

An Investigation of Bullying in Social Networks

A Thesis

Presented to

The Mathematics and Computer Science Department

Colorado College

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Arts

Yichun Shi

Faculty Advisor: Andrea Bruder

May 2016

An Investigation of Bullying in Social Networks

Yichun Shi

Faculty Advisor: Andrea Bruder

Abstract

This study examined the behavior of a bullying epidemic on different social network structures using ideas from network theory, graph theory, and stochastic epidemic modeling. Three aspects of the bullying epidemic were investigated: the predictors for duration of the epidemic, the impact of different initial conditions on the epidemic, and the impact of different network structures on the epidemic. Overall, the more connected the network and the stronger the connection between each individual, the longer the bullying epidemic last. Introducing a more popular student as the first bully to the population would also lead to a longer duration for the bullying epidemic. These results suggest that teachers could educate students on the negative consequences of bullying, which could weaken the connections between susceptible students and the bully, and thus decrease the impact of the bullying epidemic in the classroom.

1 Introduction

Bullying is defined as a form of aggression in which a bully physically and/or psychologically harasses and abuses the victim repetitively. In a study of over 7,000 sixth through tenth graders in the United States, approximately 25% of the students admitted to bullying schoolmates at least once during the current school term and nearly 10% were engaged in bullying at least once a week [3]. Bullying could be either direct or indirect, and the five most common types of bullying are direct verbal harassment, physical attack, spreading rumors, social exclusion, and cyberbullying [6]. It is worth noting that as the use of social networking sites (SNS) continue increasing among adolescents; cyberbullying has become more notorious than before. In a study of 677 high school students, over half of the students reported being a victim of cyberbullying within the last year [1]. Compared to traditional bullying, cyberbullying could have more long-term effects on the victims.

The negative consequences of bullying do not only impact the victims, but also the bullies themselves. Youth who bully others tend to show higher levels of conduct problems and dislike of school, whereas the victims demonstrate higher level of insecurity, depression, isolation, physical and mental symptoms, and low self-esteem. There is growing evidence showing that the bullying victimization may lead to long-term mental health consequences including sadness, anxiety, depression, self-harm, and suicidal attempts [2]. The bullies risk addiction, injury and criminal convictions in their adulthood. [4, 5].

Last year, Katy Martinez and Professor Andrea Bruder studied the dynamics of the bullying behavior in a population by designing an epidemic model of ordinary differential equations that models the behavior of bullying. The SEBNR bullying model presented by Bruder and Martinez was created by taking inspiration from a model used to analyze the spread of influenza. In this paper, we further applied the theory from their research and studied the behavior of the bullying epidemic on social network structures. We constructed an SEBNR stochastic model by taking inspirations from graph theory, network theory, stochastic modeling, and probability. Furthermore, we examined this stochastic SEBNR bullying model on different social network structures, which were constructed using data collected in a classroom setting at Colorado College, and studied the behavior of the bullying model on various types of network structures.

2 Literature Review

2.1 Epidemic Models

A model is a greatly simplified representation of a real system. Epidemic models are used to gain general understandings of transmission dynamics through a population, to reach general qualitative conclusions, or for real-time use in particular epidemics [11]. In order to construct epidemic models, the compartmentalization of individuals based on their disease status is required. Notice that any of the details of the progression of the

epidemic are neglected while constructing epidemic models. However, the simplification has proved to be successful for a long time [7]. Using these assumptions, we could derive two sets of differential equations that are the basis of almost all mathematical epidemiology models: the deterministic susceptible-infectious-recovered (SIR) model,

$$\begin{cases} \frac{dS}{dt} &= -\frac{\beta IS}{N} \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{cases}$$

and the susceptible-infectious-susceptible (SIS) model

$$\begin{cases} \frac{dS}{dt} &= -\frac{\beta IS}{N} + \gamma I \\ \frac{dI}{dt} &= \frac{\beta IS}{N} - \gamma I \end{cases}$$

Generally, we have three assumptions for both SIR and SIS models. The first assumption is homogeneity of hosts, which means all individuals have the same characteristics. The second one is uniform mixing, which means each individual has a equal probability to contact other individuals in the population. The third assumption is no vital dynamics. In the SIR model, individuals begin as susceptible (S), then move to the infected compartment (I) after being infected, and then finally move to the recovered compartment (R) due to recovery from the infection. Comparing to the SIR model, individuals do not gain immunity to the disease after recovery from infection in the SIS model. Thus there is no recovered class in the SIS model, and recovered individuals move back into the susceptible class (S). In both models, N stands for the total population. In the SIR model, $N(t) = S + I + R$. In the SIS model, $N(t) = S + I$. Parameter β is the contact rate and $\frac{I}{N}$ stands for the proportion of the infected individuals in the population. Therefore, $\frac{\beta IS}{N}$ is the rate at which susceptible individuals encounter infected individuals and become infected. We could name $\lambda = \beta \frac{I}{N}$ where λ is the force of infection, and the SIR model would simplify to

$$\begin{cases} \frac{dS}{dt} &= -\lambda S \\ \frac{dI}{dt} &= -\lambda S - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{cases}$$

γ is the rate of recovery and γI stands for the rate at which infected individuals are removed from the I class and either move into the R class (for the SIR model) or the S class (for the SIS model). The basic reproductive number (R_0) is the mean number of secondary infections caused by a single infected individual in an entirely susceptible population [9]. To compute R_0 , notice $S_0 = N$ when the disease first begins spreading. To begin the epidemic, we need

$$\frac{dI}{dt} > 0$$

Therefore,

$$\begin{aligned} \frac{\beta IS_0}{N} - \gamma I &> 0 \\ \beta I - \gamma I &> 0 \end{aligned}$$

Thus,

$$R_0 = \frac{\beta}{\gamma}$$

In general, if $R_0 \leq 1$, the introduction of one infectious individual will cause only minor outbreaks in large populations. If $R_0 > 1$, for a large population, major outbreaks will occur with probability $1 - z_\infty$, which is strictly between 0 and 1. z_∞ represents the probability of a minor outbreak within a very large population. The larger R_0 is, the less likely the outbreak will be a minor one [9]. Generally, we have two assumptions for both SIR and SIS models. The first assumption is homogeneity of hosts, which means all individuals have the same characteristics. The second one is uniform mixing, which means each individual has a equal probability to contact other individuals in the population.

2.2 Graph Theory

Graph theory is an important mathematical tool in a variety of subjects such as operational research, electrical engineering, and architecture. A graph could be considered as an ordered pair $G = (V(G), E(G))$, where $V(G)$ is a nonempty finite set of nodes (or vertices), and $E(G)$ a set of distinct unordered pairs of distinct elements of $V(G)$ called edges. A pair $e = \{v, w\}$ is an element of $E(G)$ if the vertices v and w are connected. Thus we could write vw as an edge. For example, in Figure 2, the graph $G_1 = (V(G_1), E(G_1))$ where $V(G_1) = \{1, 2, 3, 4, 5, 6\}$ is the set of nodes and $E(G_1) = \{12, 23, 26, 56, 35, 34, 45\}$ is the set of edges. We could also assign numerical value to each edge in a graph to represent the strength of the connection between each pair of vertices. These numerical values, named weights, are usually positive, real numbers. The degree of a vertex v in a graph stands for the number of edges incident to vertex v . For example, vertex 2 in G_1 has a degree of 3.

There are various categories of graphs. The most common type is the simple graph, in which a given pair of vertices can be connected by at most one edge. For example, both graph G_1 and graph G_2 in Figure 1 and Figure 2 are simple graphs. If there are multiple edges joining one pair of vertices, the graph is a multigraph. If we not only allow multiple edges joining one pair of nodes, but also allow the nodes to join themselves by edges, which creates a self-loop, we call this type of graph a pseudograph. For the interest of this paper, we will only study the simple graphs.

