Modeling Rumour Spread

Nicholas Mankowski

December 13th, 2023

1 Introduction

A rumour has started to spread at Ohio Wesleyan. The rumour is that Ohio Wesleyan's premier fraternity, Oozma Kappa, will not have their annual pumpkin carving contest due to a shortage of pumpkins in the Delaware area. This is of course not true, as farmers indicate a spectacularly average year for pumpkin crop yields. This rumour is being spread by Dr. McCurdy in the math department, as he wants revenge against Dr. Guo's show-stopping pumpkin from last year. Dr. McCurdy cannot begin to think what will happen if he loses the pumpkin carving contest two years in a row, so he has no other option than to try and stop the event. Dr. McCurdy begins spreading this rumour in his math modeling class, but some of the students in this class do not believe the rumour and do not spread it, while others in the class are telling everybody around the school. These reactions are rather telling, and offer us a glimpse into how the rumour will spread across the university. In this writeup, we will explore how Dr. McCurdy's rumour will spread around the campus population by using SIR-like modeling techniques.

2 Description of Mathematics

We will use a variation of the SIR modeling technique to analyze the spread of the rumour throughout Ohio Wesleyan. We can create three different categories to describe the possible status of a student/faculty member at Ohio Wesleyan. These categories are as follows:

- H: Somebody who has heard the rumour, and is actively spreading it.

- I: Somebody who has not heard the rumour.

- R: Somebody who has heard rumour, but is refusing to spread it.

We can represent the pathways between the different statuses of people as the following:

- $H \rightarrow I$: People who were spreading the rumour, but have forgotten the rumour. This is purely dependent on the number of people who are spreading the rumour, so we can describe this rate as δH.

- $H \rightarrow R$: People who were spreading the rumour, but have since refused. This is dependent on the number of people who are spreading the rumour, so we can describe this rate as βH.

- $I \rightarrow H$: People who have heard the rumour and begin to spread it. This is dependent on the interaction between people who have heard the rumour and those who haven't, so we can describe this rate as αIH.

- $R \rightarrow H$: People who have previously refused to spread the rumour, but begin to spread it. This is dependent solely on the number of people who refuse to spread the rumour, so we can describe this rate as εR.

- $R \rightarrow I$: People who were refusing to spread the rumour, but have since forgotten it. This is purely dependent on the number of people who are currently refusing to spread the rumour, so we can describe this rate as γR.

With this information, we can represent our model as the following graph:

Figure 1: HIR Model Diagram
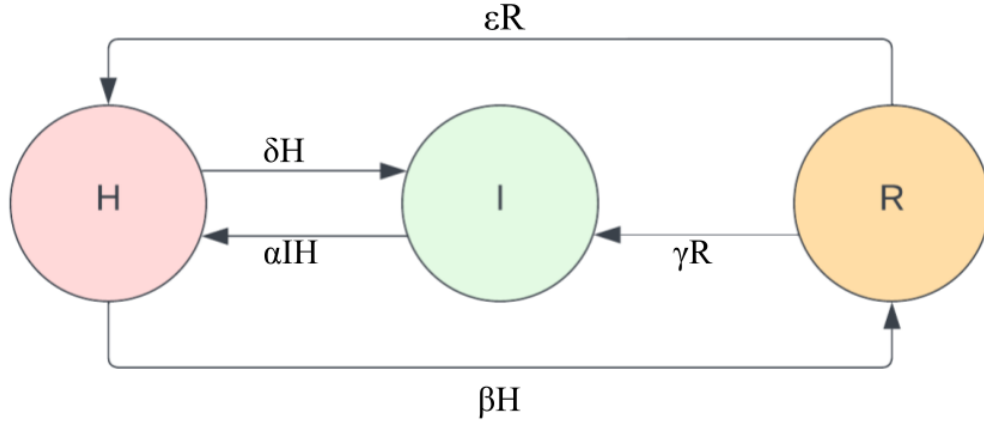
Using this model, we can set up three differential equations to represent the rate of change for each node. Doing this, we have the following system of differential equations:

$$\frac{dH}{dt} = \varepsilon R + \alpha IH - \delta H - \beta H \qquad\qquad = \varepsilon R + \alpha IH - (\delta + \beta)H$$

$$\frac{dI}{dt} \qquad\qquad\qquad\qquad\qquad\qquad = \delta H + \gamma R - \alpha IH$$

$$\frac{dR}{dt} = \beta H - \varepsilon R - \gamma R \qquad\qquad\qquad = \beta H - (\varepsilon + \gamma)R.$$

