Modeling the Spread of Rumors in Networks

A Thesis
Presented to
The Math and Computer Science Department
Colorado College
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Arts

Sanjay Roberts
Faculty Advisor: Andrea Bruder

May 2016

Modeling the Spread of Rumors in Networks

Sanjay Roberts
Faculty Advisor: Andrea Bruder

Abstract

This paper examines the propagation of rumors over various social network structures. We create two modified compartmental SIR models by bridging concepts from graph theory, network theory, and epidemic modeling. We incorporate these models in three complex networks: Complete Graph, Small World, and Preferential Attachment. Then, we conduct numerical simulations through modeling software to study the speed, intensity, duration, and extent of the spreading rumor through each structure. Finally, we compare the results between the simulations and pinpoint the underlying characteristics of each network. Our results show that large centralized hubs are more effective in rumor spreading than small, dispersed, but highly connected communities. We also find that by increasing the infection rate or creating more connections within a network both lead to a faster spreading and overall larger final rumor size.

1 Introduction

A rumor is a piece of "information or a story that is passed from person to person but has not been proven to be true" [24]. Rumors have afflicted all societies throughout the course of history. They penetrate small social circles and afflict entire nations. The spreading of misinformation has had undesirable effects in changing public opinion, causing large economic loss, creating an atmosphere of skepticism around factual evidence, and inducing irrational behavior in individuals or large groups. Rumors encompass: political smear campaigns (FOX News stating Obama is a Muslim), conspiracy theories (the attacks of September 11 were an inside job), and religious and territorial conquest (the Romans defaming the Phoenicians about their supposed child sacrificial rituals in order to gain general population approval for grounds of invasion). We have even seen the spread of rumors in our daily lives these past few months: Presidential candidate Donald Trump distorting New Jersey Muslims' celebrations and saying they were cheering on the 9/11 attacks, Yik Yak posts around the University of Missouri about race riots and KKK threats leading to increased police presence in the area, David Daleiden creating a misconstrued video about Planned Parenthood and abortion which gained massive popularity, possibly leading to Colorado Springs Planned Parenthood attack by Robert Dear.

In this paper we review the areas of mathematics that contribute to rumor propagation modeling then study rumor dynamics with an adapted SIR epidemic model on three complex social networks: Complete Graph, Small World, and Preferential Attachment. We run multiple numerical simulations for each network and analyze the final rumor size, duration, intensity, and speed within each network structure. Then we compare the outcomes between networks and locate which parameters intrinsic in the network lead to each result.

The rest of the paper is organized as follows. In section 1 we introduce graph theory, its terminology, configuration, and application in social networks. Section 2 describes network theory, the formation of three types of networks, and the dimensions and specifications unique to each network. Section 3 familiarizes us with the basic SIR compartmental model including the corresponding set of ordinary differential equations for dynamical modeling. Section 4 gives a formulation of the SIR Rumor model over network structures with mean field equations. In section 5 we present numerical simulations and interpret the results. In section 6, we close the paper with a conclusions.

The study of rumor spreading first started in the 1960s with Daley and Kendall, who derived their model from infectious disease models. They were the first to apply the terms ignorant, spreader and stifler in likening rumor propagation to epidemic compartmental modeling for a susceptible, infected, and recovered population. Using a set of differential equations, they built a stochastic model with a closed and homogenously mixed population that circulated the rumor through random contact between individuals based on specified probabilities [19]. Maki and Thompson presented a derivative of this model by creating a modified deterministic DK model using directed contacts (where rumors can only spread through specific pathways), discrete time, and Markov Chains. These early models did not take into account the influence of

the structural properties of a network or the underlying effects of social interactions in a network on rumor propagation, but instead focused on a simplified, uniform model structure. Nevertheless, both models have been used extensively for quantitative studies of rumor propagation. Sudbury found the equilibrium points of the MT model and gave the equations that predict when a rumor will cease to exist within a population [22]. Lefevre and Picard analyzed the final size and duration of a rumor within a population [20]. Nekovee et al. integrated complex social networks within the rumor model and introduced a forgetting mechanism based on the attractiveness of remembering a rumor. They ran numerical simulations to examine the epidemic threshold and other dynamics on random graphs and scale-free networks to see which of the two was more inclined to spread rumors [14]. Wang et al. created a SIRaRu model by differentiating the success of a spreading rumor from the inevitability of an infectious disease by classifying individuals into two stifler categories: *Ra* individuals as people who accept the rumor but choose not to spread it, and *Ru* individuals as people who do not accept the rumor in the first place, both groups allowing ignorant individuals to bypass the spreader stage. They ran this model over homogenous and inhomogenous networks and showed that the network structure has a significant bearing on the final rumor size. Then, they incorporated a random and targeted immunization strategy to mitigate large-scale rumor propagation [15]. Li and Ma studied a delayed SIR rumor propagation feature for emergency situations [21]. Wang et al. proposed a SIR rumor model over network structures that combined individuals with a network medium for rumor propagation. They studied the increased dynamical behavior of rumor spreading not just by word of mouth but other avenues such as online sites like Facebook, Twitter, and Reddit [9].

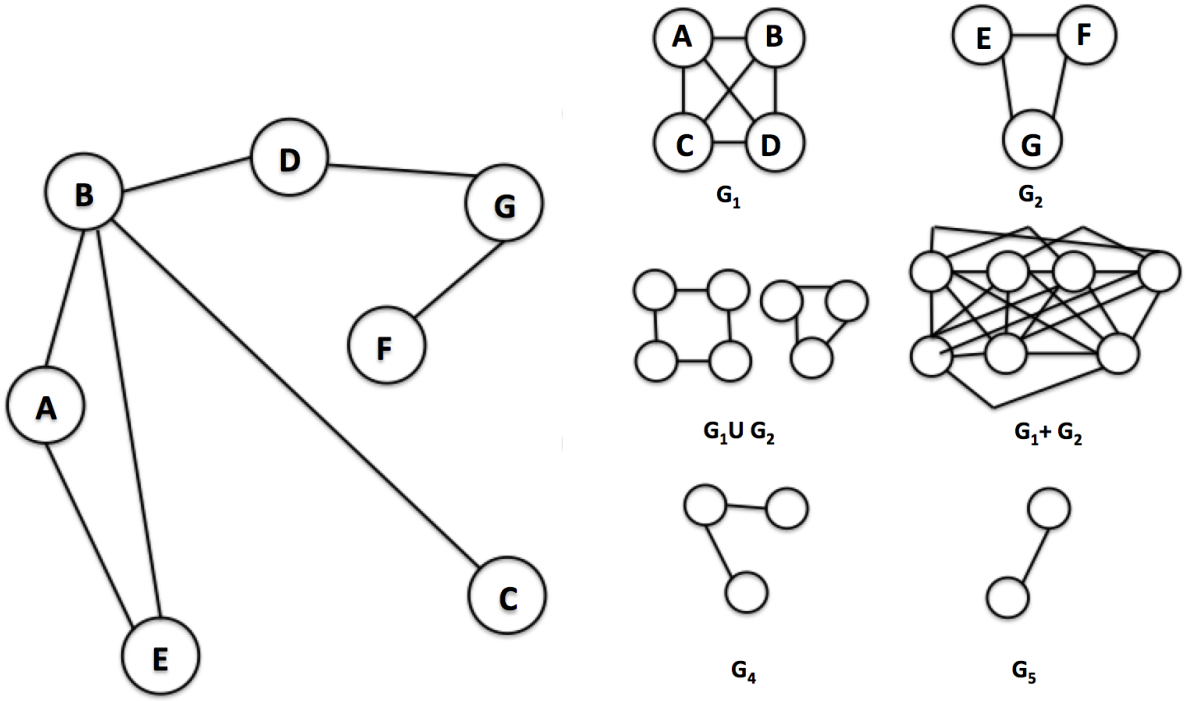
Since the 1970's, network structures have been getting a lot of attention due to their varied applications in the social sciences, mathematics, and computer science. These structures have been identified in neural networks, power grids, website linkages, and collaboration graphs between researchers as well as actors [23]. Their initial implementation in modeling was used to study the dissemination of information, ideas, news, and rumors. The development of complex network theory has brought modeling into a new period of research, allowing for more intricate models that can merge features like variable spreading rates between individuals or rules that govern the spread of disease through a diverse population to more accurately reflect real life networks. Watts and Strogatz presented the Small World (SW) network structure as a hybrid of regular and random networks using a rewiring procedure, then invited others to study disease dynamics over these network structures [23]. Zanette first employed complex network theory to study rumor propagation on small world networks [16]. Barabási and Albert presented a network structure that incorporated features of power law scaling seen in real networks to create the Barabási-Albert (BA) network. It is a model that rivals Small World in accuracy, especially in the use of social networks [12]. Newman et al. led an empirical investigation of virus propagation over a constructed email social network [18]. Following Newman, Ebel et al. built a social network from an email server of 60,000 emails and 5000 students and noted how the network analysis fit the mold of a scale-free model with small world characteristics [17]. Csányi et al. showed that human social networks are likely to be constructed of a hybrid of several networks with differing characteristics. They used Hungarian social networking site and found that it follows a combination of SW and BA networks [13].

2 Literature Review

2.1 Graph Theory

Graph Theory is an area of discrete mathematics discovered by Leonard Euler in his paper *Seven Bridges of Königsberg* written in 1736. It was considered a recreational area of mathematics with few applications until the advent of computer science, which opened the doors to theoretical chemistry, electrical engineering, and operations research [3]. Recently, it has been used in mathematical modeling to study the algorithmic and structural properties of network structures. Representing a problem as a graph can provide a different point of view and, in some cases, make the problem much simpler. Graphs excel at depicting the logical or physical linkages within a system. Combined with the techniques of Network Theory, graph theoretic representations of complex systems can be analyzed to study the structure in a system of interacting agents. Airline routes and transportation networks, friendship groups within a high school, and information dissemination through the Internet are but a few examples in which this applies [1].

A *graph* G is a pair $(V(G), E(G))$, where $V(G)$ is a non-empty finite set of elements called *nodes* and E(G) is a finite set of unordered pairs of distinct elements of V(G) called *edges* [2]. A graph is a mathematical structure used to model relationships between a collection of objects. The nodes are the set of objects within the graph, and the edges are the links that connect the objects to each other. We say that the *order* of G, or the number of nodes in G, is $|V(G)|$ and the *size* of G, or the number of edges in G, is $|E(G)|$ [3]. In Figure 1(a), the nodes of G are $V(G) = \{A, B, C, D, E, F, G\}$ and the edges of G are $E(G) = \{ab, bc, ae, be, bd, dg, fg\}$. The order and size of G is $|V(G)| = 7$ and $|E(G)| = 7$ respectively. Two nodes of a graph that have an edge joining them are called *adjacent*. In our example, nodes B and D are adjacent. We specify the *degree* of a node as the number of edges connected to that node, for instance $deg(B) = 4$ [2].



(a) A graph with 7 nodes and 7 edges.

(b) Adding, subtracting, and taking the union on a set of graphs.

Figure 1: A few example graphs noting terminology and arithmetic operations.

Now that we have the terminology for the elements of a graph, we can arrange ways to traverse the graph. Given an initial node n_0 in our graph G, we can create an *edge-sequence* of the form $n_0n_1n_2...n_{m-1}n_m$. We define a *path* as an edge-sequence in which all the nodes and edges are distinct, and a *circuit* as a path that returns to the initial node. We designate the *length* of a path or circuit as the number of edges from the beginning to end, for example in Figure 1(a) a path of length 5 is $e - a - b - d - g - f$ while a circuit of length 3 is $b - a - e - b$. The *distance* of a graph is the shortest path between two nodes, for instance the distance of C to A is 2 while a path of length 3 from C to A is $c - b - e - a$ [2].