In a given graph, we say two vertices v, w are adjacent if they are connected by an edge vw . The total number of vertices w that are adjacent to vertex v is called the degree of v and is denoted by k_v . We denote the mean degree of a graph as $\langle k \rangle$. For a graph with N vertices, we define

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_v = \frac{k_1 + k_2 + \dots + k_N}{N}$$

which implies the following formula

$$\langle k \rangle = \frac{2 |E(G)|}{N}$$

For example, in graph G_1 , $k_3 = 3$ and $\langle k \rangle = \frac{2|7|}{6} = \frac{2+3+2+3+3+1}{6} = \frac{14}{6}$.

A graph could also be represented by a matrix. For a graph G with vertices labeled $\{1, 2, 3, \dots, n\}$, its adjacency matrix A is the $n \times n$ matrix whose ij -th entry is the number of edges joining vertex i and j . If graph G has edges labeled as $\{1, 2, 3, \dots, m\}$, its incidence matrix M is the $n \times m$ matrix whose ij -th entry is 1 if vertex i is incidence to edge j , and 0 otherwise [8]. For example, the labeled graph G_1 has the following adjacency and incidence matrices:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix} M = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

A path in a graph G is a sequence of vertices $P = (i_1, i_2, i_3, \dots, i_m)$ such that each of the pairs $\{i_1, i_2\}, \{i_2, i_3\}, \dots, \{i_{m-1}, i_m\}$ is an edge of G . For example, for the sequence $(1, 2, 6, 5)$ is a path while the sequence $(1, 2, 5, 6)$ is not a path in graph G_1 . The length of path $P = (i_1, i_2, i_3, \dots, i_m)$ is $l = m - 1$. A graph G is connected if for every pair of vertices, there is a path between them [9, 10]. For example, Graph G_1 in Figure 1 is connected while G_2 in Figure 2 is not connected since vertex 1 cannot be reached by a path from vertex 2.

3 Methods

3.1 Network Theory

The studies of network structures mainly concentrate in two fields: social science and graph theory. In social science, the use of network analysis helps to understand the spread of new ideas and innovations in a society.

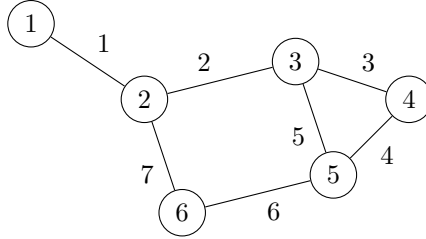


Figure 1: Graph G_1

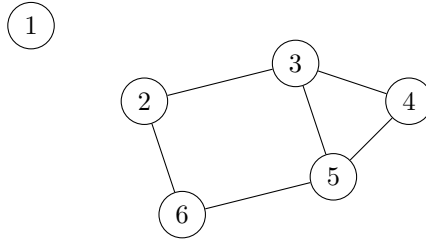


Figure 2: Graph G_2

Furthermore, social dynamics could usually be explained by analyzing the social networks that are behind them. The study of graph theory provides useful tools to quantify and describe networks [7].

Multiple types of networks have been studied for epidemiology purposes. The most popular ones are random networks, lattices, small world, spatial, and scale-free. Each of these networks distributes individuals differently in space and forms connections between individuals in different fashions. In order to construct a model that could largely capture the dynamics of social networks in real world, the scale-free network is the most ideal out of these five types. Comparing to other types of network, the scale-free network displays preferential attachment, which mimics the nature of certain individuals who have significantly more connections or friends in comparison to others. Other types of networks such as lattices show little variation in neighbor sizes. To construct a scale-free network, a new node is added into the existing network at each time step. The new individuals will exhibit preferential likelihood, which means they are more likely to connect with nodes that are adjacent to more vertices in the network. Thus, the degree of nodes displays power-law distribution in the scale-free network. The popular individuals in scale-free networks are called “hubs”. The major hubs are connected with smaller hubs, and the smaller hubs are connected to individuals with even less connections. This process mimics the formation of social networks since people tend to make friends with more popular individuals. Because of the heterogeneity in the degree of each individual, it is interesting to investigate the spread of epidemics on the scale-free network. For the popular individuals in the network, there are mainly two consequences in terms of the spread of epidemics. First, they are at greater risk of infection. Second, they could transmit the disease to many others. Studies has shown that infection is concentrated among individuals with the highest degree [7].

For epidemic models such as SIR and SIS, we assume homogeneity of hosts and uniform mixing. However, in the scale-free network, there exists extreme heterogeneity of hosts due to preferential attachment. Therefore, it is also important for us to look at fully connected networks while studying the spread of epidemics on a network. We denote K_n for a complete graph with n vertices. K_n contains $\binom{n}{2} = \frac{n(n-1)}{2}$ edges. In the complete graph, every pair of hosts (i, j) is connected by a unique edge, which means each individual has a equal probability in contacting the other individuals. Thus, uniform mixing is presented in the complete graph. Since there is no preferential attachment, each individual in the network has the same characteristics and therefore the assumption of homogeneity of hosts holds[9].

3.2 Bullying Model

Kaityn Martinez and Professor Andrea Bruder proposed a model for bullying in schools by considering bullying as an epidemic. In this model, the total population is categorized into five classes: susceptible individuals (S), exposed individuals (E), bullies (B), non-bullies (N), and recovered individuals (R). The

base model for this system is shown in Figure 3. Notice that in this bullying model, individuals in the E compartment are considered as the victims of bullying. In this paper, we derived and studied the stochastic bullying model using informations from this deterministic model. In the stochastic model, individuals in compartment E are considered as the witnesses of the bullying behavior, but not necessary the victims of bullying. The explanation for changing the definition for the E compartment is given in the Model and Methods section of this paper. The deterministic model is constructed using the following system of equations:

$$\begin{cases} P(t) = S + E + B + N + R \\ \frac{dS}{dt} = -\beta SB + dR - cS \\ \frac{dE}{dt} = \beta SB - kE \\ \frac{dB}{dt} = pkE + cS - \alpha B \\ \frac{dN}{dt} = (1-p)kE - \eta N \\ \frac{dR}{dt} = \eta N + \alpha B - dR \end{cases}$$

β stands for the rate of infection, d stands for the rate of recovered children losing their immunity to bullying, c is the proportion of susceptible children becoming bullies without being bullied, k is the likelihood of remaining in the exposed class, p is the probability of becoming a bully after being a victim of bullying, η is the recovery rate for non-bullies, and α is the recovery rate for bullies.

In this model, uniform mixing and homogeneity of hosts are assumed. There are three classes that are categorized as the diseased populations: $E(t)$, $B(t)$, and $N(t)$. However, only $B(t)$, which is the bully class, is infectious. Both bullies and non-bullies recover and then move into the same compartment for recovered individuals. However, the recovery rate is different for these two classes. Due to the social dynamics in real life, immunity for bullying does not last for the recovered class in this model. The basic reproduction number for this model is

$$R_0 = \frac{\beta p}{\alpha}$$

By testing the value of R_0 , Martinez concluded that when $R_0 < 1$, the model results in a disease free equilibrium, which means the disease will die out. If $R_0 > 1$, the model results in an endemic equilibrium, which means the disease will remain permanently endemic in the population [12].

Table 1: Parameters and their Units

Parameter	Meaning	Units
β	rate of infection	$\frac{1}{\#of\ individuals \times time}$
c	proportion of susceptible population that spontaneously become bullies	$\frac{1}{time}$
p	probability of becoming a bully after being bullied	no units
k	$\frac{1}{k}$ is the amount of time spent exposed to bullying	$\frac{1}{time}$
η	$\frac{1}{\eta}$ is the amount of time spent as a non-bully	$\frac{1}{time}$
α	$\frac{1}{\alpha}$ is the amount of time spent as a bully	$\frac{1}{time}$
d	proportion of children that lose their immunity to bullying	$\frac{1}{time}$

3.3 Stochastic Modeling and Probabilities

Two types of models are usually used for epidemic modeling: the deterministic model and the stochastic model. Although the deterministic model is widely used for epidemic modeling, there are several limitations within the model. First, the outputs from the deterministic model depend completely on set parameter values and initial conditions. For example, in the deterministic SIR model, we can determine the characteristics of the epidemic by the value of R_0 , which is completely based on the value of parameters β (rate of infection), and γ (rate of recovery). Thus, in deterministic modeling, the potential differences in the outputs that are caused by random changes in the parameter values are completely neglected. Second, in deterministic models, we have to assume that both the number of infected individuals and the number of possible contacts with susceptible individuals is large. Therefore, the deterministic model does not perform well for modeling the beginning and the end of an epidemic, where the number of infected individuals is relatively small. The deterministic model is also not ideal for modeling epidemics on a network where the number of available

susceptible individuals is small. Third, the approximation of individuals is continuous in deterministic models, which could potentially cause mistakes. For example, Murray et al. [15] purposed a deterministic model to predict the dynamics of rabies among foxes in England. Their model predicted that the number of infected foxes would rapidly increase until the number of susceptible foxes was too low and then the disease would die out. However in real life, it turned out that the disease never completely die out and was able to regenerate from arbitrarily small amounts of residual infected individuals. Thus, the stochastic modeling would be a better approach to predict the epidemics for the above scenarios [11][14].

