Modeling the Spread of a Rumor

Matthew Mercurio
Professor David Meyer

Introduction

While the Merriam-Webster dictionary defines a rumor as “a statement or report current without known authority for its truth”, the academic definition of a rumor is a bit harder to pin down and varies slightly depending on the discipline.[1,2] One of the earliest mentions of rumors in academic literature is "A Psychology of Rumor", published by Robert H. Knapp in 1944. Knapp analyzes thousands of rumors printed in the Boston Herald during World War II. He found that rumors all have three main characteristics in common. One, rumors are primarily spread by word of mouth and thus have a high chance of being distorted or plain untrue. Two, rumors always contain ‘information’ about a “person, happening or condition”. Three, rumors reflect "the emotional needs of the community." Essentially this means that the rumor is relevant to the population as a whole and is not a private or personal matter. He classified rumors into three categories: pipe dream rumors, fear rumors, and wedge driving rumors. Pipe dream rumors are positive and reflect the hopes and dreams of the community, whereas fear rumors and wedge driving rumors are negative, targeting the communities fears in general and a specific person respectively. Interestingly enough, Knapp found that negative rumors were more likely to be spread than positive rumors.[3] Despite the fact that Knapps definition of a rumor seems pretty straightforward, there are a lot of nuances to consider. For example, in [DiFonzia, Bordia 2017], the context and contents of the words “rumor”, “gossip”, and “urban legend” are compared and contrasted with great detail despite all three terms conveying a similar meaning in everyday life.[4]

However, despite this confusion in academia, most people need little introduction to the concept of a rumor. It is safe to assume that as long as humans have been spreading information around, they have been spreading dubious information around. Most people can think back to a time in their life when they were the victim of a nasty rumor, and how it affected their life negatively. Moreover, rumors can have an enormous impact on public opinion on a larger scale; evidence exists that suggests that rumors have had a significant effect on national elections.[5] With the increased usage of social media and mass communication in the recent decade, information, and hence misinformation has the potential to reach more people than ever before. As a result, successfully modeling the spread of a rumor and how the rumor is potentially stopped is more relevant than ever before.

The earliest model of a rumor was introduced in 1965 by Daley and Kendall (DK). DK noticed that rumors are spread similarly to diseases and thus constructed a model based on leading SIR Epidemic Models at the time. DK constructed a stochastic model, which they were able to analyze the success of using Markov Chains. In each model, they divided their population into three categories:

A. people who have heard the rumor.
B. people who have heard the rumor and are actively spreading it.

C. people who have heard the rumor but no longer spread it.

These categories are purposely similar to the three categories commonly seen in epidemic models: susceptible, infectious, removed.[6] While the original Daley-Kendall model has been tweaked and improved upon many times over since the original paper was released, most notably the Maki–Thompson model, rumor models are still by and large modeled of off Daley and Kendall's model, with the three categories commonly referred to as Ignorant, Spreaders, and Stiflers.[7]

Models today can further be divided into two more categories: network and non-network models. Network models are run on inhomogeneous networks, meaning each person in the population has a different probability of interacting with each other, while Non-Network models are run on homogenous populations, meaning each person in the population has an equal chance of interacting with every other person in the population.[7]

Model

Goal

Our goal is to attempt to model the spread of a rumor. We hope we can introduce a large amount of randomness into the model, but still get interesting results that mirror real life and other rumor models that have been created in the past. We define how the rumor spreads based on social reputation and how surprising the rumor is. While other factors will be present in the model, we are mainly interested in tracking these two.

Set Up

The model is a non network agent based simulation of the spread of a rumor. Each person in the model will be randomly assigned a reputation between 0 and 1. Reputations of the overall population will be distributed according to a Pareto Distribution with $\alpha = 1.16$. Hence roughly 20% of the population will be in the 80th percentile or above. The fact that this 80-20 split occurs at $\alpha = 1.16$ is called the Pareto Principle. Similar to many rumor models previously, it will be based on the Daley-Kendall model with the population divided into three classes: Ignorants, Spreaders and Stiflers. However, the rumor will be targeted at a random specific person in the population[2]. We will call this person the "victim". Initially, one person will be randomly chosen to be a Spreader. We will call this person the "bully".

We will assume that all rumors spread have negative connotations and that they are relevant to the overall population regardless of the reputation of the victim and the bully[3]. We will also assume that the truth of the rumor will always be ambiguous to the people in the population. Nothing can prove or disprove the rumor at any time to influence how the rumor is spread.