We can label graphs by the way their elements interact with each other. A graph is *connected* if there exists a path between any pair of nodes. A graph is *simple* if there does not exist multiple edges between nodes or self-loops within a node. For example, our graph in Figure 1(a) is connected and simple. A *complete* graph, C_n is a simple graph in which every pair of nodes is adjacent. Given n nodes, we have $\frac{1}{2}n(n-1)$ edges. For example, in Figure 1(b), G_1 is complete since it has 4 nodes and $\frac{1}{2}4(4-1) = 6$ edges. A graph is *regular* if every node has the same degree, for instance G_2. Nodes in a regular graph are called *homogeneous*.

We can perform operations between and on graphs to modify them. Given $G_1 = (V(G_1), E(G_1))$ and $G_2 = (V(G_2), E(G_2))$, we can take the *union* $G_1 \cup G_2 = (V(G_1) \cup V(G_2), E(G_1) \cup E(G_2))$. We can take the *sum* of $G_1 + G_2$ by taking the union of the graphs and drawing an edge from each node of G_1 to each node

of G_2. We can also carry out *subtraction* on a graph by deleting an edge e or a node n together with all the edges connected to n, for instance in Figure 1(b), $G_4 = G_2 - gf$ and $G_5 = G_2 - E$.

We can form *disconnecting sets* and *separating sets* on a connected graph whose removal of edge-sets and node-sets respectively will disconnect G. In Figure 2, the set of edges $\{cf, ce, de, df\}$ is a disconnecting set for G while the set of nodes {C, D, E} is a separating set for G. These removals form G_1 and G_2 respectively. We define the *edge-connectivity* as the smallest number of deleted edges in order to disconnect G, and similarly, *vertex-connectivity* as the smallest number of deleted nodes in order to disconnect G [2]. In Figure 2, the edge connectivity of G is 2 since we can form the disconnecting set $\{ab, ad\}$ while the vertex connectivity of G is 2 since we can form the separating set $\{B, D\}$. These removals form G_3 and G_4 respectively. Using these operations, we are able to isolate interesting or important sections of a graph and study their structural properties. Specifically we can study the entirety of the graph, called *global metrics*, or a smaller region of the graph, called *local metrics*.

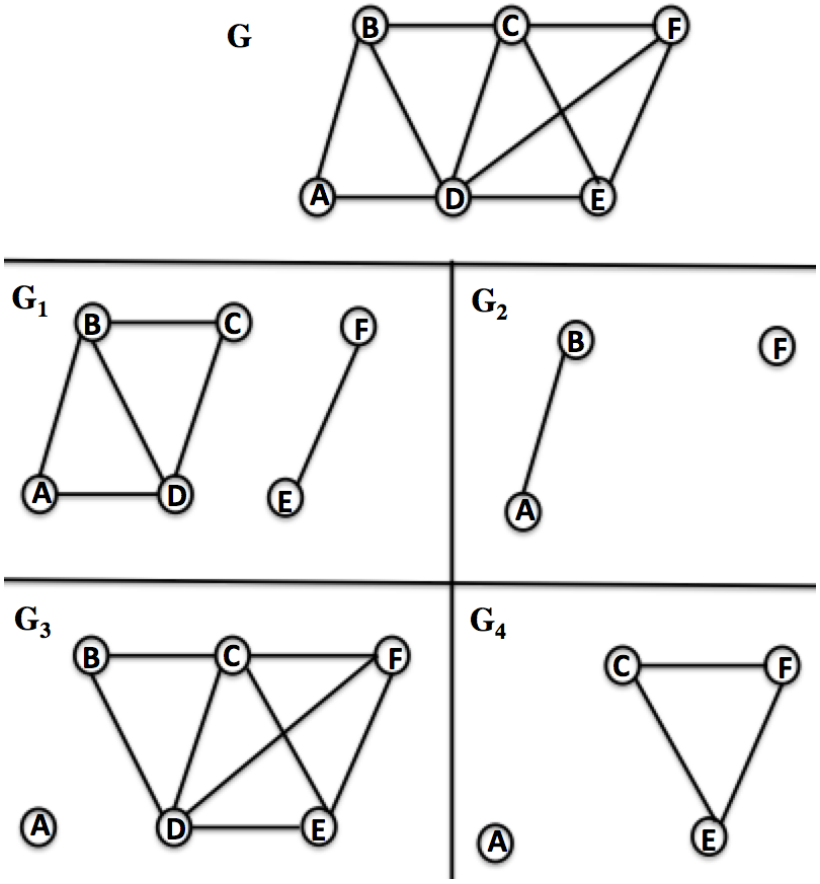
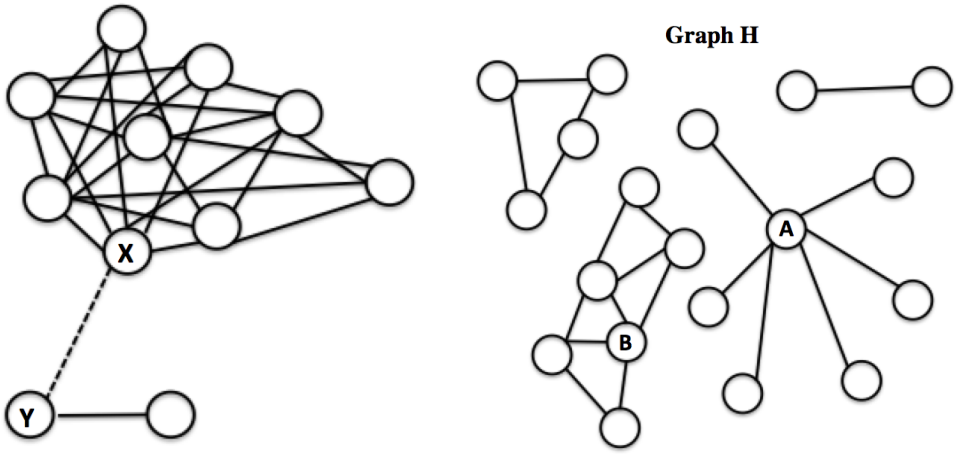


Figure 2: A depiction of disconnecting sets and separating sets on a graph.

A subset of the nodes and edges in a graph may contain certain characteristics, or relate to each other in particular ways. We call these disjoint, connected subgraphs *components*, since they are internally connected and independent pieces of the graph [1]. For example, in Figure 3(a), when we remove the edge xy we are left with a giant component containing X, and the small subgraph containing Y. We will see in the next section how large complex networks usually contain one giant component and how studying the connections between components gives insight into the underlying structure of a network. To further illustrate, in Figure 3(b), Graph H is composed of four distinctively different components. For instance, component A has a high concentration of nodes connected to a central node while component B has more connections between individual nodes but a larger degree of separation between nodes on opposite sides of the component.



(a) The subgraph containing X is a component within a larger graph.

(b) Graph H has four disconnected components.

Figure 3: Examples of disassembling graphs for further analysis.

2.2 Network Theory

Network theory is the study of complex interacting systems that can be represented by graphs. Specifically, social network theory observes social dynamics within social networks of organizations, groups, or websites by representing individuals as nodes and interactions between individuals as edges. Using graph-theoretic techniques, it has been used to describe the evolution and spread of ideas in societies [4]. By fashioning distinct models based on the allotment and connectivity of edges and nodes, networks can be made to resemble real datasets and common properties can be exhibited between different network domains [1]. Two ways to differentiate network types involve the *clustering coefficient* and the *average path length*. We will cover Random Networks, such as Erdős – Rényi model, Small World Networks, and Preferential Attachment Networks, such as the Barabási – Albert model.

First, we define *triadic closure* to be the increased likelihood of two people who share a common friend within a social network to more likely be friends themselves at some point in the future [1]. This is based on the observation that there is a greater opportunity for these two people to meet due to the mutual friend. Then we define the clustering coefficient for an individual n in the graph G as,

$$cc(n) = \frac{\text{number of pairs of adjacent neighbors}}{\text{number of pairs of neighbors}} = \frac{\text{number of triadic closures containing n}}{\text{number of possible triadic closures containing n}},$$

where $0 \leq cc(n) \leq 1$. In other words, the clustering coefficient of a node is defined as the probability that two randomly selected friends of the node are friends with each other [1]. We can take the clustering coefficient of the graph by taking the average of the clustering coefficient of all $n \in G$ so that,

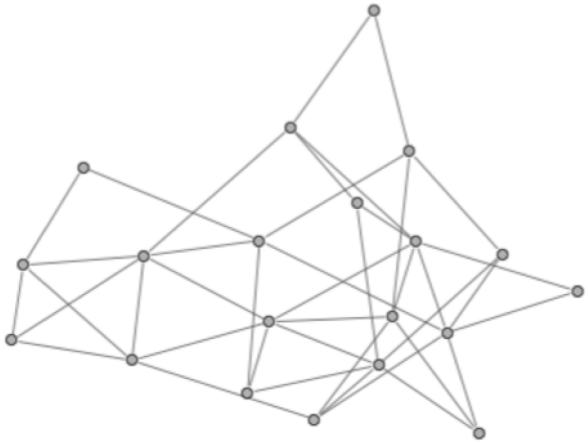
$$CC(G) = \frac{1}{|V|} \sum_{n \in V} cc(n).$$

The larger the clustering coefficient, the stronger triadic closure there is between nodes. Next, we define the *path length* between two nodes $m, n \in G$ as the length of the shortest path from m to n, denoted by $d(m, n)$. Thus, the average path length of a connected graph G is,

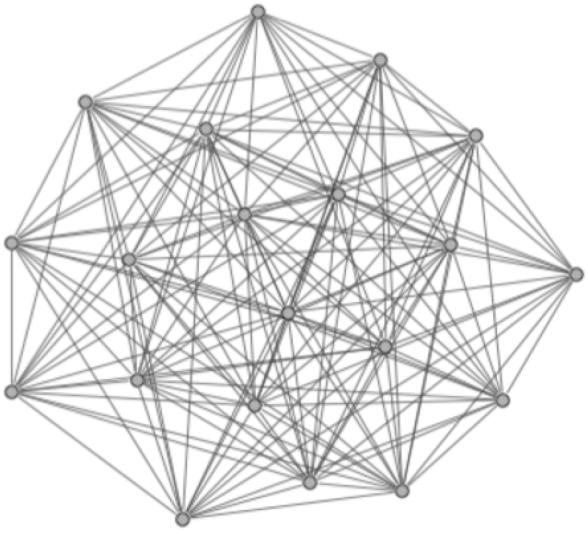
$$L(G) = \frac{1}{n(n-1)} \sum_{m,n \in V} d(m, n).$$

The Erdős – Rényi (ER) model is constructed by connecting nodes randomly. It is denoted by $G(n, p)$, where n are nodes and for any two nodes chosen there is a rewiring probability p that they are adjacent. We include edges between nodes independent from every other edge. Each node has about the same degree,

