(t) - \frac{2}{163}I1(t)R(t) \quad I(0) = 1 \tag{22}$$

$$\frac{dI2}{dt} = \frac{2}{163}S2(t)(I1(t) + I2(t)) - \frac{2}{163}I2(t)R(t) \quad I(0) = 0 \tag{23}$$

$$\frac{dR}{dt} = 0.2I1(t) + \frac{2}{163}I1(t)R(t) + \frac{2}{163}I2(t)R(t) \quad R(0) = 4.w \tag{24}$$

So now we suppose there are 10 percent of the students would not go to class and interact with the instructors, though they would interact with their fellow students regularly. The introduction of this new group of student should presumably prolong the time for the rumor to disappear. Let's plot the solution curves and table to see the change.

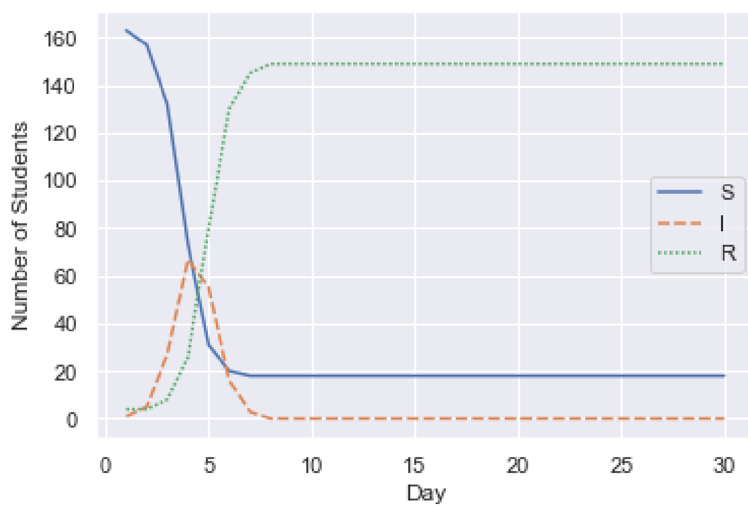

Figure 5: We combine S1 and S2, I1 and I2 when plotting the curves. We could see that the solution curves do not vary much from the first variation's.

Day	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29
S	163	132	31	18	18	18	18	18	18	18	18	18	18	18	18
I	1	27	55	3	0	0	0	0	0	0	0	0	0	0	0
R	4	8	80	145	149	149	149	149	149	149	149	149	149	149	149

Figure 6: The table shows us the values of the curves in the 30 day range. Only the odd days are shown.

But the number of students had never been infected only decreased by 2. To explain this, a reasonable speculation is that the counter-infection power of the instructors, with $k_i = 0.2$ is too strong after all. And it should also due to the fact that 10 percent of the students is not a large proportion, and their conception of the rumor would be cleared by peers over time. So it would eventually not affect the overall trend significantly.

4 Connection with Class

The rumor spread model mimics the spread of infection in a population. The variables ‘S’ (The susceptible), ‘I’ (The infected), and ‘R’ (The rational) all have an analogical role in the epidemic SIR model. Their rates of change are dependent of each other. Thus, we model the spread of rumor within a population by constructing an ordinary differential equation system. At the beginning of the project, I assumed that I grows in proportional of S, so it became a linear differential equation system which I can manually solve with the knowledge in Chapter 3 and 6 in the textbook. The calculation provided me with hints of the growth trend of the curves in the model.

But with the knowledge has been taught in the class, we are not able to fully solve it because of the nonlinear nature of the system. While exactly why we cannot find an analytical solution for the system is to be determined in future revision of the project, for now, we could only seek for numerical methods, which is used in the odeint method of the Scipy library.

5 Discussion

So now we have constructed one model and its two variation for the spread of rumor in the scenario of our class. Though the model is based on many ideal assumptions, it models the complex problem of spread of rumor and disease within a population with relatively simple differential equation systems. Hopefully it can provide readers with meaningful, incisive insights into both an interesting topic of rumor spread and the current concerned one with the pandemic.

References

[1] S. Florkowski and R. Miller, *6-018-s-exploringsirmodel*, May 2018. [Online]. Available: https://www.simiode.org/resources/4795.