For simplicity and for ease of computation we will assume that the rumor is spread person to person.$_1$ Hence we will not try to scale this model up to try and represent a social media network. Since the rumor is only spread person to person, the number of people each person in the population interacts with per day will be normally distributed with a mean of 1% of the total population and a standard deviation of .5% of the total population. Had we tried to model an online network, we probably would have used a Pareto Distribution as it would better represent the number of followers or friends people usually have in a network.

Notice that the subscripted sentences in the last few paragraphs together imply that the rumor being spread is consistent with Knapp's definition of a "wedge driving" rumor. Since the rumor is directly scandalous to the victim and the information that it contains is relevant to the overall population, the victim's reputation is subtracted by the amount of people who believe the rumor, multiplied by a surprise factor constant. The surprise factor is a preset value which we will test multiple values for.

Using Knapp's third assumption: that the rumor is relevant "the emotional needs of the community." allows us to make a huge assumption about the population. We assume that the population is largely homogeneous in their hopes, dreams, way of life, favorite sports team etc. While human variability no doubt plays a huge role in the way information is spread, it is impossible to quantify the wide range of human experiences and doing so would distort what the data tells about the variables we have chosen to define. It also ignores the fact that plenty of homogeneous communities exist around the world and rumors still spread throughout their communities as well.

Also note that each step is a completely arbitrary unit of time. The number of people each person interacts with in a step is chosen so that relatively little change happens in each step so we can better observe how the rumor spreads over time.

Simulation Method

The different roles each person can have will be listed here:

1. Ignorant (I) - Has not heard the rumor will not spread it
2. Spreader - believes the rumor (S_b) - Spreads the fact that the rumor is true
3. Spreader - disbelieves the rumor (S_d) - Spreads the fact that the rumor is false
4. Stifler (R) - Does not spread the rumor but still believes/disbelieves with it

In contrast to the classical Daley-Kendall Model however, our model will have two different types of spreaders. Only the first type of Spreader (S_b) will be present in the simulation until the victim hears the rumor for the first time. Once the victim hears the rumor for the first time, they will immediately become a spreader of the second type (S_d). However, they cannot ever become a stifler.

Interaction Rules

To explain the model we will go class by class starting with Ignorants. For the purposes of this explanation, the category on the left most side will be assumed to be the person receiving information from the other person. In the actual simulation, when two people come into contact with each other, they will both go through the same process, essentially it happens twice, however it will still count as one interaction for each person when determining how many interactions each person has had in each step. This is because exactly what happens depends on who is sharing information to the other.

<u>Ignorants:</u>

When two ignorants come into contact with each other, nothing happens, however it is still considered an interaction. Similarly if an ignorant comes into contact with a sifler, no information is shared between them as the stifler does not spread the rumor willingly.

When an ignorant comes into contact with a spreader, the spreader will attempt to spread the rumor to the ignorant, it doesn't matter what type of spreader they are. The probability of that actually happening is determined by the spreader's belief. If the spreader believes the rumor then
The ignorant will switch with probability β, in converse, if the spreader does not believe the rumor, then the probability of the ignorant becoming a spreader is δ.

<u>Spreaders:</u>

When two spreaders meet, what happens depends on if they share the same belief. If they do share the same belief, then both of them will switch to stiflers with probability λ. The two spreaders will switch independently of one another so it is possible that none, one or both of the spreaders will become stiflers.

When two spreaders meet who have different beliefs, they try to convince each other of their respective opinion. The exact same process happens with two spreaders, they switch their opinions with probability β or δ depending on the other spreaders belief. However,if a spreader fails to convince another spreader of their belief the they become a stifler with probability λ.

If a spreader meets a stifler then it is impossible for the spreader to convince the stifler to change their mind. However the stifler can still change the spreaders belief with probability β or δ depending on the stiflers belief.

Stiflers:

The only interaction not covered still is when a stifler interacts with a stifler. When this happens no information is shared about the rumor but it still counts as an interaction for both people.

Switch Probabilities

Now we will actually define β, δ, and λ.

Lambda: Lambda is set at the beginning of the simulation and stays constant throughout. Multiple values for lambda are tested.