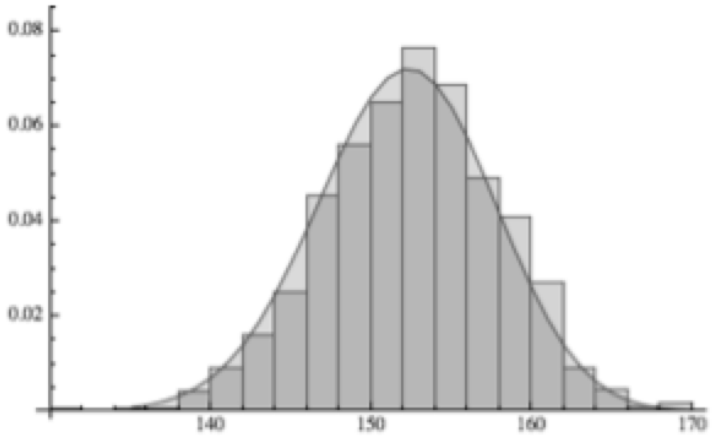
$p = \frac{\langle k \rangle}{n}$ where $\langle k \rangle$ is the average degree of a node and p is the expected degree of all nodes in ER. If p is small, the graph tends to be disconnected with many components, such as the graph $G(20, 0.2)$ in Figure 4(a). If p is large, the graph tends to be well connected, such as the graph $G(20, 0.8)$ in Figure 4(b). In general, the degree distribution of the network will follow a normal distribution with very few nodes of large or small degree. We see an example of this in Figure 4(c), where the degree for a node in $G(20, 0.8)$ follows a normal distribution. We determine the probability of a node having degree h by using the binomial distribution: $\binom{n-1}{h} p^h (1-p)^{n-1-h}$, where each node has $n-1$ opportunities to create an edge between other nodes. A graph $G(n, p)$ on average has $\binom{n}{2} p$ edges. The graph in Figure 4(a) has 41 edges while the graph in Figure 4(b) has 157 edges. The clustering coefficient of an ER graph is $CC(G(n, p)) \approx p = \frac{\langle k \rangle}{n}$. For example, $CC(G(20, 0.2)) = 0.23$ and $CC(G(20, 0.8)) = 0.82$. ER models exhibit a small average shortest path length but tend not to be clustered, for instance $L(G(20, 0.2)) = 2.12$ and $L(G(20, 0.8)) = 1.17$. Due to low clustering, ER graphs are usually not realistic for real world networks.



(a) Graph $G(20, 0.2)$ has a low rewiring probability.



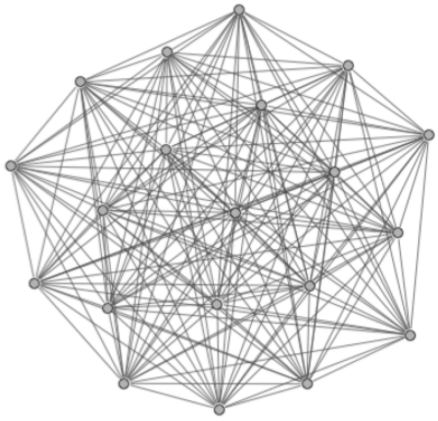
(b) Graph $G(20, 0.8)$ has a high rewiring probability.



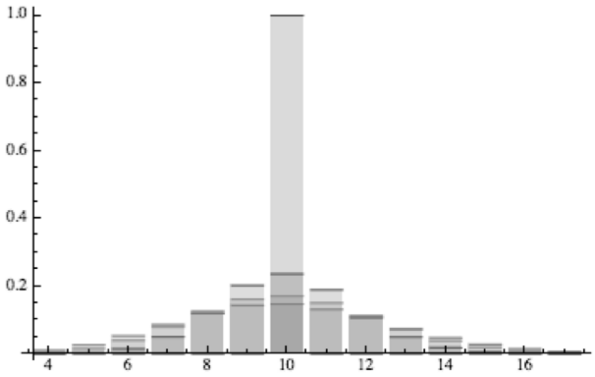
(c) Degree distribution of ER graphs follows a normal distribution.

Figure 4: Erdős – Rényi Graphs

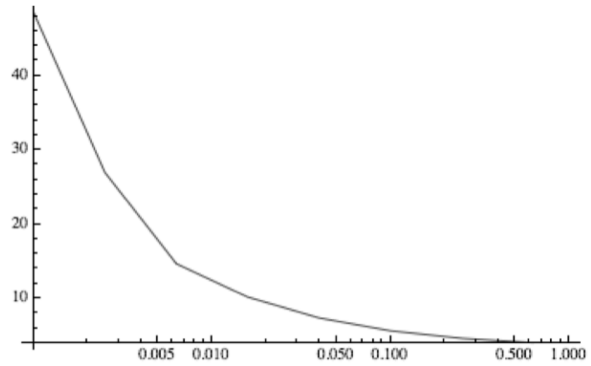
Small World networks were designed to mimic real world phenomena by demonstrating short path lengths as well as high clustering, seen by an average path length $L(G) \approx 1$ and a clustering coefficient $CC(G) \approx 1$ respectively. These networks are defined by $G(n, p, k)$ where p is the rewiring probability and k is the mean degree distribution of nodes. Figure 5(a) displays a Small World network $G(20, 0.1, 9)$ with $L(G) = 1.05$, $CC(G) = 0.94$ and an edge count of 180. The high clustering coefficient is due to the existence of *communities*, or highly connected hubs. As the probability of rewiring increases, there exists a greater likelihood for a SW network to create shortcuts between these communities. Communities tend to attract the members of other communities forming even bigger communities, where paths between these communities tend to be short. This occurrence is seen in our parameters: As p increases, $L(G)$ decreases quickly and $CC(G)$ decreases logarithmically slower than $L(G)$. This allows $L(G)$ to decrease greatly with only small decreases in $CC(G)$. Figures 5(c) and 5(d) show the graphs of $L(G)$ and $CC(G)$ as a function of p for $G(5000, p, 5)$. This relates to the node degree distribution for $G(500, p, 5)$ in Figure 5(b), which accounts for the formation of these communities since there exist nodes with a much higher degree of connections within the network. The existence of these short paths between highly connected nodes have consequences for the high potential speed with which information or diseases can spread through a social network [5].



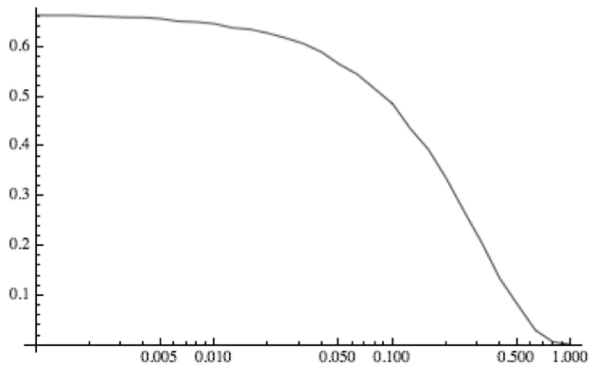
(a) Graph $G(20, 0.1, 9)$ with $L(G) = 1.05$, $CC(G) = 0.94$ and edge count of 180



(b) Degree distribution of Small World graphs



(c) Path length as a function of rewiring probability

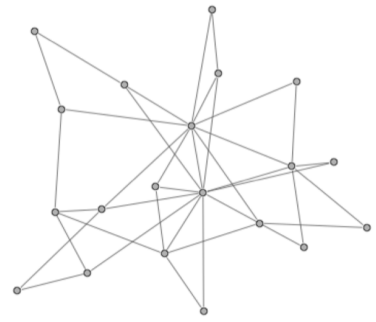


(d) Clustering as a function of rewiring probability

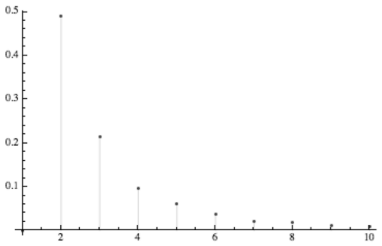
Figure 5: Small World Graphs

The Barabási – Albert model generates scale-free networks using a preferential attachment model and growth. A scale-free network has a power law degree distribution of the form $P(k) \sim k^{-\gamma}$, where $P(k)$ is defined as the probability that a node, n, chosen at random has degree k or, equivalently, as the fraction of nodes in the graph having degree k. These networks are widely observed in real world systems. A preferential attachment network works on the basis that some quantity is distributed among individuals according to how much they already have, deemed 'the rich get richer' model. Growth means that the number of nodes in the network increases over time. Thus, a BA graph is denoted by $G(n, k)$ where a new node with k edges is

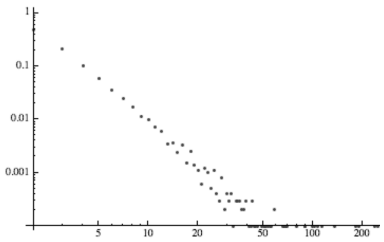
added at each step. In the case of social networks, a node with many edges represents a popular person with numerous relationships. When a newcomer enters the community, she is more likely to form a relationship with one of the well-known people rather than one of the relatively unknown individuals, thus forming a new edge to this already highly connected node. These highest degree nodes are called *hubs*, and have a stronger ability to connect with nodes added to the model. The average path length of a BA model is $L(G) \sim \frac{\ln N}{\ln \ln N}$, which is shorter than $L(G)$ for the ER model. The clustering coefficient distribution for BA models decrease as the node degree increases. This means that low degree nodes belong to very dense components, which in turn, are connected by hubs [6]. The BA graph in Figure 6(a), $G(20, 2)$, has an $L(G) = 2.12$, $CC(G) = 0.38$, and an edge count of 37. The BA graph in Figure 6(d), $G(20, 11)$, has an $L(G) = 1.19$, $CC(G) = 0.83$, and an edge count of 154. Figures 6(b) and 6(e) show the node degree distribution for each graph respectively. Figures 6(c) and 6(f)s display the power law nature of each degree distribution for each graph respectively.



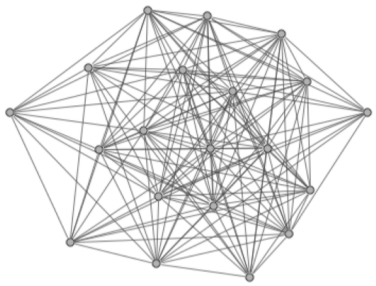
(a) Graph $G(20, 2)$ with $L(G) = 2.12$, $CC(G) = 0.38$, and edge count of 37



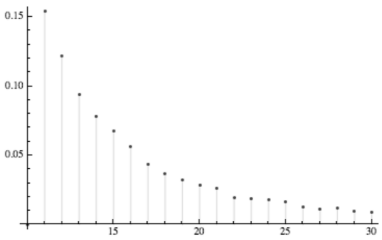
(b) Degree distribution of $G(20, 2)$



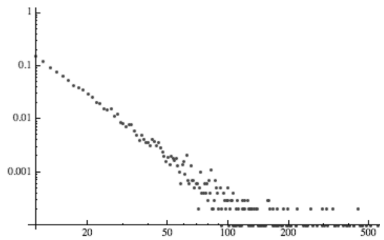
(c) Power law nature of $G(20, 2)$



(d) Graph $G(20, 11)$ with $L(G) = 1.19$, $CC(G) = 0.83$, and edge count of 154.