To construct a stochastic model for analyzing the spread of a disease in a population, we first need to use binomial models to estimate the probability for each individual to move from one class to another. To construct a simple binomial model, let p be the probability of transmitting the disease during a contact between a susceptible individual and an infected individual. Denote $q = 1 - p$ as the probability of a susceptible individual not being infected after contact with an infected host. q could also be named as the escaping probability. Notice in the binomial model, each contact is assumed to be independent from each other, which indicates that the previous contacts will not affect the successive contacts during the epidemic. Let n be the number of contacts, then the probability of a susceptible individual escaping infection after n contacts is $q^n = (1 - p)^n$. Thus, the probability of a susceptible individual getting infected after n contacts is $1 - q^n = 1 - (1 - p)^n$.

With the knowledge of the simple binomial model, we can now construct a chain binomial model. A chain binomial model is derived from a simple binomial model by assuming the infection spreads from individual to individual in a population in discrete time, producing chains of infections governed by the binomial probability distribution. In this section we will study the Reed-Frost Model, which is a classic example of the chain binomial model. In the Reed-Frost Model, we assume individuals pass through three states: susceptible (S), infected (I), and recovered (R). When a susceptible individual is infected via contact with an infected individual, he/she will become infectious for only the next time period, and then gain permanent immunity. There is no vital dynamics in the population. Given $S_t = s_t$ and $I_t = i_t$ at the previous time step, we write the transition probability of getting $I_{t+1} = i_{t+1}$ new infected individuals at time $t + 1$ as

$$Pr(I_{t+1} = i_{t+1} | S_t = s_t, I_t = i_t) = \binom{s_t}{i_{t+1}} (1 - q^{i_t})^{i_{t+1}} q^{i_t(s_t - i_{t+1})}, s_t \geq i_{t+1} \quad (1)$$

where

$$\binom{s_t}{i_{t+1}} = \frac{s_t!}{i_{t+1}!(s_t - i_{t+1})!}$$

is the number of ways to choose the supposed i_{t+1} new cases out of the s_t susceptible individuals. The number of susceptible, and recovered individuals at time $t + 1$ are

$$S_{t+1} = S_t - I_{t+1} \quad (2)$$

$$R_{t+1} = R_t + I_t = \sum_{n=0}^t I_n \quad (3)$$

The probability for the chain $\{i_0, i_1, i_2, \dots, i_T\}$ is thus given by

$$\begin{aligned} Pr(I_1 = i_1 | S_0 = s_0, I_0 = i_0) Pr(I_2 = i_2 | S_1 = s_1, I_1 = i_1) \dots Pr(I_T = i_T | S_{T-1} = s_{T-1}, I_{T-1} = i_{T-1}) \\ = \prod_{t=0}^{T-1} \binom{s_t}{i_{t+1}} (1 - q^{i_t})^{i_{t+1}} q^{i_t(s_t - i_{t+1})} \end{aligned} \quad (4)$$

To demonstrate a simple example of the Reed-Frost model, consider a closed population of three individuals. Suppose $I_0 = 1$, $S_0 = 2$. Then at $t = 1$, the infected individual infects one, both, or neither of the two susceptible individuals. From Equation 1, we obtain

$$Pr(I_1 = 1) = \binom{2}{1} (1 - q^1)^1 q^{1(2-1)} = 2(1 - q)q = 2pq$$

$$Pr(I_1 = 2) = \binom{2}{2} (1 - q^1)^2 q^{1(2-2)} = (1 - q)^2 = p^2$$

$$Pr(I_1 = 0) = \binom{2}{0}(1 - q^1)^0 q^{1(2-0)} = q^2$$

Now, we can form four possible chains for the numbers of infected individuals at $t = 0, 1, 2$. The first one is $\{1, 1, 1\}$ and the chain probability for this chain is

$$Pr(I_1 = 1)Pr(I_2 = 1) = (2pq) \binom{1}{1}(1 - q^1)^1 q^{1(1-1)} = 2p^2q$$

The second one is $\{1, 1, 0\}$ and the chain probability is

$$Pr(I_1 = 1)Pr(I_2 = 0) = (2pq) \binom{1}{0}(1 - q^1)^0 q^{1(1-0)} = 2pq^2$$

The third one is $\{1, 2, 0\}$ and the chain probability is

$$Pr(I_1 = 2)Pr(I_2 = 0) = (p^2) \binom{0}{0}(1 - q^2)^0 q^{2(0-0)} = p^2$$

The fourth one is $\{1, 0, 0\}$ and the chain probability is

$$Pr(I_1 = 0)Pr(I_2 = 0) = (q^2) \binom{1}{0}(1 - q^0)^0 q^{0(1-0)} = q^2$$

In this paper, the movement of individuals between each compartment of the model was determined by selecting random numbers from either a Bernoulli distribution (also known as the coin toss distribution) or a multinomial distribution. The Bernoulli distribution could be interpreted as a binomial distribution where the probability of one success from a trial is

$$Pr(X = 1) = \binom{1}{1}p^1(1 - p)^{1-1} = p$$

For the multinomial distribution, suppose there could be k different outcomes. Denote the number of each outcome i ($i = 1, 2, \dots, k$) and let p_i be the probability for each outcome, then for non-negative integers x_1, \dots, x_k ,

$$Pr(X_1 = x_1 \text{ and } \dots \text{ and } X_k = x_k) = \begin{cases} \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}, & \text{when } \sum_{i=1}^k x_i = n \\ 0, & \text{otherwise} \end{cases}$$

4 Model and Methods

4.1 Constructing Social Network Structures

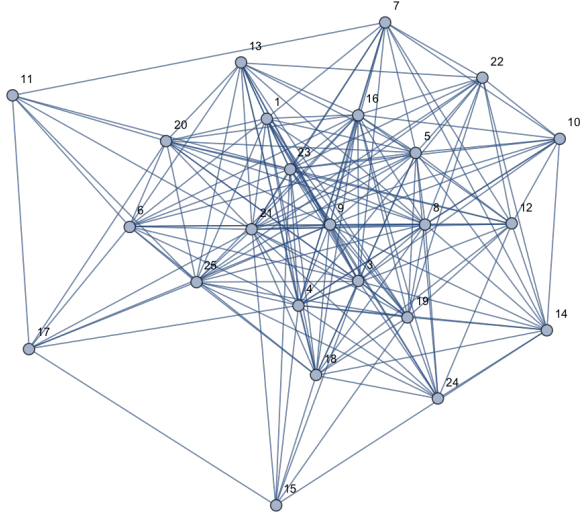
Three different social network structures are studied in this paper: the complete graph K_{24} (Figure 3.c), the friendship graph F_1 where each edge has the same weight (Figure 3.a and 3.b), and the friendship graph with different edge weight F_2 (Figure 3.a and 3.b). All three network structures were constructed using Mathematica, which is a mathematical computation software.

We were interested in studying the 24-node complete graph K_{24} because each node in the graph is connected to all other nodes in the graph and the weights of the edges between each pair of nodes is equal. These two characteristics of K_{24} are considered equivalent to the two model assumptions for standard deterministic epidemic models: uniform mixing and homogeneity of hosts. The complete graph K_{24} was constructed using built-in commands for creating complete graphs in Mathematica. K_{24} contains 24 vertices and 276 edges. Every edge in K_{24} has a weight of 1.