Trust Probability - The probability that person m trusts person n. This is constant across all possible tulips of people not including the victim such that $m \neq n$. If the reputation of person n is higher than the reputation of person m then the trust probability defaults to one. Trust probability is a ratio between two people, ensuring that people of similar social status have a high trust probability between each other. If the reputation of person n is high than the reputation of person m, then the trust probability defaults to one ensuring people of a high social status have a high trust probability among all people.

$$P_{m,n}(trust) = \frac{\text{reputation}_n}{\text{reputation}_m}$$

Information of an Event - The amount of information a person receives when an event of probability p_i occurs.[8] Also known as the amount of surprise. At $p_i = 0$ the surprise is not defined which makes sense intuitively since it is impossible for an event with probability zero to happen. As p_i goes to one, the surprise decreases until the surprise is 0 for an event with probability 1. It is important to note that the amount of surprise is not a probability, it is actually measured in its own units depending on what log base is used. The choice of $\log_2$ will be explained shortly.

$$I_i = -log_2 p_i$$

Probability of the Rumor Being True - Since we set the surprise factor at the beginning of the simulation, we can calculate how true the rumor sounds from the surprise factor by plugging in the surprise factor into the inverse of the Information of an event. We could have chosen any logbase, but $\log_2$ makes the most intuitive sense, because when the surprise factor is 1, the probability of the rumor being true is .50. As the SF increases, the truth probability decreases. It is important to emphasize that this probability is only defined based on how true the rumor sounds to the population in general. Since we assume the truth of the rumor is completely ambiguous to the people in the simulation, the only way a person can deduce whether a rumor is true is by how surprising it sounds. A good way to understand this is by imagining if you were talking to a friend on the phone and they tell you that they won the lottery. Since that is extremely unlikely you will be less inclined to believe your friend based on no evidence. However, if you see your friend a few days later driving down the street in an expensive sports car, then the possibility of them having won the lottery goes up significantly. The truth probability in our model is based on the phone call, not when you see the car.

$$P_{\text{true}} = 2^{-SF}$$
$$P_{\text{false}} = 1 - 2^{-SF}$$

Final Probability Parameters

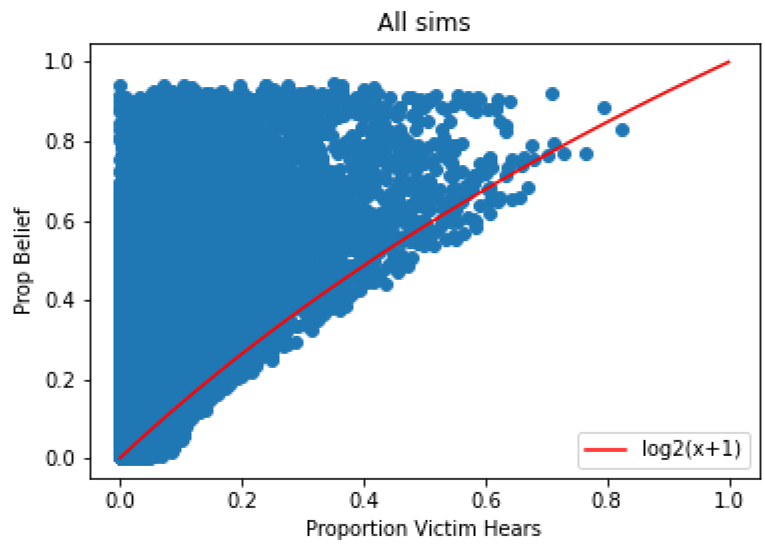
Combining the two above probabilities we have our final probability parameters β and δ. β being the probability that a person believes someone who is telling them that the rumor is true, and δ being the probability that a person does believe someone who is telling them that the rumor isn't true. The below probabilities require that the trust probability and the truth probability are independent. We know this is true because the truth and false probabilities are constant and thus they cannot depend on the trust probability. Since β and δ only occur when two people interact, they must be based on information that is at hand at the time of the interaction.

$$\beta = P(\text{trust}_{m,n} \wedge \text{true}) = P_{m,n}trust * P_{\text{true}}$$
$$\delta = P(\text{trust}_{m,n} \wedge \text{false}) = P_{m,n}trust * P_{\text{false}}$$
$$\lambda = \text{Stifle Probability}$$