(e) Degree distribution of $G(20, 11)$



(f) Power law nature of $G(20, 11)$

Figure 6: Barabási – Albert Graphs

2.3 SIR Model

Infectious diseases are the second leading cause of death worldwide [8]. The spread of infectious diseases is an intricate phenomenon that depends on many complex factors such as the type of host, disease, and environment. Specifically, social contact networks pose an interesting avenue in which diseases are likely to spread from person to person. Mathematical models have incorporated network theory to simulate disease propagation by using *contact networks* which represent each individual as a node and draws an edge between two individuals who come in contact with each other to more accurately model disease transaction [1]. Models are able to represent simplified real systems as mathematical constructs by selectively choosing which parameters of study to focus on while ignoring minor factors or conditions imposed on a real system that are not pertinent to the results being studied. For example, a model can incorporate a simplistic *closed population*, meaning it does not account for new individuals entering the system via births, deaths, immigration or emigration, but rather studies the interactions between an exclusive group of individuals. These models can predict the numerous attributes of disease spread such as the duration of an epidemic as well as the total number of infected individuals. After determining propagation avenues within the system, models can interpret the data on a disease and design control measures within the system, such as quarantine,

behavior modification, and vaccination, which can be implemented to curb the intensity of spread [8].

In the late 1800s, Sir Ronald Ross developed the first mathematical models, specifically for the epidemiology of malaria, which laid the foundations for modern application of infectious disease modeling. Given a population of *hosts*, or individuals that are susceptible to a disease, and introducing a *pathogen*, or disease-causer within the hosts, Ross wanted to see how an outbreak could be modeled within a population [8]. He noted that some diseases led to an *endemic equilibrium*, where the pathogen was maintained within the population causing small outbreaks over the course of many years, while some diseases led to full-blown epidemics, where there is an intense outbreak among the population followed by the disappearance of the pathogen [7]. His research influenced the first basic SIR model by Kermack and McKendrick. This model assigns each individual in a closed population to a compartment according to its disease status at time t: $S(t)$ is the susceptible population at time t, $I(t)$ for the infectious population, and $R(t)$ for the recovered population. Interactions between individuals are determined by a set of ordinary differential equations in which a set of parameters is chosen to dictate the probability of interactions [4]. At each time step, individuals can move between compartments until the disease has run its course. Time can be measured in two approaches based on the resulting needs of the model. By using *Discrete time*, our model measures each variable once at each allotted time step resulting in a finite number of measurements. On the other hand, by using *Continuous time*, time is viewed as a continuous variable resulting in an infinite number of measurements, usually concluding in a larger yet more comprehensive dataset than discrete. To simplify, the model assumes transmission of the disease between any two hosts is a random event. It assumes *Homogeneity of Hosts* in which all hosts have the same individual characteristics, and *Uniform Mixing* in which each host is equally likely to make contact with each other host [8]. This model is able to map the pattern of spread for different pathogens and predict the future course of an outbreak in another population [8].

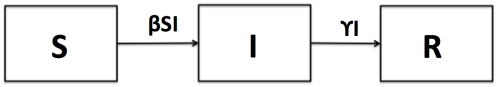


Figure 7: A basic SIR compartmental model. β is the probability of a susceptible individual contracting a disease from an infected individual, thus transitioning from the S compartment to the I compartment. γ is the probability of an infected individual recovering from a disease and transitioning from the I compartment to the R compartment

The basic model for a closed and constant population is given by the set of differential equations [7]:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \\ N &= S + I + R = 1 \end{aligned} \tag{1}$$

where N is the total population size and the initial condition for each compartment is $S(0) = S_0 \approx N$, $I(0) = N - S_0 \approx 0$, and $R(0) = 0$. For purposes of simplifying our model, we set $N = 1$ to normalize our population. In the first time step we can introduce an initially infected host, called the *index case* [8].

An individual in the S compartment is susceptible to infection by coming into contact with an infectious individual. An individual in the I compartment is infectious and has some rate of infecting susceptible individuals [1]. Each infected individual transmits the pathogen to b susceptible individuals per unit time. This *contact rate* is denoted by $\beta = \frac{b}{N}$. The number of new cases per unit time, called *secondary infections*, is $bI\frac{S}{N} = \beta SI$ [7]. By making a model assumption of *permanent immunity*, an individual who moves from the I compartment to the R compartment is removed from consideration, since the individual no longer poses a threat to the future of infection. R-stage individuals are neither infectious nor infectable since they are either dead or have permanent or temporary immunity [4]. The *rate at which infected individuals recover or die* from the disease is γ. If d is the *duration* of the infection in an individual within the model, then $\gamma = \frac{1}{d}$. We can determine the prevalence of the disease at each time step by $\frac{I(t)}{N}$. We can follow the expected

number of new infections per infected individual by the *basic reproductive rate*, $\mathcal{R}_0$. $\mathcal{R}_0$ is derived by

$$\frac{\text{time between recoveries}}{\text{time between contacts}} \times \text{total population size.}$$

Time between recoveries $= d = \gamma^{-1}$. Time between contacts $= \frac{1}{\beta}$. Thus $\mathcal{R}_0 = \frac{\beta N}{\gamma}$ [7]. If $\mathcal{R}_0 > 1$ then the disease will persist within the population. If $\mathcal{R}_0 < 1$ then the disease will die out after a finite number of time steps. In disease prevention measures, the goal is to reduce $\mathcal{R}_0$ to below 1. The final size of a major outbreak depends on $\mathcal{R}_0$.

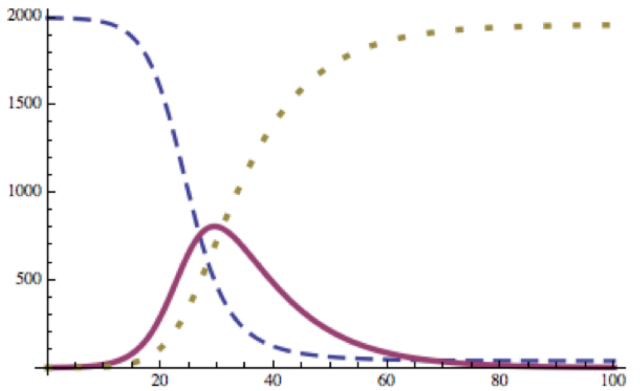


Figure 8: Graph of a basic closed population SIR model with permanent immunity. $N = 2000$, $\beta = .0002$, $\gamma = .1$, $t = 100$, $\mathcal{R}_0 = 4$, with initial conditions $S(0) = 2000$, $I(0) = 0$, and $R(0) = 0$. The susceptible population is the dashed line, the infected population is solid line, the recovered population is the dotted line.

The long-term behavior of our SIR model can be determined by looking at the *equilibria* and the corresponding *stability* at each equilibrium point. Equilibria are points where the function does not change with time, for instance when the ordinary differential equation $\frac{dx}{dt} = 0$. Stability refers to the model's response to moving a bit away from the equilibrium point. We say an equilibrium is *locally stable* if a small perturbation results in a solution that tends back toward the equilibrium point. Similarly, an equilibrium point is *locally unstable* if a small perturbation results in a solution that moves away from the equilibrium point. If a perturbation results in our point orbiting the equilibrium or staying a fixed distance away from the equilibrium, it is denoted *neutrally stable*.

For the SIR differential equations, one equilibrium is given at the point,

$$\frac{dS}{dt} = -bI\frac{S}{N} = 0.$$

This means either $S = 0$ or $I = 0$. If $S = 0$ then we have no susceptible population to infect. Thus $I = 0$, which means $S = N = 1$ since our population is normalized. This means our population is disease free; our pathogen has not infected any individuals. We say that $(S_1^*, I_1^*, R_1^*) = (N, 0, 0) = (1, 0, 0)$ is the *disease free equilibrium*.

We obtain the second equilibrium point by setting,

$$\frac{dI}{dt} = \beta SI - \gamma I = 0.$$

By factoring this equation, we see $0 = \beta SI - \gamma I = I(\beta S - \gamma)$. In this case, either $I = 0$ or $S = \frac{\gamma}{\beta}$. If $I = 0$, then we have the above case. If $S = \frac{\gamma}{\beta}$, then we can find the other equilibrium by solving for I and R. Thus to solve for I, we take,

$$\begin{aligned} N &= S + I + R \\ N &= \frac{\gamma}{\beta} + I + R \\ I &= N - R - \frac{\gamma}{\beta}. \end{aligned} \tag{2}$$

Then we solve for R by setting $\frac{dR}{dt} = 0$,

$$\begin{aligned} \frac{dR}{dt} &= \gamma I \\ 0 &= \gamma(N - R - \frac{\gamma}{\beta}) \\ R &= N - \frac{\gamma}{\beta}. \end{aligned} \tag{3}$$

Now we can find I by substituting (3) into (2),

$$\begin{aligned} I &= N - R - \frac{\gamma}{\beta}. \\ I &= N - (N - \frac{\gamma}{\beta}) - \frac{\gamma}{\beta} \\ I &= 0. \end{aligned} \tag{4}$$

We say that $(S_2^*, I_2^*, R_2^*) = (\frac{\gamma}{\beta}, 0, N - \frac{\gamma}{\beta}) = (\frac{\gamma}{\beta}, 0, 1 - \frac{\gamma}{\beta})$ is the *endemic equilibrium*.

Now we can linearize our equations $\frac{dS}{dt}$, $\frac{dI}{dt}$ to find the stability of our disease free point. This means shifting the equilibrium to the origin and redefining our variables as $S^* = N - S$ and $I^* = I$. Thus our new differential equations become:

$f_1 = \frac{dS^*}{dt} = \frac{dN}{dt} - \frac{dS}{dt} = 0 - (-bI\frac{S}{N}) = bI\frac{S}{N} = b\frac{I^*}{N}(N - S^*) = bI^* - b\frac{I^*S^*}{N}$

$f_2 = \frac{dI^*}{dt} = \frac{dI}{dt} = bI\frac{S}{N} - \gamma I = b\frac{I^*}{N}(N - S^*) - \gamma I^* = bI^* - b\frac{I^*S^*}{N} - \gamma I^*.$

Then we take the Jacobian matrix of our equations:

$$J(s,i) = \begin{pmatrix} \frac{\partial f_1}{\partial s} & \frac{\partial f_1}{\partial i} \\ \frac{\partial f_2}{\partial s} & \frac{\partial f_2}{\partial i} \end{pmatrix} = \begin{pmatrix} -\frac{bI^*}{N} & -\frac{bS^*}{N} \\ \frac{bI^*}{N} & \frac{bS^*}{N} - \gamma \end{pmatrix} = \begin{pmatrix} -\beta I^* & -\beta S^* \\ \beta I^* & \beta S^* - \gamma \end{pmatrix}$$

Let $s = N = 1$ and $i = 0$ for our disease free point.

$$J(1,0) = \begin{pmatrix} 0 & -\beta \\ 0 & \beta - \gamma \end{pmatrix}$$

Our eigenvalues are: $\lambda_1 = 0$ and $\lambda_2 = \beta - \gamma$. If our eigenvalues are negative or zero, then the equilibrium is stable. Otherwise, the equilibrium is unstable if the eigenvalues have a positive part. If $\mathcal{R}_0 > 1$, then our point is stable and if $\mathcal{R}_0 < 1$, then our point is unstable. Thus λ_1 is stable, and λ_2 is stable if $\mathcal{R}_0 < 1$, i.e. if $\gamma > \beta$, and unstable if $\mathcal{R}_0 > 1$, i.e. if $\beta > \gamma$. Recall, if $\mathcal{R}_0 < 1$ then no epidemic occurs and if $\mathcal{R}_0 > 1$, then an epidemic takes place [8].
Let $s = \frac{\gamma}{\beta}$ and $i = 0$ for our endemic equilibrium point.

$$J(\frac{\gamma}{\beta}, 0) = \begin{pmatrix} 0 & -\gamma \\ 0 & 0 \end{pmatrix}$$

Our eigenvalues are: $\lambda_1 = 0$ and $\lambda_2 = 0$. Thus, our equilibrium point is stable.