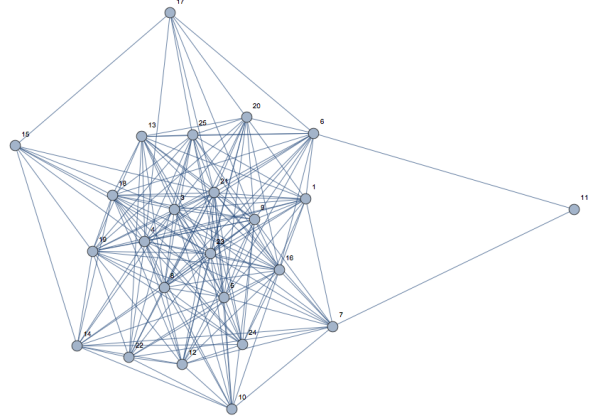
The information used to construct F_1 and F_2 were obtained from responses to a 25-question survey, which is attached in the Appendix section of this paper. 25 students from the MA325 Graph Theory class completed the survey. In the survey, students were asked to rate their relationship with every other student in class on a scale of 1 to 5. A rating of 1 means no interaction between the two individuals, 2 means the two people know each other but barely have conversed, 3 stands for frequent conversations but never spending time

together, 4 stands for occasionally spending time together, and 5 means spending time together frequently. In F_1 and F_2 , an edge was added between two vertices if one of the two students provided a rating 2 or above for their relationship. In F_1 , every edge was assigned with a weight of 1. In F_2 , the weight of each edge was determined based on the ratings of the relationships between each pair of students. The corresponding edge weight for each pair of ratings is presented in Table 1. F_1 and F_2 are shown in Figure 3.a. Both F_1 and F_2 contain 24 vertices and 186 edges. Data from Student 2 was omitted when constructing F_1 and F_2 due to the inconsistency between the ratings he provided and others provided about him. In F_1 and F_2 , the most popular individual is Student 21, with 22 connections in total. The least popular individual is Student 11, which only has 7 connections in the network. On average, each student in the network had 7.75 connections. An 11-person clique (a fully connected subgraph) was also observed in F_1 and F_2 .

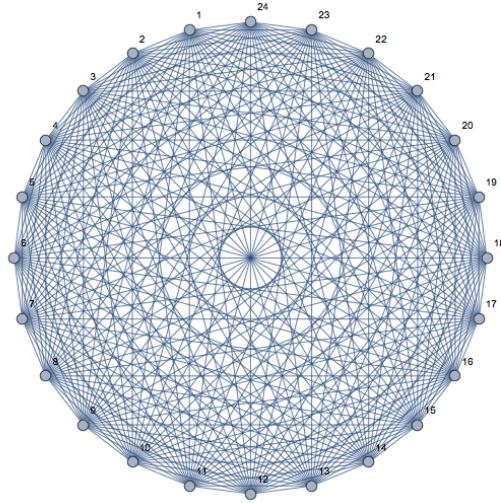
A modified version of F_1 and F_2 (Figure 3.b) was also created by removing 5 connections from Student 11. Both the modified F_1 and F_2 contain 24 vertices and 181 edges. The reason for the modification is explained in Section 4.2. In the modified F_1 and F_2 , each student has 7.54 connections.



(a) Friendship graph F_1 and F_2 . Both F_1 and F_2 have 24 nodes and 186 edges. In F_1 , each edge has a weight of 1. In F_2 , each edge was assigned with a weight based on the friendship survey rating. The edge weights in F_2 ranges from 0 to 1.



(b) The modified friendship graphs F_1 and F_2 . 5 connections of Student 11 were removed. Both the modified F_1 and F_2 contain 24 nodes and 181 edges. In the modified F_1 , each edge has a weight of 1. In the modified F_2 , each edge was assigned with a weight based on the friendship survey rating. The edge weights in the modified F_2 ranges from 0 to 1.



(c) The complete graph K_{24} . K_{24} contains 24 nodes and 276 edges. Each edge was assigned with a weight of 1.

Figure 3: Network structures F_1 , F_2 , modified F_1 , modified F_2 , and K_{24}

Table 2: Corresponding edge weights in F_2 for each pair of rating from the friendship survey

Ratings	1	2	3	4	5
1	0	0.125	0.25	0.375	0.5
2	0.125	0.25	0.375	0.5	0.625
3	0.25	0.375	0.5	0.625	0.75
4	0.375	0.5	0.625	0.75	0.875
5	0.5	0.625	0.75	0.875	1

4.2 Implementing the Bullying Model on Network Structures

The bullying model was implemented on the three network structures F_1 , F_2 , and K_{24} using Mathematica. The five compartments in this epidemic model are Susceptible (S), Exposed (E), Bully (B), Non-bully (N), and Recovered (R). Again, in this model, the movements of individuals between each compartment were determined by selecting random numbers from either a Bernoulli distribution or a multinomial distribution.

Recall that for a individual in the S compartment, he/she will have the probability of becoming exposed to bullying if he/she is connected to a bully. The susceptible individual could either move forward to the E compartment, or remain in the S compartment for the next time step. Thus at each time step, the movement for individuals in the S compartment is determined by randomly selecting either a 0 or 1 from a Bernoulli distribution, where 0 represents remaining in the S compartment and 1 represents moving to the E compartment. For each individual in the S compartment, denote the probability of moving to the E compartment as p_1 and the probability for staying in the S compartment as $1 - p_1$. For K_{24} and F_1 , $p_1 = 1 - p_1 = 0.5$. For F_2 , p_5 equals the weight of the edge between the susceptible individual and the bully. This assumption is supported by Albert Bandura's Social Learning Theory, which indicates that the closer the relationship between the child and the model, the more likely that the child will observe and perform behaviors that the model demonstrates. Thus, it is safe to assume that the closer the relationship between the bully and the susceptible individual, the more likely that the susceptible individual will learn and perform bullying behaviors by observing the bully. [17]

The movement of individuals in the E compartment at each time step is determined by drawing a random number from a multinomial distribution since the individual could either move to the B compartment, N compartment, or remain at the E compartment at each time step. The probability for each of these outcomes are p_2 , p_3 , and p_4 . For all three network structures, $p_2 = 0.2$, $p_3 = 0.3$, and $p_4 = 0.5$. Similar to individuals in the S compartment, the movement of individuals in compartment B, N, and R are also determined by drawing either a 0 or 1 from the Bernoulli distribution at each time step. Individuals in the B and N compartments could either move forward to the R compartment, or remain as a bully or non-bully for this time step. Recovered individuals (R) could either become susceptible again or remain recovered. For individuals in B, N, or R, the probability of moving to the next epidemic compartment is p_5 . The probability for remaining in the same compartment is $1 - p_5$. For F_1 , F_2 and K_{24} , $p_5 = 1 - p_5 = 0.5$.

90 trials of simulations were conducted on F_1 and F_2 , and 30 trials were conducted on K_{24} . Each trial lasted 100 days. 1 initial bully and 23 susceptible individuals were introduced to the population at day 1 of each trial. In order to investigate the impact of different initial conditions on the epidemics, 3 different individuals were introduced as the initial bully for simulations on F_1 and F_2 . Student 21, who was the most popular individual in both F_1 and F_2 (degree 22), was introduced as the initial bully for the first 30 trials of simulation on F_1 and F_2 . Student 12 (degree 15) was introduced as the initial bully for the second 30 trials on F_1 and F_2 . After removing 5 edges that were connected to Student 11, the modified Student 11 (previously degree 7, now degree 2) was introduced as the initial bully for the last 30 trials of simulations on F_1 and F_2 . We removed 5 connections from Student 11 because we were interested in investigating whether the initial bully having a very low number of connections would significantly change the progress of the epidemic. For the complete graph K_{24} , since each individual had an identical number of connections (23 connections), individual 12 was set to be the initial bully for all 30 trials.