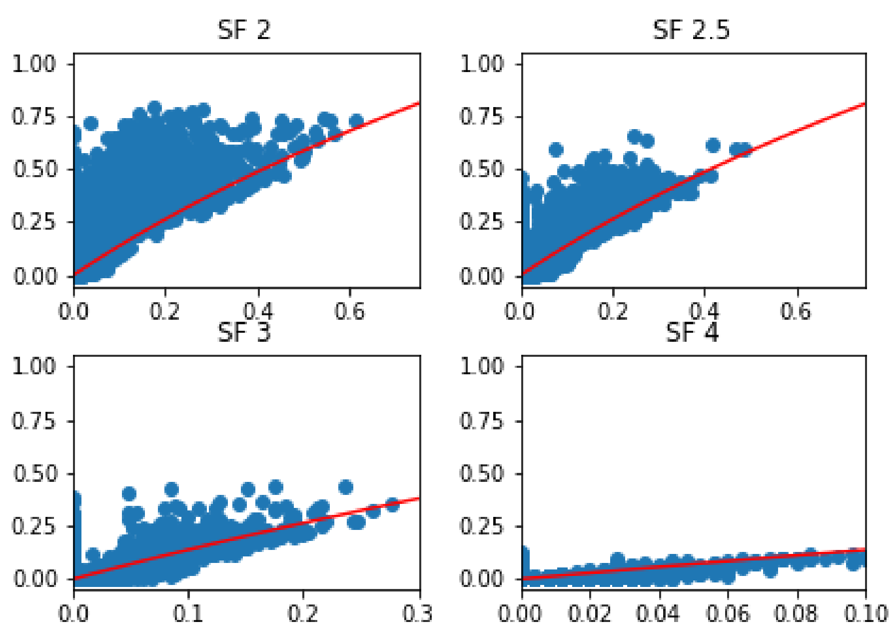
Tests and Results

We tested the simulation by running it over and over again for multiple parameters. We changed population size, the surprise factor, and stifle probability. We tested population sizes [100, 250, 500, 1000] people. We tested the surprise factor for [1, 1.5, 2, 2.5, 3, 5]. We tested the stifle probabilities for [.15, .3, .45, .6]. For each unique combination, the simulation was run 250 times producing 19,200 total data points. In early testing it was found that 20 steps was enough for a rumor to take its course and top out at a certain point. Early data exploration resulted in extremely variable results when testing on the data set in its entirety, when grouped the data into data sets on population size, the data behaved similarly, indicating that the model scales well to larger population sizes.

The reputation of the victim and the reputation of the bully had an almost zero correlation with the proportion of believers at the end of the simulation. In fact the only random variable that we looked at that had a strong correlation was the proportion of people who believed the rumor at the time the victim first heard the rumor. We will call this Victim Proportion (VP). The scatter plot of the VP and the proportion of believers among all plots was completely upper triangular with no below a well defined hypotenuse but completely random above. The hypotenuse of the triangle is fit almost perfectly by the line $y = log_2(x + 1)$. I have included the plot below.



At first glance it seems that the distribution of points is closely related to the true probabilities, which in turn is determined by the surprise factor. When we group the data by the different shock factors we tested we found an interesting trend.



As we can see in the figure above, as the SF goes up the proportion of believers goes down, that was pretty expected. The fact that the percentage of believers that the victim hears the rumor at goes down is also expected as when it is harder to get people to believe the rumor, as ignorants don't believe the rumor and thus don't become spreaders and thus can never turn into stiflers. The fact that the variability and proportion of believers over all population sizes goes down as the surprise factor goes up is due to the fact the 2^{-x} term in β and δ Gets very small and causes the acceptance $\beta \rightarrow 0$ and $\delta \rightarrow 1$ and thus the probability of people spreading the fact that the rumor is true is very high and the rumor has a much lower probability of being spread.

Conclusion

In trying to reduce the amount of factors in my model in order to better isolate how reputation and surprise factor contributed to the spread of a rumor, the attempt was not a success. The surprise factor ended up affecting the model disproportionately compared to the reputation probability. This could also possibly due to the fact that the reputation parameter was not variable enough, sampled according to a pareto distribution, the majority of the sample is in the bottom 50 percentile which make comparing the reputations as a simple ratio not very productive in affecting the model

Acknowledgements

I'd like to thank my friend Anna cause she's the best.

1. "rumor." *Merriam-Webster.com*. Merriam-Webster, 2021.
2. Pendleton, S.c. (1998), 'Rumor research revisited and expanded', *Language & Communication*, vol. 1. no. 18, pp. 69-86.
3. Knapp, Robert H. “A Psychology of Rumor.” *Public Opinion Quarterly*, vol. 8, no. 1, 1944, p. 22., https://doi.org/10.1086/265665.
4. DiFonzo, N. and Bordia, P., 2007. Rumor, Gossip and Urban Legends. *Diogenes*, 54(1), pp.19-35.
5. Brian E. Weeks, R. Kelly Garrett, Electoral Consequences of Political Rumors: Motivated Reasoning, Candidate Rumors, and Vote Choice during the 2008 U.S. Presidential Election, *International Journal of Public Opinion Research*, Volume 26, Issue 4, Winter 2014, Pages 401–422,
6. Daley, D.J., and Kendal, D.G. 1965 Stochastic rumors, J. Inst. Maths Applics 1, p. 42.
7. Ndii, Meksianis. (2018). Mathematical Models for the Spread of Rumors: A Review. 10.1201/9781315148700-8.
8. MacKay, David J.C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press..