3 Model derivation

3.1 The Simple Rumor Model

Social network analysis examines the structure of relationships between entities such as groups of people, social media sites, and scholarly publications. Specifically, this analysis has been key in understanding the propagation of news and rumors [10]. Though authoritative sources and expert opinions usually curb the

intensity of propagation, rumors continue to have a great impact on society in their power to shape public opinion [11]. Rumor propagation has been compared to the spread of computer viruses or epidemic diseases. Studying rumor propagation mechanisms and the network structures in which they are disseminated are important for reducing the effect of rumor spreading.

Mathematical models have tried to capture this phenomenon and, presently, have been combining a modified SIR model with Network Theory. For example, in a social network, individuals are nodes and contacts between different people are edges. The standard rumor-spreading model is the Daley-Kendall model, a modified SIR model. From a fixed social network of $N = S + I + R$ individuals we have three groups: ignorants, spreaders, and stiflers. The ignorants, S, have never heard the rumor. The spreaders, I, are actively spreading the rumor to ignorants. The stiflers, R, have heard the rumor but are no longer interested in spreading it [9]. Two types of rumor spreading mechanisms can be used. The *push-model* assumes that only spreaders actively seek ignorants to inform them of the rumor. The *push-pull model* assumes the pull-model but additionally ignorants contact spreaders in attempts to learn of the rumor [11].

The rumor is propagated through the population by pair-wise contacts. A spreader who contacts an ignorant attempts to "infect" the ignorant with the rumor determined by a 'probability of infection' λ. If successful the ignorant becomes a spreader. When a spreader interacts with another spreader or stifler, the initial spreader will become a stifler with a probability σ, due to the initial spreader realizing the rumor has lost its 'shock value'. With a certain probability δ, the spreader will forget or become uninterested in spreading the rumor, thus becoming a stifler. Nothing occurs when an ignorant meets another ignorant, or an ignorant meets a stifler [9]. Fig 8 depicts a compartmental model which shows the flow of rumor propagation. Our differential equations for this model are:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\lambda S(t)I(t) \\
\frac{dI(t)}{dt} &= \lambda S(t)I(t) - \sigma I(t)(I(t) + R(t)) - \delta I(t) \\
\frac{dR(t)}{dt} &= \sigma I(t)(I(t) + R(t)) + \delta I(t)
\end{aligned} \tag{5}$$

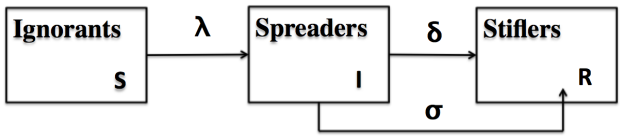


Figure 9: λ is the probability of a S-I interaction. δ is the spreader's rumor forgetting rate. σ is the probability of a spreader interacting with a stifler or another spreader.

3.2 Rumor Spreading Model Over Complex Networks

To create a rumor spreading model over complex networks, we modify the simple rumor model by including the average degree of the network, $\langle k \rangle$, into our equations:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\lambda \langle k \rangle I(t)S(t) \\
\frac{dI(t)}{dt} &= \lambda I(t)S(t) - \sigma \langle k \rangle I(t)(I(t) + R(t)) - \delta I(t) \\
\frac{dR(t)}{dt} &= \sigma \langle k \rangle I(t)(I(t) + R(t)) + \delta I(t)
\end{aligned} \tag{6}$$

Then, we combine these equations with the adjusted specifications inherent in each network, such as the path length and clustering coefficient. In the next section, we will analyze rumor propagation through homogeneous networks, meaning that all nodes within the model share the same characteristics. Every edge in each model has an equally weighted probability for rumors to pass through them. We use the simple rumor model for the Complete Graph simulations, with reasonings which will be explained in greater detail

later. Complete Graphs contain all possible edges between their nodes so that every node is connected to every other node. We use the complex networks model for Small World and Preferential Attachment graph simulations.

To get a general idea of how this model works, we see that as the spreading rate, λ, increases SIR graph shifts left since the outbreak becomes more intense. The spreader curve gets larger over a smaller time period, while the susceptible curve falls. As we increase the average mean degree, $\langle k \rangle$, the graph shifts left as more individuals have a higher average degree of connections allowing for a rumor to hold a greater chance of spreading through different avenues of individuals. As we increase the probability of a spreader interacting with a stifler or another spreader, σ, the spreader curve drops as the stifler curve grows due to the increased rate at which spreaders become stiflers. At a certain threshold, if λ is low and σ is high, then the spreader category can die out. As we increase the rate of forgetting a rumor, *delta*, we find a similar graphical interpretation to σ.

4 Results of Numerical Simulations and Discussion

To perform a numerical analysis, I linked the simulation programs *Netlogo IONTW* to *Mathematica*. Netlogo IONTW excels at modeling and running SIR model simulations over complex networks. It also calculates and records important metrics pertaining to the network before and after simulating the disease, such as the expected mean degree, clustering coefficient, path length, and the basic reproductive rate. Only the constants λ (not the same λ from our models above), d, and the infection rate can be varied to produce the outcome of our parameters within the Netlogo program, so I focused on these parameters for running simulations and comparing results. Also, note that each of these parameters controls a different aspect within each model, which will be explained later. I kept a closed and constant population of $N = 100$ for all experiments. Netlogo uses a stochastic algorithm to create and simulate its models, so replicating and studying one particular (deterministic) model with identical parameters is difficult. Though it is able to run batch processes and give the outcome results for each simulation, it is not able to parse the data at each time step to demonstrate the inner mechanics of how a simulation obtains its results. Mathematica, on the other hand, is not equip with a complex network package, even though it has an awesome differential equations solving and modeling engine. By linking the two programs together, I was able to track and report on the propagation of the disease at each time step and generate a basic SIR graph for each network.

4.1 Complete Graph Model

For the Complete Graph model, I ran 4 experiments with differing infection rates. The experiments consisted of 250 simulations, tracking the number of susceptible, infectious, and recovered at each discrete time step for 250 time steps separately. After collecting the data, I took the mean of each category for each corresponding time step (for example, the number of infected at t=35 for all simulations) and plotted a line graph for each S, I, and R curve then compared the results using the collected data.

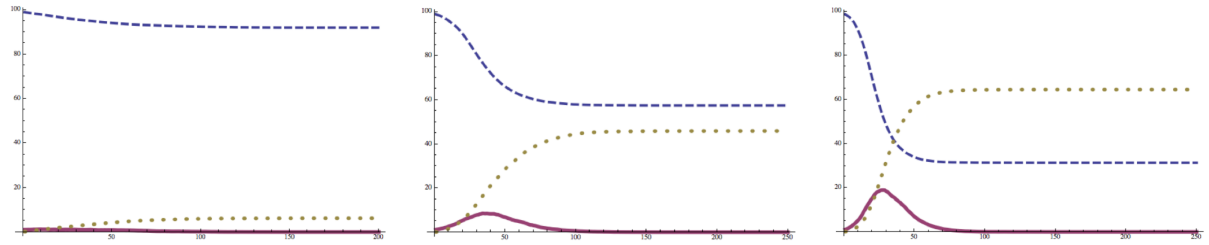
Table 1: Complete Graph Constants and Parameters and their Effects

Constant	Effect
λ	No effect because Complete graph will always have a 100% clustering coefficient
d	No effect because Complete graph will always have a path length of 1
$\langle k \rangle$	Does not vary because expected mean degree for Complete graph is always $N - 1 = 100 - 1 = 99$

4.1.1 The Effect of Increasing the Infection Rate β

The basic reproductive rate, $\mathcal{R}_0$, varies in correlation with the infection rate. As the infection rate increases, $\mathcal{R}_0$ increases, and we see the duration of the rumor go from short to long to constant. We see a similar result for the time of the infection's peak. This could be because the rumor with $\mathcal{R}_0 = 1$ fizzles out within the population quickly since it is not able to gain traction and spread. When $\mathcal{R}_0 = 2$, the rumor lasts longer in the population since the rumor is stronger, infecting people over a longer period of time (think of the flu). Then, when $\mathcal{R}_0 = 3$ the rumor spreads quickly since it is highly infectious, but fizzles out at the same time

as $\mathcal{R}_0 = 2$. This may be due to the possibility of the model passing a certain threshold of population (in this case 65%), where there ends up being a smaller pool of susceptible individuals that can become infectious as more individuals become infectious or removed. The percentage of spreaders at the peak and the total removed individuals shoot up drastically from $\mathcal{R}_0 = 1$ to $\mathcal{R}_0 = 2$, but then level out from $\mathcal{R}_0 = 2$ to $\mathcal{R}_0 = 3$ since we have a small finite population N.



(a) $\mathcal{R}_0 = 1$ (b) $\mathcal{R}_0 = 2$ (c) $\mathcal{R}_0 = 3$

Figure 10: Increasing the infection rate in Complete Graph results in an increase in $\mathcal{R}_0$. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.2 Small World Model

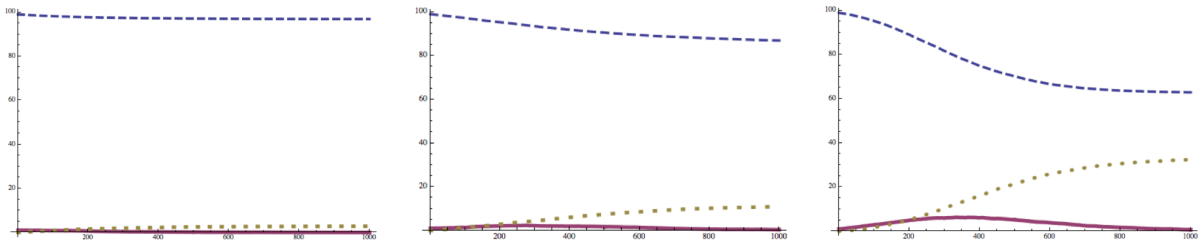
For the Small World model, Netlogo combines an Erdős – Rényi graph with a Nearest Neighbor graph (meaning the nodes are arranged in circle instead of a lattice or square or random placement in space). I ran four experiments varying the infection rate, three experiments varying λ, and three varying d. I ran 250 simulations with 1000+ time steps for each experiment.

Table 2: Small World Constants and their Effects

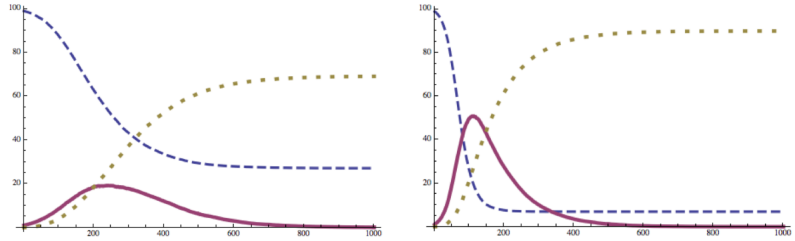
Constants	Effects
λ	affects the rewiring probability and the expected mean degree
d	A node is connected to the d nearest neighbors to each side of every node
$\langle k \rangle$	Expected mean degree of each node in the network

4.2.1 The Effect of Increasing the Infection Rate β

As we increase the infection rate, our parameters $\langle k \rangle$, $CC(G)$, and $L(G)$ stay approximately the same; only $\mathcal{R}_0$ increases as the infection rate increases. When $\mathcal{R}_0 > 1.5$, major rumor epidemics occur ($> 25\%$ of population). When $\mathcal{R}_0 < 1.5$, the rumor tends to die out quickly, not affecting a large percentage of the population at any given time. The time of the rumor peak occurs later as $\mathcal{R}_0$ approaches 2, then occurs sooner with more total percentage of the population becoming infected as $\mathcal{R}_0 > 2$. Similarly with the duration of the infection for major outbreaks, as $\mathcal{R}_0$ increases, the duration decreases and the intensity of the outbreaks increase. As $\mathcal{R}_0$ grows, a larger percentage of total individuals get infected by the rumor. Beyond $\mathcal{R}_0 = 5$, the entire population appears to succumb to the rumor.