We were interested in investigating three different questions. First, on the same social network structure, how would the different initial conditions affect the behavior of the bullying epidemic? Second, given the same initial condition, how would the epidemic evolve differently on different network structures? Last, are there any significant variables that could predict how long the bullying epidemic would last? Four variables were collected from each simulation: the Number of Days to Eliminate Bullying, the Maximum Number of Bullies in a Day, the Number of Exposed Individuals on Day 2, and the Weighted Average of Bullies per Day. The units for these variables are presented in Table 2. The Number of Days to Eliminate Bullying is determined by the total number of days it took to obtain no exposed individual or bully in the network during each trial. The Maximum Number of Bullies in a Day was determined by the biggest possible number of bullies observed in a day during each trial. The Number of Exposed Individuals on Day 2 was studied because of its similarity to the basic reproduction number R_0 from standard deterministic epidemic models. This variable accounts for the number of exposed individuals observed on the second of each trail. The

Weighted Average of Bullies per Day for each trial was calculated using the following formula:

$$\begin{aligned} \text{weighted average} &= \frac{\sum_{i=1}^{100} \text{number of bullies on day } i}{100 \text{ days}} \\ &= \frac{0 \cdot (\text{number of days with 0 bullies}) + \dots + 8 \cdot (\text{number of days with 8 bullies})}{100 \text{ days}} \end{aligned} \quad (5)$$

Table 3: Parameters and their Units

Parameter	Meaning	Units
Number of days to Eliminate bullying	The duration of the bullying epidemic in each trial.	<i>day</i>
Max Number of Bullies in a Day	The maximum number of bullies that was observed in a single day during each trial	<i>bully</i>
Number of Exposed on Day 2	The number of exposed individuals on the second day of each trail	<i>bully</i>
Weighted Average	The average number of bullies on a day for each trial	$\frac{\text{bullies}}{\text{days}}$

5 Results

5.1 K_{24}

For 2 out of the 30 trials conducted on K_{24} , the bullying epidemic did not terminate during the 100-day period. By setting Number of Days to Eliminate Bullying = 101 for these two trials, the mean of Number of Days to Eliminate bullying for the 30 trials on K_{24} was 34.94 days. By removing these two trials, the mean for Number of Days to Eliminate Bullying was 30.32 days. The average for the Maximum Number of Bullies in a Day was 5.13 bullies. The average for the Number of Exposed Individuals on Day 2 was 11.8 bullies. The mean of the Weighted Average was 0.706 bullies/day.

5.2 F_1

First, Student 21 (degree 22) was introduced as the initial bully and 30 simulations were conducted. For 2 out of the 30 trials, the bullying epidemic did not terminate during the 100-day period. By setting Number of Days to Eliminate Bullying = 101 for these two trials, the mean for Number of Days to Eliminate Bullying for the 30 trials on F_1 was 32.23 days. By removing these two trials, the mean for Number of Days to Eliminate Bullying was 27.32 days. The average for Maximum Number of Bullies in a Day was 4.83 bullies. The average for Number of Exposed on Day 2 was 11.03 bullies. The mean for the Weighted Average was 0.62 bullies/day.

By introducing Student 12 (degree 15) as the initial bully, the bullying epidemic ended during the 100-day period for all 30 trials. On average, it took 21.7 days to completely terminate bullying in the network. The mean for the Maximum Number of Bullies in a Day was 4.23. The mean for the Number of Exposed Individual on Day 2 was 7.43. The mean for the Weighted Average was 0.484 bullies/day.

Student 11 with 5 connections removed (degree 2) was introduced as the initial bully for the last 30 trials conducted on F_1 . The bullying epidemic was successfully terminated during the 100-day period for all 30 trials. On average, it took 9.43 days to eliminate bullying. The mean for the Maximum Number of Bullies in a Day was 2.0 bullies. On average, the Number of Exposed Individuals on Day 2 was 0.867, which was significantly less than the means from the previous 60 trials. However, this is not surprising since after

removing 5 connections, Student 11 was only connected to 2 students, which indicates that he/she could expose 0, 1, or 2 students to bullying. The mean for the Weighted Average was 0.138 bullies/day.

5.3 F_2

For the first 30 trials conducted on F_2 , Student 21 (degree 22) introduced as the initial bully. On average, it took 21.33 days to completely eliminate bullying for the first 30 trials on F_2 . The bullying epidemic was successfully eliminated for all 30 trials. The mean for Maximum Number of Bullies in a Day was 4.13. The average Number of Exposed Individuals on day 2 was 10.5, and the mean for Weighted Average was 0.617 bullies/day.

For the second 30 trials conducted on F_2 , Student 12 (degree 15) was introduced as the first bully to the network. The bullying epidemic was successfully terminated during the 100-day period for all 30 trials. The Mean for Number of Days to Eliminate Bullying was 21.9 days. The mean for Maximum Number of Bullies in a Day was 3.6. On average, 5.1 students became exposed to bullying on the second day of each trial. The mean for the Weighted Average was 0.361 bullies/day.

With Student 11 (degree 2) introduced as the initial bully, bullying was successfully eliminated from F_2 for all 30 trials. On average, it took 10.2 days to end bullying in the network. The average for Maximum Number of Bullies in a Day is 1.967. The mean for Number of Exposed Individuals on Day 2 was 0.633. The mean for the Weighted Average was 0.147 bullies/day.

Table 4: Results from introducing Student 21 (22 connections) as the first bully

Variable	Network Structure	Mean	Standard Deviation
Number of days to terminate bullying	F_1	32.23	27.20
	F_2	21.33	19.48
	K_{24}	34.97	27.21
Maximum number of bullies in a day	F_1	4.83	1.64
	F_2	4.13	1.36
	K_{24}	5.13	1.48
Number of exposed on day 2	F_1	11.03	2.30
	F_2	10.50	1.96
	K_{24}	11.8	2.16
Weighted Average	F_1	0.62	0.53
	F_2	0.37	0.38
	K_{24}	0.71	0.55

Table 5: Results from introducing Student 12 (15 connections) as the first bully

Variable	Network Structure	Mean	Standard Deviation
Number of days to terminate bullying	F_1	29.70	18.51
	F_2	21.90	15.97
Maximum number of bullies in a day	F_1	4.23	1.46
	F_2	3.60	1.81
Number of exposed on day 2	F_1	7.43	1.76
	F_2	5.10	1.30
Weighted Average	F_1	0.48	0.35
	F_2	0.36	0.31

Table 6: Results from introducing the modified Student 11 (2 connections) as the first bully

Variable	Network Structure	Mean	Standard Deviation
Number of days to terminate bullying	F_1	9.43	12.56
	F_2	10.20	14.00
Maximum number of bullies in a day	F_1	2.00	1.60
	F_2	1.97	1.43
Number of exposed on day 2	F_1	0.87	0.57
	F_2	0.63	0.62
Weighted Average	F_1	0.14	0.23
	F_2	0.15	0.24

6 Discussions

6.1 The Behavior of the Bullying Epidemic on Different Social Network Structures

Compare to the total 180 simulations conducted on F_1 and F_2 , the 30 simulations conducted on the complete graph K_{24} had the largest means for all four variables collected: Number of Days to Eliminate Bullying, Maximum Number of Bullies in a Day, Number of Exposed Individuals on Day 2, and the Weighted Average. Because K_{24} was fully connected (each individual had 23 connections) and each edge was assigned with a weight of 1, bullies in K_{24} could cause more susceptibles' exposure to the bullying behavior. The increased number of exposed individuals could then allowed more movement from the E compartment to the B compartment during each trial. Thus on average, more exposed individuals were expected to be observed on the second day for trials conducted on K_{24} , which explained the higher average for Number of Exposed Individuals on Day 2. The mean for the Maximum Number of Bullies in a Day for trials conducted on K_{24} was also expected to be higher than the mean from trials conducted on F_1 and F_2 . The mean of the Weighted Average was also larger for trials on K_{24} than on F_1 and F_2 . This is due to the longer duration of the bullying epidemic on K_{24} , which implies that there were more days with 1 or more bullies observed in each 100-day trial executed on K_{24} .