While this system of differential equations may be difficult or even impossible to solve analytically, we are able to discretize the equations to produce an approximate solution to the system. This technique has the added benefit that all systems will have a similar approach in solving and will not depend on the equations themselves, as is often the case with solving differential equations analytically. In order to produce this approximation, we must convert

our continuous differential equations to discrete difference equations. Doing this, we have

$$\frac{\Delta H}{\Delta t} = \frac{H(t + \Delta t) - H(t)}{\Delta t} = \varepsilon R(t) + \alpha I(t)H(t) - (\delta + \beta)H(t)$$
$$\frac{\Delta I}{\Delta t} = \frac{I(t + \Delta t) - I(t)}{\Delta t} = \delta H(t) + \gamma R(t) - \alpha I(t)H(t)$$
$$\frac{\Delta R}{\Delta t} = \frac{R(t + \Delta t) - R(t)}{\Delta t} = \beta H(t) - (\varepsilon + \gamma)R(t),$$

where t is our current time, Δt is our discrete time step, ΔH is our discrete step in H, ΔI is our discrete step in I, and ΔR is our discrete step in R.

Now that we have a discretized system, we can solve each difference equation to give us the next time step using the data from the previous time step. Doing this, we obtain

$$H(t + \Delta t) = \Delta t(\varepsilon R(t) + \alpha I(t)H(t) - (\delta + \beta - 1)H(t))$$
$$I(t + \Delta t) = \Delta t(\delta H(t) + \gamma R(t) - \alpha I(t)H(t)) + I(t)$$
$$R(t + \Delta t) = \Delta t(\beta H(t) - (\varepsilon + \gamma - 1)R(t),$$

which gives us an estimation of our differential equation solution. Using this, we can find our approximate curves iteratively given an initial condition.

3   Applications in Code

Given our explicit discretized equations derived in 2, we can iterate through values using a `for` loop in MATLAB in order to produce an approximate solution to our system of differential equations. In order to accomplish this, we must first declare our parameters $\alpha, \beta, \gamma, \& \varepsilon$. We can accomplish this as follows:

```
1  % Set up parameters
2  alpha = 0.001;
3  beta = 0.04;
4  gamma = 0.01;
5  delta = 0.01;
6  epsilon = 0.005;
```

Now that we have our parameters set up, we can create our corresponding variables to handle time. To do this, we will need an initial time, a maximum time, and a time step. We will take our initial time to be 0, our maximum time to be 100, and our time step to be 1. We do this as follows:

```
1  % Set up time variables
2  deltaT = 1; % Time step
3  startTime = 0; % Initial time
4  maxTime = 100; % Maximum time(non inclusive)
```

Note: We do not want our maximum time to be inclusive, since we are counting from zero. If we set up our maximum time to be inclusive, we will have a mismatch in array lengths

when plotting our solution.

In order to set up our vectors to ultimately hold our approximations, we must create a vector for each value that holds $\frac{t_{\max}-t_{\min}}{\Delta t}$ values where $t_{\max}$ is the maximum time value and $t_{\min}$ is the minimum time value. So, we declare our vectors as the following:

```matlab
1  % Set up arrays that will hold our values of H, I, & R.
2  H = ones(1, (maxTime - startTime) / deltaT);
3  I = ones(1, (maxTime - startTime) / deltaT);
4  R = ones(1, (maxTime - startTime) / deltaT);
```

Now, we must set up our initial conditions. Unfortunately, due to low enrollment numbers, Ohio Wesleyan has a total of 200 students and faculty. Thus, we must have 199 people unaware of the rumour and 1 person (Dr. McCurdy) spreading the rumour initially. This also means that we initially have nobody refusing to spread the rumour. Thus, we can set up our vectors containing our data and their initial conditions as follows:

```matlab
1  H(1) = 1; % The original person spreading the rumour
2  I(1) = 199; % Initial population of people not spreading rumour
3  R(1) = 0;
```

Now that we have our initial conditions set up, we are able to iterate through our difference equations to produce an approximation of our solution. We can do this in MATLAB by doing the following:

```matlab
1  % Iterate through each time step.
2  % We must exclude the last value, since we are setting the value at t + 1,
3  % so iterating until length(H) would cause us to go out of bounds.
4  for t = 1:length(H) - 1
5      % Calculate the next value of H, I, & R based on the previous values using ...
       our difference equations.
6      H(t + 1) = deltaT * (epsilon * R(t) + alpha * I(t) * H(t) - (delta + beta) ...
       * H(t)) + H(t);
7      I(t + 1) = deltaT * (delta * H(t) + gamma * R(t) - alpha * I(t) * H(t)) + I(t);
8      R(t + 1) = deltaT * (beta * H(t) - (epsilon + gamma) * R(t)) + R(t);
9  end
```