(a) $\mathcal{R}_0 = 0.75$ (b) $\mathcal{R}_0 = 1.4$ (c) $\mathcal{R}_0 = 1.9$

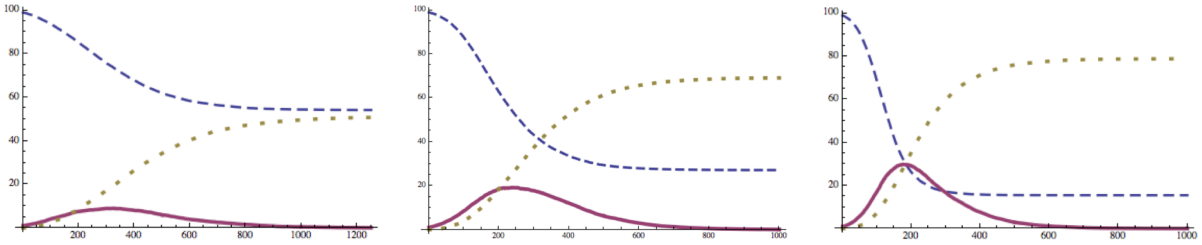


(d) $\mathcal{R}_0 = 2.9$ (e) $\mathcal{R}_0 = 4.7$

Figure 11: Increasing the infection rate in Small World results in an increase in $\mathcal{R}_0$. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.2.2 The Effect of Increasing λ

If we manipulate our constant λ, we see an increase in our parameters $\langle k \rangle$ and $\mathcal{R}_0$, but a decrease in $CC(G)$ and $L(G)$. The edge density increases. The time of peak infection decreases, signifying the enhanced speed of the spread of the rumor through the population. This also corresponds to a shorter duration time as λ increases. Also the intensity and strength of the rumor increases noted from the larger final rumor sizes and the percentage of the population affected during the peak of infection. The total number of removed individuals tapers off as λ increases due to less susceptibles present in the total population.

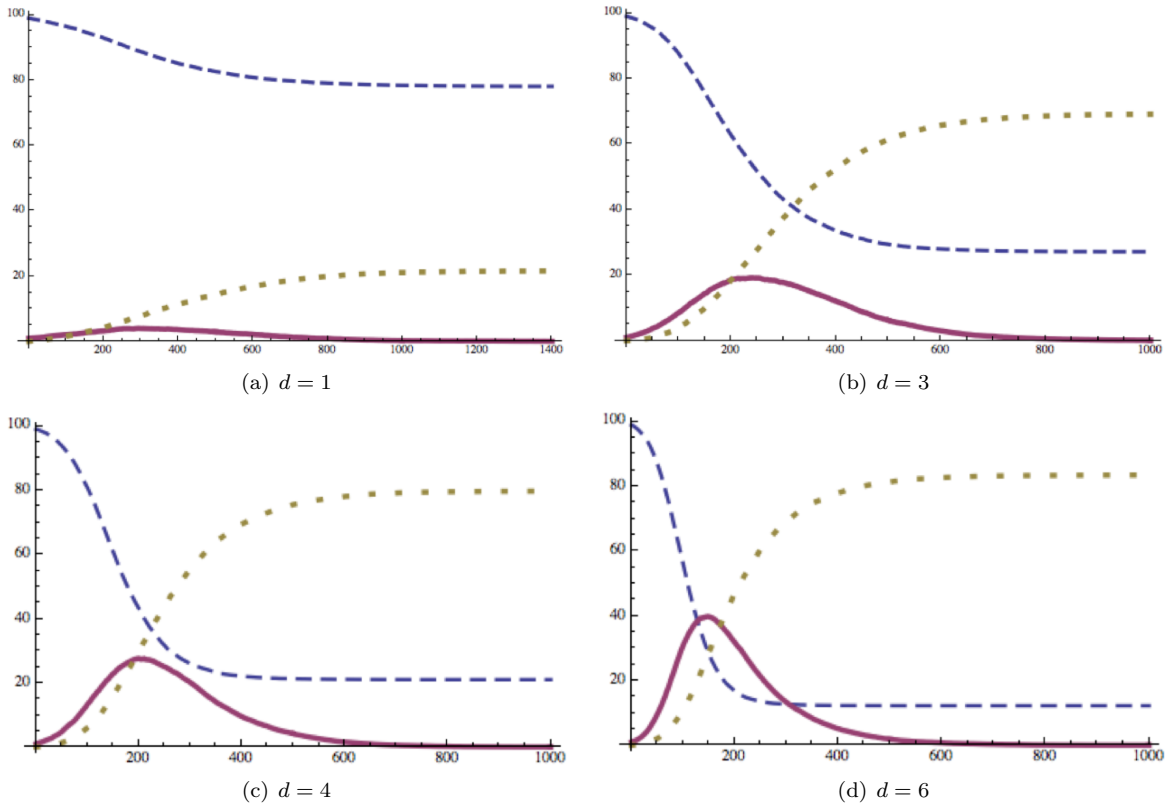


(a) $\lambda = 1$ (b) $\lambda = 2.5$ (c) $\lambda = 5$

Figure 12: Increasing λ in Small World results in an increase of the mean degree and $\mathcal{R}_0$, and a decrease in the clustering coefficient and path length. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.2.3 The Effect of Increasing d

We get results similar to λ as d is increased, except that $CC(G)$ increases instead of decreases as d gets larger. d seems to have a dramatic effect in increasing $CC(G)$, by about 7% when adding 1 more edge to each node (i.e. d going from 3 to 4). It appears that the larger $CC(G)$ is, the longer it takes for the rumor to peak, yet the intensity of the outbreak and final rumor size end up the same, simply shifting the spreader curve to the left.



(a) $d = 1$ (b) $d = 3$

(c) $d = 4$ (d) $d = 6$

Figure 13: Increasing d in Small World results in an increase of the mean degree, $\mathcal{R}_0$, and the clustering coefficient, while a decrease occurs in the path length. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.2.4 Discussion on Parameter's Effects on Small World

As λ increases, the average number of edges on a node increase causing more connections between nodes within the network. Counter intuitively, when we increase λ, we actually dilute the already strong clustering that occurs within this network. This occurs because increasing λ can cause there to be a greater likelihood of long connections (rewirings) existing from nodes across the network that break up the small clustered components that may exist on the edges of the network where there are many short connections within closely spaced nodes. Thus $CC(G)$ decreases as λ increases. As λ increases, the existence of more edges within the model cause $L(G)$ to decrease since there exist more avenues for shorter lengths between nodes. As d increases, $CC(G)$ increases since greater clustering forms with nearby nodes. As with λ, as d increases, $L(G)$ decreases for similar reasons. We can compute the rewiring probability as $p = \frac{\lambda}{N-1}$. The expected mean degree of the network can be computed within the network as $\langle k \rangle = 2d + \lambda - \frac{dN\lambda}{N-1}\frac{2}{N} = 2d + \lambda - \frac{2d\lambda}{N-1}$.

4.3 Preferential Attachment Model

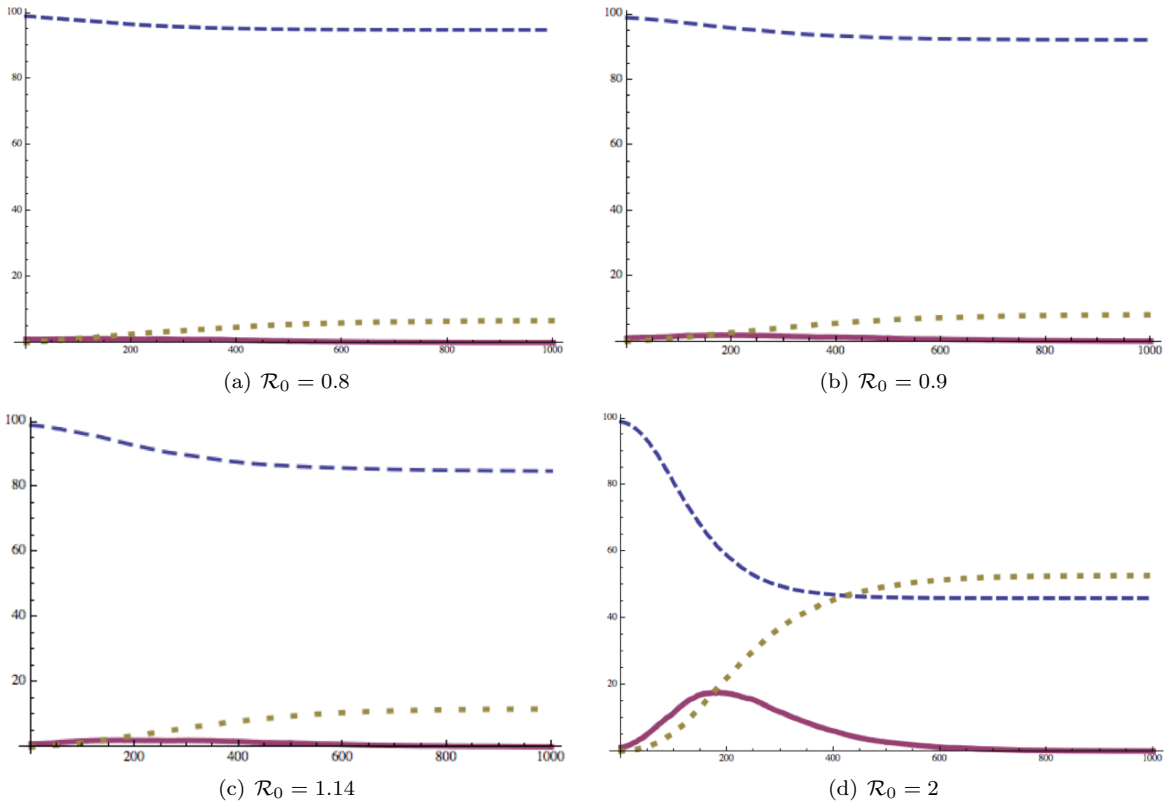
For the Preferential Attachment model, I ran four experiments varying the infection rate, four varying λ, and 3 varying d. Similar to Small World, I ran 250 simulations with 1000+ time steps for each experiment. This graph initially begins with a complete graph of n nodes, then adds the remainder $N - n$ nodes to the graph one node at a time. Each new node added randomly attaches to a certain number of nodes within the current model in proportion to the degree of each node from the previous iteration of adding nodes (rich get richer).