Given the initial condition of Student 21 (degree 22) introduced as the first bully, simulations conducted on F_1 had higher means than on F_2 for all four variables. This trend was also observed in trials conducted on F_1 and F_2 where Student 12 (degree 15) was set as the initial bully. These observations were not surprising since every edge in F_1 was assigned with a weight of 1 while edges in F_2 were assigned with weights ranging from 0.125 to 1. In F_1 , each susceptible individual that was connected to a bully faced a probability of 0.5 of moving to the exposed compartment at each time step. In F_2 , this probability equaled to the weight of the edge between the susceptible individual and the bully, and only 74 out of the 186 edges had weights of 0.5 or above. Since over half of the edges in F_2 were assigned with weights less than 0.5, bullies in F_2 could potentially cause less susceptible individuals becoming exposed to bullying, which led to shorter duration of the bullying epidemic. One unusual observation was that by introducing the modified Student 11 (previously degree 7, now 2) as the initial bully, the 30 simulations on F_2 had a slightly higher mean for Number of Days to Eliminate Bullying than trials on F_1 . However, trials on F_1 demonstrated larger means for the three other variables: Maximum Number of Bullies in a day, Number of Exposed Individuals on Day 2, and the Weighted Average of Bullies per day. This observation could be caused by the stochastic nature of the bullying epidemic model. Since only 30 trials were conducted on each network structure using each initial condition, the data gathered from these trials might not accurately describe the total population. In order to better describe the overall population, more simulations should be conducted with each set of initial conditions and network.

6.2 Predictors for the Duration of the Bullying Epidemic

In order to find significant variables that could predict the duration of the bullying epidemic on social networks, three variables (Maximum Number of Bullies in a Day, Number of Exposed Individuals on Day 2, and Weighted Average) were tested using regression analysis where the number of days to eliminate bullying

in each trial was set as the dependent variable. Three regression analyses were conducted for testing each variable. Data sets collected from trials conducted on F_1 , F_2 and K_{24} with Student 12 (degree 15 in F_1 and F_2 , degree 23 in K_{24}) as the initial bully were used for the regression analyses in this section.

First, Maximum Number of Bullies in a day was used as the independent variable. The p-values from the regression analyses using data from F_1 and F_2 were less than 0.05, which indicated that the null hypothesis could be rejected. Thus, in these two regression, the Maximum Number of Bullies in a Day is a statistically significant variable. In the regression analysis using data from K_{24} , the p-value was larger than 0.05 ($p = 0.063$) and thus the null hypothesis could not be rejected. The R-squared values from the analyses using data from F_1 and K_{24} were low (26% and 11%), which implied that only a small amount of the data in the duration of the bullying epidemic could be explained by the linear regression models. In the regression model with data from trials on F_2 , the R-squared value was slightly higher (R-sq = 44.19%). Based on the findings above, we concluded that the maximum number of bullies in a day from each trial was not a good predictor for the duration of the bullying epidemic.

The Number of Exposed individuals on Day 2 was tested for the regression. The p-values from the regression analysis using data from F_2 and K_{24} (p-value = 0.877 and p-value = 0.778) were very large, which strongly suggest that there were no linear relationships between the independent and dependent variables. This suggested that Number of Exposed individuals on Day 2 is not a statistically significant variable. The extremely low R-squared values (0.09% and 0.29%) further confirmed this observation. The p-value = 0.063 from regression using data from F_2 was slightly larger than 0.05, thus the null hypothesis was failed to be rejected. The R-sq = 11.81% suggested that only 11.81% of the data points of days to eliminate bullying could be explained by number of exposed individuals on day 2. Based on the observations above, we determined that there is little or no linear relationship between the duration of the bullying epidemic and the number of exposed individuals on day 2 of each trial.

The Weighted Average was then tested as the independent variable. The p-value for all three regression analyses was less than 0.001. Therefore, the null hypothesis, which has no linear relationship between the dependent and independent variable, could be rejected. Thus the Weighted Average is a statistically significant variable. The R-squared values from all three regression analyses were high (89.62%, 92.22%, and 92.20%), which implied that most of the data points for the duration of the bullying epidemic could be predicted using the linear regression models. Thus, the Weighted Average can nicely describe the duration of the bullying epidemics using the regression formula. This is not surprising since the weighted average of bullies per day for each trial was calculated by summing the bullies on each day and then divide by 100. Thus the longer the epidemic last in a trial, the larger then weighted average would be for that trial. Although there were strong correlations between the weighted average and the duration of the epidemic, the weighted average did not provide useful information to predict the duration of the epidemic.

6.3 Impacts of Different Initial Conditions on the Bullying Epidemic

On F_1 and F_2 , three different initial conditions were examined: Student 21 (degree 22), Student 12 (degree 15), and Student 11 with reduced connections (degree 2). 5 connections were removed from Student 11 since we were interested in investigating whether introducing a highly unconnected individual as the initial bully would significantly decrease the duration of the bullying epidemics and the first wave of exposed individuals caused by the first bully in the network.

On F_1 , the mean for the duration of the bullying epidemic was slightly higher for trials with Student 21 (degree 22) as the initial bully than trials with Student 12 (degree 15) as the first bully. However, for trials on F_1 , the average of the number of exposed individuals on day 2 with Student 12 as the initial bully was significantly smaller than trials with Student 21 as the initial bully. The same patterns was also observed for trials on F_2 . One possible explanation for these observations is that since Student 12 had 7 less connections than Student 21, he could expose less susceptible individuals to his bullying behavior on the second day of each trial. Thus, a significant decrease in the mean of number of exposed individual on day 2 was observed when changing the initial condition from Student 21 to Student 12. However, since on average each student in the network had 7.75 connections, the students that were exposed on the second day by Student 12 were still capable of maintaining the high number of bullies in the network, which would then lead to more exposure to the bullying behavior inside the network throughout each trial. Thus, there was not a significant decrease in the mean of days to eliminate bullying when switching the initial condition

from Student 21 to Student 12. By introducing the modified Student 11 (degree 2) as the initial bully, only 0, 1 or 2 students could be exposed to bullying on the second day. Thus a significant decrease in the mean of number of exposed individual on day 2 was observed for trials on both F_1 and F_2 . Since Student 11 could only exposed a maximum of 2 student on the second day of each trial, there could only be a maximum of 3 bullies on the third day of each trial. This greatly decreased the possibility for the bullying epidemic to persist inside the network. Thus, a significant decrease in the average of days to eliminate bullying was observed for trials on both F_1 and F_2 with Student 11 introduced as the first bully.

For trials on both F_1 and F_2 , a slight drop in the mean of the maximum number of bullies in a single day was observed when changing the initial bully from Student 21 (degree 22) to Student 12 (15 connections). By changing the initial bully to Student 11 (previously degree 7, now degree 2), the average of the maximum number of bullies in a single day dropped significantly for trials both on F_1 and F_2 . Again, since both Student 21 and Student 12 were capable of keeping a high number of exposed individuals on the second day of each trial, the mean of the duration of the epidemic with both initial conditions were quite similar. Therefore, similar mean for the maximum number of bullies in a single day should be observed with both initial conditions. Since Student 11 was not capable of maintaining a high number of students exposed to bullying on the second day, the average duration of the epidemic is much lower and a much smaller number for the maximum number of bullies during each trial should be observed.

7 Conclusions

The stochastic SEBNR bullying model could be used to examine the dynamics of bullying behavior on any social network structures. In this paper, we have shown that the more connected the individuals in the network and the stronger the connection between each individual, the longer the bullying epidemic would persist on the social network structures. On the same network structure, introducing different individuals as the first bully to the network would lead to different behaviors of the bullying epidemic. The more connections the initial bully has, the higher the chance that more students who were susceptible would become exposed to bullying, and thus the longer the duration of the bullying epidemic on the network structure. However, if two different students that are both highly connected within the network structure were introduced as the first bullies, the difference between the duration of the bullying epidemic would be not very significant. If compare simulations where a highly connected student and a highly unconnected student introduced as the first bullies, the duration of the bullying epidemic would be significantly longer for simulations with the highly connected student as the first bully.

What do we learn from these findings? First, these observations provide us suggestions on how to treat bullying in networks such as middle school classrooms. Since bullying would last longer on more connected networks, teachers could reduce the impact of bullying to adolescents by weakening the connections between students. For example, the teacher could educate students on the negative impacts of bullying behavior. After recognizing the negative consequences of bullying, susceptible students might weaken their connections to the bully, which then decrease the impact of the bullying epidemic on the whole network. For students that were already exposed to bullying, the education about bullying could lower their probability of turning into a bully themselves.