[2] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, *et al.*, "Scipy 1.0: Fundamental algorithms for scientific computing in python", *Nature methods*, vol. 17, no. 3, pp. 261–272, 2020.

6 Code

This is the code I wrote to find the solutions of the differential equation system and plot the curves and table.

```
from scipy.integrate import odeint
import numpy as np
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
sns.set()


def pend_ini(y, t, b, k):
    S, I, R = y
    dsdt = -b * S * I
    didt =  b * S * I - k * I
    drdt = k * I
    dydt = [dsdt, didt, drdt]
    return dydt

def pend_v1(y, t, b, k_i, k_s):
    S, I, R = y
    dsdt = -b * S * I
    didt =  b * S * I - k_i * I  - k_s * R * I
    drdt = k_i * I  + k_s * R * I
    dydt = [dsdt, didt, drdt]
    return dydt

def pend_v2(y, t, b, k_i, k_s):
    S1, S2, I1, I2, R = y
    ds1dt = -b * S1 * (I1 + I2)
    ds2dt = -b * S2 * (I1 + I2)
    di1dt =  b * S1 * (I1 + I2) -  k_i * I1 - k_s * R * I1
    di2dt =  b * S2 * (I1 + I2) - k_s * R * I2
```

```
    drdt = k_i * I1 + k_s * R * I1 + k_s * R * I2
    dydt = [ds1dt, ds2dt, di1dt, di2dt, drdt]
    return dydt


def plot_curve(sol, sol_c=None):
    df = pd.DataFrame(sol, columns=['S',  'I', 'R']).astype(int)
    df.index = df.index + 1
    ax = sns.lineplot(data=df) #plot the curves
    ax.set_xlabel('Day')
    ax.set_ylabel('Number of Students')

def plot_table(sol, sol_five_var=1):
    if type(sol_five_var) != int:
        df = pd.DataFrame(sol_five_var,
                    columns=['S1', 'S2', 'I1','I2', 'R']).astype(int)
        df.index = df.index + 1
        df.index.name = 'Day'
        keep = np.arange(0, 29, 2)
        df= df.iloc[keep].reset_index().T

        fig, ax = plt.subplots()

        # hide axes
        fig.patch.set_visible(False)
        ax.axis('off')
        ax.axis('tight')
        ax.table(cellText=df.values, rowLabels=df.index, loc='center')
        fig.tight_layout()
        plt.show()

    df = pd.DataFrame(sol, columns=['S',  'I', 'R']).astype(int)
    df.index = df.index + 1
    df.index.name = 'Day'
    keep = np.arange(0, 29, 2)
    df= df.iloc[keep].reset_index().T

    fig, ax = plt.subplots()

    # hide axes
    fig.patch.set_visible(False)
    ax.axis('off')
    ax.axis('tight')
    ax.table(cellText=df.values, rowLabels=df.index, loc='center')
    fig.tight_layout()
    plt.show()
```

```python
def plot(b=2/163, k_i=0.2, k_s=2/163, y0=[163, 1, 4],
                    t=np.linspace(1, 30, 30), table=False, model='initial'):
    if model == 'initial':
        sol_c = odeint(pend_ini, y0, t, args=(b, k_i))
    elif model == 'v1':
        sol_c = odeint(pend_v1, y0, t, args=(b, k_i, k_s))
    elif model == 'v2':
        y0=[163 * 0.9, 163 * 0.1, 1, 0, 4]
        sol = odeint(pend_v2, y0, t, args=(b, k_i, k_s))
        sol = np.array(sol).T
        sol_c = np.zeros((3, sol.shape[1]))
        sol_c[0] = sol[0] + sol[1]
        sol_c[1] = sol[2] + sol[3]
        sol_c[2] = sol[4]
        sol_c = sol_c.T
        sol = sol.T
    if table:
        if model != 'v2':
            plot_table(sol_c)
        else:
            plot_table(sol_c, sol_five_var=sol)
    plot_curve(sol_c)


# plot the model under the initial assumption
plot(k_i=0.2, table=True)

# plot mode variation 1
plot(k_i=0.2, model='v1', table=True)

# plot mode variation 2
plot(model='v2', table=True)
```