Table 3: Preferential Attachment Constants and their Effects

Constant	Effect
λ	This model begins with an initial complete graph of λ-nodes
d	Each new node added will have d edges
$\langle k \rangle$	Expected mean degree

4.3.1 The Effect of Increasing the Infection Rate

As we increase the infection rate, our parameters $\langle k \rangle$, $CC(G)$, and $L(G)$ stay approximately the same; only $\mathcal{R}_0$ increases as the infection rate increases. When $\mathcal{R}_0 < 1.2$, the infection duration lasts longer the closer $\mathcal{R}_0$ approaches 1.2, then it shifts back to a medium duration as $\mathcal{R}_0$ approaches 2. The time of peak infection shifts left, then right, then ends up in the middle where it continues to stay as $\mathcal{R}_0 > 1.2$. The total removed and percentage of spreaders at peak time steadily increase as $\mathcal{R}_0$ increases. The rumor peak and rumor duration tend toward the middle after $\mathcal{R}_0$ gets above 1.2 since when $\mathcal{R}_0$ is below a certain threshold, the rumor gains momentum but is still not strong enough to permeate through the population. Instead it slowly infects individuals and stays within the population for longer periods of time. Once $\mathcal{R}_0$ passes the threshold, then it returns to its normal spreading pattern, permeating throughout the population at a quicker rate, then fizzling out at a faster rate as less susceptibles that can be infected are present in the population.



(a) $\mathcal{R}_0 = 0.8$ (b) $\mathcal{R}_0 = 0.9$

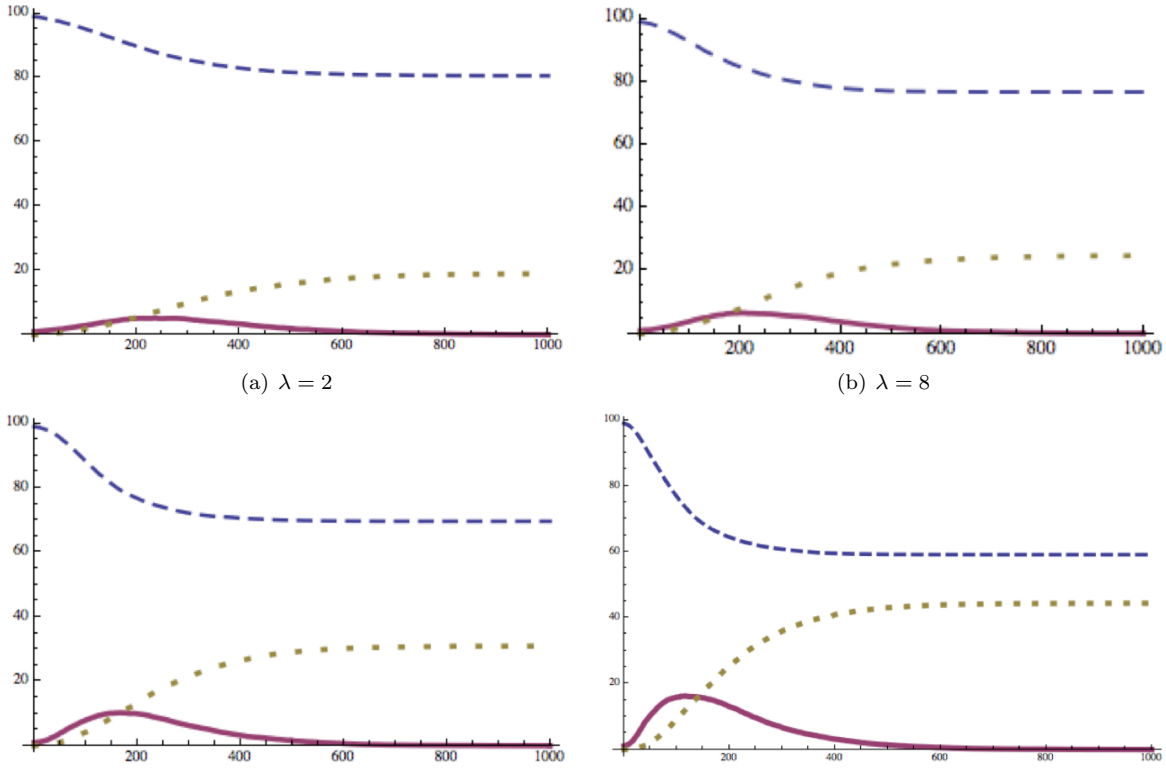
(c) $\mathcal{R}_0 = 1.14$ (d) $\mathcal{R}_0 = 2$

Figure 14: Increasing the infection rate in Preferential Attachment results in an increase in $\mathcal{R}_0$. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.3.2 The Effect of Increasing λ

If we manipulate our parameter λ, we see an increase in $\langle k \rangle$, $CC(G)$, and $\mathcal{R}_0$ but a decrease in $L(G)$. The time of the rumor's peak is sooner (graph shifts to the left). As λ increases the intensity of the outbreak increases, as seen through the increases in the percentage of spreaders at peak time and the total removed. The duration of the infection is shorter then starts to get longer between $14 < \lambda < 25$. The peak infection

time occurs once the higher degree nodes get infected then start infecting the lower degree nodes. In Small World $CC(G)$ decreases when we increase λ, but in Preferential Attachment, $CC(G)$ increase. This occurs because increasing λ creates a larger initial closely clustered ($CC(G) = 1$) complete graph. Since d is constant in this particular experiment, the initial complete graph is the determining factor of $CC(G)$.



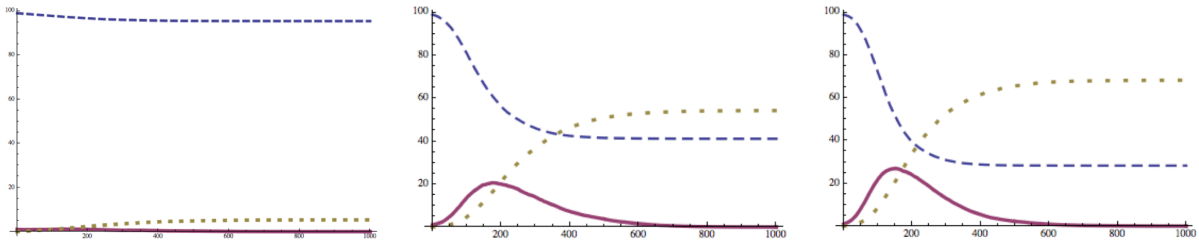
(a) $\lambda = 2$ (b) $\lambda = 8$

(c) $\lambda = 14$ (d) $\lambda = 25$

Figure 15: Increasing λ in Preferential Attachment results in an increase of the mean degree, $\mathcal{R}_0$, and the clustering coefficient, while a decrease in path length. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.3.3 The Effect of Increasing d

Similarly to above, if we manipulate our constant d, we see an increase in $\langle k \rangle$, $CC(G)$, and $\mathcal{R}_0$, and a decrease in $L(G)$. This makes sense since there are $(N-\lambda)d$ edges present in each experiment, so as d grows, the expected mean degree gets larger, the clustering coefficient gets larger, and the path length gets shorter. The total removed and spreader percentage at peak time both increase as d increases. The peak infection time shifts right then left, similarly with the infection duration.



(a) $d = 1$ (b) $d = 4$ (c) $d = 5$

Figure 16: Increasing d in Preferential Attachment results in an increase of the mean degree, $\mathcal{R}_0$, and the clustering coefficient, while a decrease occurs in the path length. The dashed line represents the number of ignorants, $S(t)$. The solid line represents the number of spreaders, $I(t)$. The dotted line represents the number of stiflers, $R(t)$.

4.3.4 Discussion on Parameter's Effects on Preferential Attachment

As we increase λ, there will be more edges present but only in proportion to the initial complete graph, $\frac{1}{2}n(n-1)$. As λ increases, the duration of the rumor gets longer between $14 < \lambda < 25$ because we have less centralized hubs with a large degree since we have up to $\frac{1}{4}$ (25 nodes) of our population in a tightly clustered group. In this sense, our graph is mimicking a Small World graph. Thus the rumor will not spread as quickly as if there were one or two large hubs that would receive the rumor and spread it to all its connecting nodes. Instead, the rumor must begin infecting the larger cluster before leaving the cluster and infecting the remaining nodes. In Preferential Attachment, it appears that as the path length gets shorter and the expected mean degree gets larger, the speed of the rumor sweeps quicker through the population after a certain threshold for each. If it is too far below this threshold, then the infection will peak and die out quickly, not effecting too large a chuck of the population. Similarly if it is high above the threshold, it will peak and die out, but will cause mass infection in the process.

4.4 Comparison Between Graphs

4.4.1 Final Rumor Size

I matched similar final rumor sizes (FRS) between all three simulations and compared the results. Referring to Table 4, we see similar percentages of spreaders at peak time for FRS$\approx$10. For FRS$\approx$30 and 50, the percentage of spreaders at the time of peak infection for Small World and Complete Graph (the intensity of the rumor) is about half the intensity for Preferential Attachment, then levels out to $\frac{3}{4}$ the intensity of Preferential Attachment for FRS $\approx$ 70. Complete Graph is much faster in reaching the peak infection time than Small World and Preferential Attachment, as well as having the shortest rumor duration of all 3 networks. This is probably due to the maximum number of edges present in all Complete Graph models which enables the quickest spread of infection. Preferential Attachment spreads the infection at a faster rate than Small World, in some cases almost doubling the infection duration and time of peak infection. Again, I would argue this is due to the highly connected nodes that exist within Preferential Attachment, and, once infected, are liable to spreading the rumor to the remainder of the network. This phenomenon is compared to the highly connected but separate communities that exist within Small World network.

Small World appears to have similar infection durations regardless of the final size of the rumor. Yet the time of infection peak occurs quicker when FRS $\approx$ 10, then longer when FRS $\approx$ 30 and 50, then quickest when FRS $\approx$ 70. This is probably due to the strength of the rumor permeating through the population. As the rumor size gets larger for Preferential Attachment, both the rumor duration and time of peak infection occur at a later period of time. We see a similar phenomenon in Complete Graph.

In general, Preferential Attachment has smaller $\langle k \rangle$, CC(G), and L(G) values than Small World in each scenario, sometimes up to half for $\langle k \rangle$ and $CC(G)$ in each case. The expected mean degree is less since we have a power law degree distribution in Preferential Attachment, and a normal distribution in Small World. The clustering coefficient is smaller since Small World is designed to have many small hubs throughout the perimeter of the network while Preferential Attachment is designed to have a few big hubs that extend and connect throughout the network. The path length is shorter for PA since with one or two big central hubs, one can jump quickly through the network to any other node by first moving to a big hub.

4.4.2 Percentage of Spreaders at Rumor Peak

I matched similar spreader percentages between all three simulations and compared the results. Referring to Table 5, we see similar results as above: $\langle k \rangle$, L(G), CC(G) are smaller in Preferential Attachment than Small World. There are very similar $\langle k \rangle$ values but much smaller values of CC(G) in Preferential Attachment. Complete Graph is the fastest in spreading the rumor between the three models. Small World has a larger final rumor size than Preferential Attachment in each case, though the intensity is the same. It appears that the rumor lasts longer and is slower in peaking in Small World than Preferential Attachment, yet more people are infected by the rumor in Small World. This is due to the fact that in Small World there exist a greater number of separate highly connected communities that spread the rumor at a slower rate, taking longer to diffuse the rumor to different parts of the network, than the one or two highly connected hubs in Preferential Attachment that spread the rumor everywhere through the network quickly. The shorter path

	$\mathcal{R}_0$	β	α	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	RD
$FRS \approx 10$									
CG = 6	1	0.0011	0.1	-	-	-	10	2	144
SW = 3	0.74 - 0.78	0.001	0.01	8.1 - 8.6	0.33 - 0.36	2.5 - 2.6	7	2	392
SW = 11	1.3 - 1.45	0.002	0.01	8.3 - 8.8	0.33 - .036	2.5 - 2.6	273	3	1000
PA = 6	0.8	0.0025	0.01	4	0.13 - 0.17	2.9 - 3	130	2	577
PA = 5	0.7	0.005	0.01	2.1	0.02	3.6 - 4.6	148	2	535
$FRS \approx 30$									
SW=32	1.8 - 2	0.003	0.01	8.1 - 8.5	0.33 - 0.37	2.5-2.6	368	5	1000
PA=31	1.76	0.005	0.01	5.26	0.34 - 0.44	2.5 - 2.7	147	10	676
$FRS \approx 50$									
CG=46	2	0.0021	0.1	-	-	-	36	9	154
SW=50	2.2 - 2.4	0.005	0.01	6.8 - 7.2	0.46 - 0.49	2.9 - 3.1	330	10	1089
PA=53	2	0.01	0.01	4	0.1 - 0.2	2.9-3.1	186	17	755
PA=54	2.6	0.005	0.01	7.8	0.16 - 0.21	2.3	198	20	831
$FRS \approx 70$									
CG=65	3	0.00312	0.1	-	-	-	34	19	153
SW=70	2.7 - 3.1	0.005	0.01	8 - 9	0.33 - 0.38	2.5 -2.7	240	20	885
PA=68	3.2	0.005	0.01	9.7	0.18 - 0.22	2.2	167	27	773

Table 4: Comparison of Final Rumor Size

length in Preferential Attachment lead in part to the quicker speed at which the rumor is disseminated. Along the same lines, a higher clustering coefficient in Small World leads to a slower speed of dissemination.