In this paper, we only examined one set of probabilities for the stochastic SEBNR model and only 30 simulations were conducted for each initial condition on each network structure. In the future, more sets of parameter values and initial conditions could be examined to develop a deeper understanding of the spread of bullying on network structure. Also, in order to obtain results that could better describe the “truth” or the population, more simulations could be conducted.

8 Acknowledgements

First, I would like to express my sincere gratitude to my thesis advisor, Professor Andrea Bruder, for her continuous support on this research project. Her guidance helped me in all time of research and writing of this thesis. I would also like to thank Professor Tricia Waters and Professor John Watkins for sharing their insights and providing great advice. I’m also grateful to all students in the MA325 Graph Theory class for supporting my thesis project.

References

- [1] Maite Garaigordobil, Vanesa Martinez-Valderrey. Effects of Cyberprogram 2.0 on ?face-to-face? bullying, cyberbullying, and empathy. *Psicothema*, 27(1): 45-51, 2015
- [2] Nazanin Alavi, Nasreen Roberts, Chole Sutton, Nicholas Axas, Leanne Repeyyi. Bullying Victimization (Being Bullied) Among Adolescents Referred for Urgent Psychiatric Consultation: Prevalence and Association With Suicidality. *CanJPsychiatry*, 60(10):427-431, 2015
- [3] Tonja R. Nansel, Mary Overpeck, Ramani S. Pilla, W. June Ruan, Brunce Simons-Morton and Peter Scheidt. Bullying behaviors among US youth: prevalence and association with psychosocial adjustment. *JAMA*, 285(16): 2094-2100, 2001
- [4] Kathleen S. Berger. Update on bullying at school: Science forgotten? *Developmental Review*, 27(1): 90-126, 2007
- [5] Johannes F. Sigurdson, A. M. Undheim, J. L. Wallander, S. Lydersen and A. M. Sund. The long-term effects of being bullied or a bully in adolescence on externalizing and internalizing. *Child and Adolescent Psychiatry and Mental Health*, 9(1): 1-13, 2015
- [6] Jing Wang, Ronald J. Iannotti, Jeremy W. Luk and Tonja R. Nansel. Co-occurrence of victimization from five subtypes of bullying: physical, verbal, social exclusion, spreading rumors, and cyber. *Journal of Pediatric Psychology*, 35(10):1103-1112, 2010
- [7] Matt J. Keeling and Ken T.D. Eames. Networks and epidemic models. *J.R. Soc Interface*, 2(4):295-307, 2005
- [8] Robin J. Wilson. *Introduction to Graph Theory*. Prentice Hall, 1996
- [9] Winfried Just, Hannah Callender and M. Drew LaMar. *Algebraic and Discrete Mathematical Methods for Modern Biology*. Academic Press, 2015.
- [10] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.
- [11] Valerie Isham. *Stochastic Models for Epidemics*. University College London, 2004.
- [12] Kaitlyn Martinez and Andrea Bruder. *Epidemiological Methods for Examining Bullying*. Colorado College, 2015.
- [13] Linda J. S. Allen and Amy M. Burgin. Comparison of deterministic and stochastic SIS and SIR models in discrete time *Mathematical Biosciences*, 163(1):1-13, 2000.
- [14] Jan Pieter Trapman. *On Stochastic Models for the Spread of Infections*, ISBN-13(2006): 978-90.
- [15] J. D. Murray, E. A. Stanley, and D. L. Brown. On the spatial spread of rabies among foxes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 229(1255): 111-150.
- [16] M. Elizabeth Halloran, Ira M. Longini, Jr and Claudio J. Struchiner. *Design and Analysis of Vaccine Studies*. Springer, 2010.
- [17] Albert Bandura. *Social Learning Theory*. (1977): 305-316.

A Appendix

The Friendship Survey.

Name:

Please select the option that best describes your relationship with him/her. Leave the question blank if you are asked to rate the relationship with yourself.

1: I have never met or talked to him/her before. 2: I have introduced myself to him/her but I have barely talked to him/her. 3: I talk to him/her regularly but have never hanged out together before. 4: I talk to him/her regularly and have hanged out together a few times. 5: I am good friend with him/her and we hang out regularly.

1. Describe your relationship with Student 1.
1 2 3 4 5
2. Describe your relationship with Student 2.
1 2 3 4 5
3. Describe your relationship with Student 3.
1 2 3 4 5
4. Describe your relationship with Student 4.
1 2 3 4 5
5. Describe your relationship with Student 5.
1 2 3 4 5
6. Describe your relationship with Student 6.
1 2 3 4 5
7. Describe your relationship with Student 7.
1 2 3 4 5
8. Describe your relationship with Student 8.
1 2 3 4 5
9. Describe your relationship with Student 9.
1 2 3 4 5
10. Describe your relationship with Student 10.
1 2 3 4 5
11. Describe your relationship with Student 11.
1 2 3 4 5
12. Describe your relationship with Student 12.
1 2 3 4 5
13. Describe your relationship with Student 13.
1 2 3 4 5
14. Describe your relationship with Student 14.
1 2 3 4 5
15. Describe your relationship with Student 15.
1 2 3 4 5
16. Describe your relationship with Student 16.
1 2 3 4 5
17. Describe your relationship with Student 17.
1 2 3 4 5
18. Describe your relationship with Student 18.
1 2 3 4 5
19. Describe your relationship with Student 19.
1 2 3 4 5
20. Describe your relationship with Student 20.
1 2 3 4 5
21. Describe your relationship with Student 21.
1 2 3 4 5
22. Describe your relationship with Student 22.
1 2 3 4 5

23. Describe your relationship with Student 23.
1 2 3 4 5
24. Describe your relationship with Student 24.
1 2 3 4 5
25. Describe your relationship with Student 25.
1 2 3 4 5