Now that we have assembled our model, we can plot H, I, & R by creating a representation of continuous time and plotting against this. Doing this, we have the following:

```matlab
1  % "Continuous" time, subtract deltaT since we do not want it to be inclusive.
2  % If we make this inclusive, it will cause errors in the length of our H, I, & ...
       R vectors,
3  % since we are counting from zero.
4  timeCts = startTime:deltaT:maxTime - deltaT;
5
6  % Plot the model
7  hold on;
8  % Plot H, I, & R
9  plot(timeCts, H, 'DisplayName', 'H, students actively spreading the rumour');
10 plot(timeCts, I, 'DisplayName', 'I, students ignorant to the rumour');
11 plot(timeCts, R, 'DisplayName', 'R, students refusing to spread the rumour');
```

```
12  % Label the graph and add a legend
13  xlabel('Time (t)')
14  ylabel('Population')
15  title('Rumour Spread Across OWU Campus and Faculty')
16  legend;
17  hold off;
```

4   Discussion

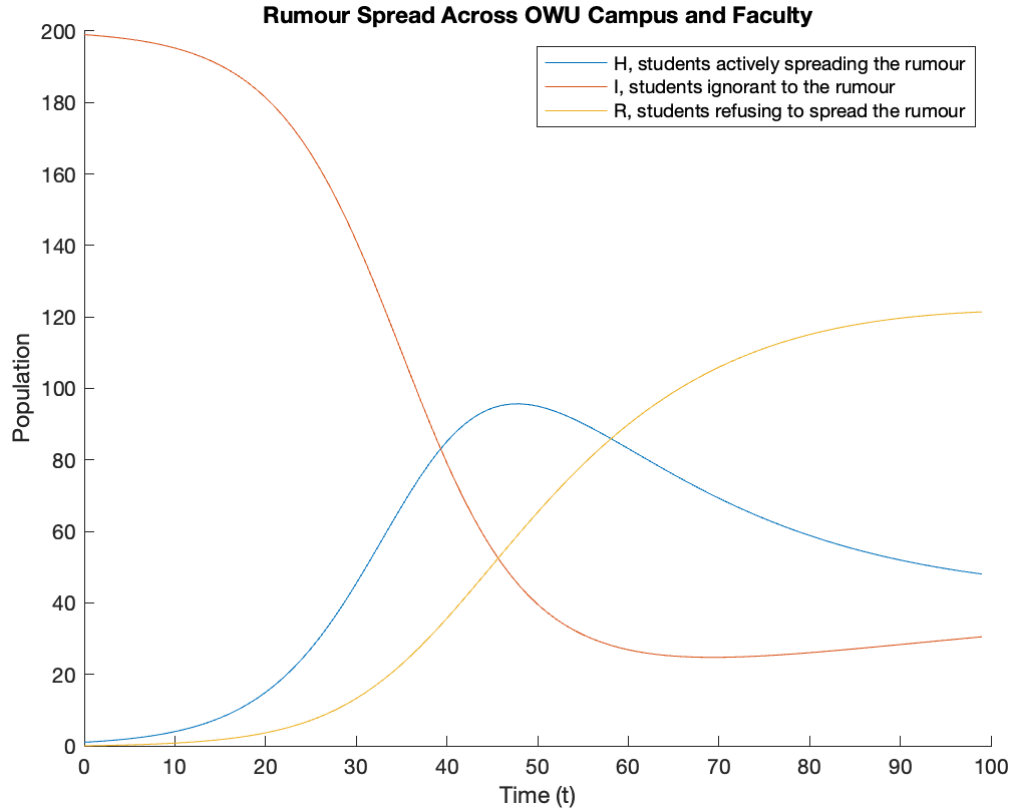By following the methodology described in 3, we produce the following graph:

Figure 2: Rumour Spread Model

4.1   Parameter Analysis

We can look at the effect that the parameters will have on our model by digging a bit deeper into the underlying math we had established and using a bit of intuition. Via the results shown in 2, we can see that the parameters,

$$\alpha = 0.001\,, \quad \beta = 0.04\,, \quad \gamma = 0.01\,, \quad \delta = 0.01\,, \quad \varepsilon = 0.005\,,$$

produce realistic curves that can be used to approximate the rumour spread. Further examining these parameters, we should note that their order is closely related to the time scale that we are looking at. For example, $\alpha = 0.001$ moves significantly more people if we are looking at a larger span of time than a smaller one. This relation to the arbitrary time scale

5

means that we should moreso focus on the relationships between these parameters, rather than their exact values. Looking at our paremeters, $\gamma = 0.01$ and $\beta = 0.04$ implies that ignorant people hearing about the rumour are much less likely to spread the rumour than they are to refuse to spread it. Thus, we are assuming our population to be fairly resistant to spreading the rumour. Similarly, $\varepsilon = 0.005$ and $\gamma = 0.01$ means that the population of people refusing to spread the rumour are much more likely to forget about the rumour than they are to begin spreading it. Overall, the ratios between these parameter values seem to match intuition regarding what masses of people will do when facing a rumour. Thus, they are reasonable for this model.