B Appendix

Listing 1: Example Mathematica code for F_2 with Student 21 as the first bully

```

1 gf = Graph[{1 <-> 3, 1 <-> 4, 1 <-> 5, 1 <-> 6, 1 <-> 7,
2   1 <-> 8, 1 <-> 9, 1 <-> 13, 1 <-> 16, 1 <-> 18,
3   1 <-> 19, 1 <-> 20, 1 <-> 21, 1 <-> 23, 1 <-> 24,
4   1 <-> 25, 3 <-> 4, 3 <-> 5, 3 <-> 6, 3 <-> 7,
5   3 <-> 8, 3 <-> 9, 3 <-> 10, 3 <-> 13, 3 <-> 14,
6   3 <-> 15, 3 <-> 16, 3 <-> 18, 3 <-> 19, 3 <-> 20,
7   3 <-> 21, 3 <-> 23, 3 <-> 24, 3 <-> 25, 4 <-> 6,
8   4 <-> 7, 4 <-> 8, 4 <-> 9, 4 <-> 10, 4 <-> 12,
9   4 <-> 13, 4 <-> 14, 4 <-> 15, 4 <-> 16, 4 <-> 17,
10  4 <-> 18, 4 <-> 19, 4 <-> 21, 4 <-> 22, 4 <-> 23,
11  4 <-> 24, 4 <-> 25, 5 <-> 6, 5 <-> 7, 5 <-> 8,
12  5 <-> 9, 5 <-> 10, 5 <-> 12, 5 <-> 13, 5 <-> 14,
13  5 <-> 16, 5 <-> 18, 5 <-> 19, 5 <-> 20, 5 <-> 21,
14  5 <-> 22, 5 <-> 23, 5 <-> 24, 5 <-> 25, 6 <-> 8,
15  6 <-> 9, 6 <-> 11, 6 <-> 13, 6 <-> 16, 6 <-> 17,
16  6 <-> 18, 6 <-> 20, 6 <-> 21, 6 <-> 23, 6 <-> 25,
17  7 <-> 9, 7 <-> 10, 7 <-> 11, 7 <-> 12, 7 <-> 14,
18  7 <-> 16, 7 <-> 21, 7 <-> 22, 7 <-> 23, 8 <-> 9,
19  8 <-> 10, 8 <-> 12, 8 <-> 13, 8 <-> 14, 8 <-> 16,
20  8 <-> 18, 8 <-> 19, 8 <-> 20, 8 <-> 21, 8 <-> 22,
21  8 <-> 23, 8 <-> 24, 8 <-> 25, 9 <-> 10, 9 <-> 12,
22  9 <-> 13, 9 <-> 14, 9 <-> 16, 9 <-> 17, 9 <-> 18,
23  9 <-> 19, 9 <-> 20, 9 <-> 21, 9 <-> 23, 9 <-> 24,
24  9 <-> 25, 10 <-> 12, 10 <-> 14, 10 <-> 16, 10 <-> 21,
25  10 <-> 22, 10 <-> 23, 11 <-> 17, 11 <-> 20, 11 <-> 21,
26  11 <-> 23, 11 <-> 25, 12 <-> 14, 12 <-> 16, 12 <-> 18,
27  12 <-> 19, 12 <-> 20, 12 <-> 21, 12 <-> 22, 12 <-> 23,
28  12 <-> 24, 13 <-> 16, 13 <-> 18, 13 <-> 19, 13 <-> 20,
29  13 <-> 21, 13 <-> 22, 13 <-> 23, 13 <-> 25, 14 <-> 15,
30  14 <-> 18, 14 <-> 19, 14 <-> 22, 14 <-> 24, 15 <-> 17,
31  15 <-> 18, 15 <-> 19, 15 <-> 21, 15 <-> 23, 16 <-> 18,
32  16 <-> 19, 16 <-> 20, 16 <-> 21, 16 <-> 22, 16 <-> 23,
33  16 <-> 24, 16 <-> 25, 17 <-> 20, 17 <-> 21, 17 <-> 25,
34  18 <-> 19, 18 <-> 21, 18 <-> 23, 18 <-> 24, 18 <-> 25,
35  19 <-> 20, 19 <-> 21, 19 <-> 22, 19 <-> 23, 19 <-> 24,
36  19 <-> 25, 20 <-> 21, 20 <-> 23, 20 <-> 25, 21 <-> 22,
37  21 <-> 23, 21 <-> 24, 21 <-> 25, 22 <-> 23, 22 <-> 25,
38  24 <-> 25
39  },
40  EdgeWeight -> {0.125, 0.25, 0.625, 0.625, 0.125,
41    0.875, 0.375, 0.875, 1, 0.625,
42    0.75, 0.875, 0.5, 0.375, 0.25,
43    0.5, 0.375, 0.375, 0.125, 0.125,
44    0.625, 1, 0.125, 1, 0.125,
45    0.125, 0.375, 0.375, 0.125, 0.125,
46    0.375, 0.125, 0.625, 0.125, 0.125,
47    1, 0.125, 0.25, 0.75, 0.875,
48    0.25, 0.5, 0.125, 0.125, 0.125,
49    0.375, 0.125, 0.625, 0.125, 0.5,
50    0.5, 0.125, 1, 0.25, 0.375,
51    0.5, 0.25, 0.25, 0.75, 0.125,
52    1, 0.125, 0.25, 1, 0.625,
53    0.125, 0.125, 0.25, 0.875, 0.25,
54    0.625, 0.25, 0.5, 0.5, 0.375,

```

```

55     0.125, 1, 0.875, 0.25, 1,
56     0.125, 1, 0.125, 0.5, 0.5,
57     0.125, 0.5, 0.75, 0.625, 0.625,
58     0.125, 0.25, 1, 0.5, 0.75,
59     0.375, 0.875, 0.25, 0.125, 0.25,
60     0.5, 0.25, 0.25, 0.125, 0.25,
61     1, 0.125, 0.625, 0.125, 0.5,
62     0.125, 0.25, 0.375, 0.25, 0.5,
63     0.25, 0.25, 0.375, 0.125, 0.125,
64     1, 0.125, 1, 0.5, 0.875,
65     0.875, 0.125, 0.375, 0.125, 0.375,
66     0.5, 0.125, 0.25, 0.125, 0.5,
67     0.375, 1, 0.5, 0.75, 1,
68     0.25, 0.5, 0.125, 0.125, 0.5,
69     0.5, 0.5, 0.25, 0.25, 0.375,
70     0.125, 0.125, 0.125, 0.25, 0.875,
71     0.625, 1, 0.625, 0.25, 0.625,
72     0.625, 0.125, 0.5, 0.125, 0.125,
73     1, 0.75, 0.25, 0.75, 0.375,
74     0.75, 0.375, 0.25, 0.5, 0.125,
75     0.125, 0.75, 0.625, 0.875, 0.25,
76     1, 0.5, 0.625, 0.5, 0.125, 0.125
77     }, VertexLabels -> "Name"]
78 M3 = WeightedAdjacencyMatrix[gf];
79 n = 0;
80 days = 100;
81 runs = 30;
82
83 Lf3 = VertexList[gf];
84 Lf3copy = Table[0, {s, 1, 24}];
85 Count3s1 = Table[Table[0, {w, 1, days}], runs];
86
87 Do[
88   Lf3[[1]] = 1;
89   Lf3[[2]] = 1;
90   Lf3[[3]] = 1;
91   Lf3[[4]] = 1;
92   Lf3[[5]] = 1;
93   Lf3[[6]] = 1;
94   Lf3[[7]] = 1;
95   Lf3[[8]] = 1;
96   Lf3[[9]] = 1;
97   Lf3[[10]] = 1;
98   Lf3[[11]] = 1;
99   Lf3[[12]] = 1;
100  Lf3[[13]] = 1;
101  Lf3[[14]] = 3;
102  Lf3[[15]] = 1;
103  Lf3[[16]] = 1;
104  Lf3[[17]] = 1;
105  Lf3[[18]] = 1;
106  Lf3[[19]] = 1;
107  Lf3[[20]] = 1;
108  Lf3[[21]] = 1;
109  Lf3[[22]] = 1;
110  Lf3[[23]] = 1;
111  Lf3[[24]] = 1;
112  Do[
113    Lf3copy = Lf3;
114    Print[Lf3]
115    Print[Count[Lf3, 1]]
116    Print[Count[Lf3, 2]]
117    Print[Count[Lf3, 3]]
118    Print[Count[Lf3, 4]]
119    Print[Count[Lf3, 5]]
120    Do[
121      If[Lf3copy[[m]] == 1,(*checks if invidual is susceptible*)
122

```

```

123 Do[
124   If[M3[[m, j]] != 0,
125     If[Lf3copy[[j]] == 3,
126
127       If [RandomVariate[ BernoulliDistribution [M3[[m, j]]]] == 1,
128         Lf3 [[m]] = 2, Lf3 [[m]] = 1,
129         n = 0]
130       , n = 0], {j, 1,
131 24}] (*when flipping coin for suseptibles , I determined p =
132 entry mj * 1/
133 2 for both this graph and the complete network so that there are \
134 consistency*)
135 ,
136 If [Lf3copy [[m]] == 2,
137   If [RandomVariate[MultinomialDistribution [1, {0.2, 0.3, 0.5}]][[
138 1]] == 1, Lf3 [[m]] = 2,
139   If [RandomVariate[MultinomialDistribution [1, {0.2, 0.3, 0.5}]][[
140 2]] == 1, Lf3[[m]] = 3, Lf3 [[m]] = 4],
141   If [Lf3copy [[m]] == 4,
142     If [RandomVariate[ BernoulliDistribution [1/2]] == 1,
143       Lf3 [[m]] = 4, Lf3 [[m]] = 5],
144     If [Lf3copy [[m]] == 5,
145       If [RandomVariate[ BernoulliDistribution [1/2]] == 1,
146         Lf3 [[m]] = 5, Lf3 [[m]] = 1],
147       If [RandomVariate[ BernoulliDistribution [1/2]] == 1,
148         Lf3 [[m]] = 3, Lf3 [[m]] = 5]
149       ]
150     ]]]
151   , {m, 1, 24}]
152 , {q, 1, days}]
153 , {k, 1, runs}]

```