Looking at our graphic first shown in 1, we can draw some conclusions about how changing our model parameters will affect our model. We can create the following conclusions based on this:

- By increasing/decreasing the value of α, we are altering the percentage of interactions between a rumour spreading student and an ignorant student that will cause that ignorant student to learn of the rumour. If we are to increase this value, it will cause more people to spread the rumour, but also could cause more people to refuse spreading the rumour via the pathway from $H \rightarrow R$.

- By increasing/decreasing the value of β, we are changing the percentage of rumour spreading students that eventually refuse to spread the rumour at each time step. A high value for this will cause less people to be spreading the rumour and more people to be refusing to spread the rumour. This is because a high β value will cause more people to flow across the pathway $H \rightarrow R$ and thus more people will refuse to spread the rumour who were previously actively spreading the rumour.

- By increasing/decreasing the value of γ, we are changing the rate at which people forget the rumour among those not actively spreading it. By having a higher γ value, there will be fewer people who are refusing to spread the rumour and more people ignorant to the rumour. This may seem to not have an effect, but a larger amount of ignorant people could possibly mean more people eventually spreading the rumour, so having this value be very large may have the effect of increasing the number of people spreading the rumour.

- By increasing/decreasing the value of δ, we are causing more people who are actively spreading the rumour to forget it. This rate is likely very small, as it is unlikely that somebody spreading the rumour would forget it, but a higher δ value will likely lead to more people ignorant to the rumour and thus less who are actively spreading it. This is because increasing δ will increase the rate at which people cross the pathway from $H \rightarrow I$.

- By increasing/decreasing the value of ε, we are increasing the rate at which people move from $R \rightarrow H$. This will increase the number of people who are actively spreading the rumour and decrease the number of people who are refusing to spread the rumour.

4.2 Model Analysis

Using the model we have produced, a rumour would spread in a way that begins with many people the rumour. Throughout most of the spread of this, there would also be a decrease in people who are ignorant to the rumour. By the end of the spread, most people will be aware of the rumour in one way or another, despite some people forgetting throughout the spread. It is also apparent that by the end, the largest group of people would be those who are refusing to spread the rumour. There would also be a peak of people who would be spreading the rumour early on. This all makes intuitive sense, as people are likely to come to their senses and realize that it is innappropriate to spread a rumour and eventually stop. Overall, this model seems to do a good job of approximating the spread of a rumour across Ohio Wesleyan's campus, but is still likely a massive oversimplification of what would happen during such a process.

4.3 Topics for Further Consideration

As discussed above, this model is most certainly a massive oversimplification of what occurs if a rumour were to spread across Ohio Wesleyan's campus. Despite making intuitive sense, there are several considerations that our model does not take into account. Here is a short list of such things:

- How does a rumour spread online as opposed to in person?

- How do people's attitudes toward a rumour change as they pass between nodes? More specifically, are people less likely to spread a rumour after they've been reminded of it several times?

- What if parameters are time dependent? For example, could it be the case that people are less likely to spread a rumour after the rumour has been spread for a long time?

We could address these concerns a variety of ways during further study. For example, the difference in spreading online versus in person could be modeled by having multiple pathways in between nodes, in particular, between the I & H nodes. These would both likely still be interactions between I & H, but the online path will likely have a larger coefficient than in person, since more people are likely to be reached by posting online.

When thinking about how somebody's attitude would change based on how many times they've been to a particular node, it follows intuition that somebody's previous experiences would influence their current decisions. We can explore this idea by modeling individual people that have their own coefficients. For the sake of model simplicity, all individuals would begin with the same starting coefficients. As a person passes between nodes, their coefficients could be slightly modified each time. To quantify this further, when somebody hears a rumour for the fourth time, they are less likely to forget it than the first time they hear it, so it follows intuitively that they would have a lower δ & γ coefficients.

When handling coefficients becoming time dependent, we could define functions that return each corresponding coefficients based on the current time. This would be rather easy to factor in to our model, and would likely greatly increase the accuracy. This is because rumours are likely to become more boring the longer they are around, so people will be less

likely to spread older rumours. It is also worth noting that the rumour could be exposed as a rumour as time passes, and this would likely effect the coefficients as well. Exactly how time effects these coefficients would likely vary wildly from rumour to rumour though, so special consideration would have to be taken while modeling this.

By factoring in these ideas, we could increase our models accuracy and get a better glimpse into how a rumour would spread throughout campus.