	$\mathcal{R}_0$	β	α	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	$R(\infty)$	RD
% I(t)$\approx$ 10									
CG=9	2	0.0021	0.1	-	-	-	36	46	154
SW=10	2.2 - 2.4	0.005	0.01	6.8 - 7.2	0.46 - 0.49	2.9 - 3.1	330	50	1089
PA=10	1.76	0.005	0.01	5.26	0.34 - 0.44	2.5 - 2.7	147	31	676
% I(t)$\approx$ 20									
CG=19	3	0.00312	0.1	-	-	-	34	65	153
SW=20	2.7 - 3.1	0.005	0.01	8 - 9	0.33 - 0.38	2.5 -2.7	240	70	885
PA=20	2.6	0.005	0.01	7.8	0.16 - 0.21	2.3	198	54	831
% I(t)$\approx$ 30									
SW=27	3.3 - 3.5	0.005	0.01	10 - 10.5	0.41 - 0.45	2.4 - 2.5	218	80	787
PA=27	3.2	0.005	0.01	9.7	0.18 - 0.22	2.2	167	68	773

Table 5: Comparison of Percentage of Spreaders at Rumor Peak

4.4.3 $\mathcal{R}_0$

I matched similar $\mathcal{R}_0$ values between all three simulations and compared the results. Referring to Table 6, we see similar results in the speed of the infection through the different networks as well as the parameters in $CC(G)$, $L(G)$, and $\langle k \rangle$. Each network has a similar final rumor size for a specific $\mathcal{R}_0$. In the Complete Graph, $\mathcal{R}_0$ can be obtained from the equation: $\mathcal{R}_0 = (\frac{\beta}{\alpha})N$ where β is the infection rate and α is the end infection rate. For discrete time models of Small World and Preferential Attachment, $\mathcal{R}_0$ can be computed as: $\mathcal{R}_0 = \frac{\beta}{\alpha+\beta-\alpha\beta}\langle k \rangle$.

	β	α	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	$R(\infty)$	RD
$\mathcal{R}_0 \approx 1$									
CG = 1	0.0011	0.1	-	-	-	10	2	6	144
SW = 1.3 - 1.45	0.002	0.01	8.3 - 8.8	0.33 - 0.36	2.5 - 2.6	273	3	11	1000
PA = 1.14	0.004	0.01	4	0.1 - 0.2	2.7 - 3	173	3	12	815
$\mathcal{R}_0 \approx 2$									
CG = 2	0.0021	0.1	-	-	-	36	9	46	154
SW = 1.8 - 2	0.003	0.01	8.1 -8.5	0.33 - 0.37	2.5 - 2.6	368	5	32	1000
PA = 2	0.01	0.01	4	0.1 - 0.2	2.9 - 3.1	186	17	53	755
$\mathcal{R}_0 \approx 3$									
CG = 3	0.00312	0.1	-	-	-	34	19	65	153
SW = 2.7 - 3.1	0.005	0.01	8 - 9	0.33 - 0.38	2.5 - 2.7	240	20	70	885
PA = 3	0.005	0.01	9	0.63 - 0.77	2.4	129	16	44	722
PA = 3.2	0.005	0.01	9.7	0.18 - 0.22	2.2	167	27	68	773

Table 6: Comparison of $\mathcal{R}_0$

5 Conclusion

In this paper we examined the dynamics of rumor spreading on three complex social networks: Complete Graph, Small World, and Preferential Attachment. In Complete Graph, we assume that each individual has an equal probability of interaction with every other individual. This is represented by an edge connection between every individual node within the network. Small World is characterized by many small groups of tightly connected individuals with a few inter-community connections. As an example, this can be thought of as you as the connector between your family, your college friends, and your work colleagues. Preferential Attachment is depicted by a few highly connected centralized hubs, which form the bulk of the network. This relation is seen through online sites such as Google or Wikipedia, which present trillions of links to other webpages throughout the Internet network, while other sites, such as a personal blog, are not so heavily linked. The clustering coefficient and the path length are common features shared by all networks. These features can be manipulated to create network structures that mimic real life phenomenon. They ultimately determine the speed, duration, intensity, and final size of a rumor spreading over a network. Simply put, the more connections that exist within a network, the more avenues there are for a rumor to spread. Similarly, the fewer number of steps between individuals to get exposure to a rumor, the faster the rumor will be able to reach more individuals.

We set up two rumor spreading models based on modified SIR contact networks and carried out numerical simulations to compare the different network structures. For all three networks our findings show that increasing the infection rate results in faster, more intense, and larger final rumor sizes. In SW, increasing the rewiring probability leads to an increase in the number of cross graph connections between communities while simultaneously decreases the strong clustering held between small communities. In affect, this SW network becomes more like a PA network, leading to more severe rumors. By increasing the number of adjacent neighbor connections, stronger clustering occurs within the network, which results in slower spreading rumors across the network. For PA, increasing the size of the initial complete graph before running the simulation leads to a greater number of hubs that may emerge within the network. These hubs act more like tightly knit communities, since they exhibit stronger clustering and slower rumor spreading. In affect, this PA network mimics the properties of a SW network. In general, we encounter a spectrum of network variation between SW and PA through tweaking the initial conditions of the simulation.

In summary, we find that the CG model is the most effective network for spreading rumors since it has the maximum number of connections, highest clustering, and lowest path length. In general the mean degree, path length, and clustering are lower in PA than SW, but the hubs in PA are more effective at rumor spreading than the communities in SW. These findings can help future studies construct hybrid networks between CG, SW, and PA for further analysis. Our results can serve as a reference for assessing a network type as well as determining methods for countering the harmful spread of rumors within other networks.

6 Acknowledgments

I would like to thank my thesis advisor Andrea Bruder of the Mathematics Department for taking the time to help and direct me through this process. I would also like to thank the Colorado College Math Department for enlightening me to the world of mathematics.

References

[1] Easley, David, and Jon Kleinberg. Networks, Crowds, and Markets Reasoning about a Highly Connected World. New York: Cambridge UP, 2010. Print.

[2] Wilson, Robin J. Introduction to Graph Theory. New York: Academic, 1972. Print.

[3] Chartrand, Gary. Graphs as Mathematical Models. Boston: Prindle, Weber and Schmidt, 1977. Print.

[4] Keeling, Matt J., and Ken TD Eames. "Networks and epidemic models." Journal of the Royal Society Interface 2.4 (2005): 295-307.

[5] Watts, Duncan J. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton, NJ: Princeton UP, 1999. Print.

[6] Barabasi, Albert-Laszlo. Network Science. S.l.: Cambridge Univ, 2016. Print.

[7] "Dynamic Models in Biology Stephen P. Ellner and John Guckenheimer."Ellner, S.P. and Guckenheimer, J.: Dynamic Models in Biology (eBook and Paperback). N.p., n.d. Web. 23 Sept. 2015.

[8] Robeva, Raina S., and Terrell L. Hodge. Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models. N.p.: n.p., n.d. Print.

[9] Wang, Jing, and Ya-Qi Wang. "SIR Rumor Spreading Model with Network Medium in Complex Social Networks." Chinese Journal of Physics 53.1 (2015): 020702-1.

[10] Doerr, Benjamin, Mahmoud Fouz, and Tobias Friedrich. "Social networks spread rumors in sublogarithmic time." Proceedings of the forty-third annual ACM symposium on Theory of computing. ACM, 2011.

[11] Xia, Lingling, et al. "Modeling and Analyzing the Interaction between Network Rumors and Authoritative Information." Entropy 17.1 (2015): 471-482.

[12] Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." science 286.5439 (1999): 509-512.

[13] Csányi, Gábor, and Balázs Szendroi. "Structure of a large social network." arXiv preprint cond-mat/0305580 (2003)

[14] Nekovee, Maziar, et al. "Theory of rumour spreading in complex social networks." Physica A: Statistical Mechanics and its Applications 374.1 (2007): 457-470.

[15] Wang, Jiajia, Laijun Zhao, and Rongbing Huang. "SIRaRu rumor spreading model in complex networks." Physica A: Statistical Mechanics and its Applications 398 (2014): 43-55.

[16] Zanette, Damian H. "Dynamics of rumor propagation on small-world networks."Physical review E 65.4 (2002): 041908. Zanette, Damian H. "Critical behavior of propagation on small-world networks."Physical Review E 64.5 (2001): 050901.

[17] Ebel, Holger, Lutz-Ingo Mielsch, and Stefan Bornholdt. "Scale-free topology of e-mail networks." Physical review E 66.3 (2002): 035103.

[18] Newman, Mark EJ, Stephanie Forrest, and Justin Balthrop. "Email networks and the spread of computer viruses." Physical Review E 66.3 (2002): 035101.

[19] Daley, D. J., and David G. Kendall. “Stochastic rumours.” IMA Journal of Applied Mathematics 1.1 (1965): 42-55.

[20] Lefevre, Claude, and Philippe Picard. “Distribution of the final extent of a rumour process.” Journal of Applied Probability (1994): 244-249.

[21] Li, Chunru, and Zujun Ma. “Dynamics Analysis of a Delayed Rumor Propagation Model in an Emergency-Affected Environment.” Discrete Dynamics in Nature and Society 501 (2015): 269561.

[22] Sudbury, Aidan. “The proportion of the population never hearing a rumour.”Journal of applied probability (1985): 443-446.

[23] Watts, Duncan J., and Steven H. Strogatz. “Collective dynamics of ‘small-world’networks.” nature 393.6684 (1998): 440-442.

[24] Dictionary, Marriam-Webster. The Merriam-Webster Dictionary. Massachusetts: Merriam-Webster, 2006. Print.

A Appendix: Table Results for Each Network

Inf Rate	$\mathcal{R}_0$	Time of Rumor Peak	% of I(t) at peak	Total Removed	Rumor Duration
0.0011	1	10	2	6	144
0.0021	2	36	9	46	154
0.00312	3	34	19	65	153

Table 7: Complete Graph results

Inf Rate	$\mathcal{R}_0$	Time of Rumor Peak	% of I(t) at peak	Total Removed	Rumor Duration
0.001	.74 - .78	7	2	3	392
0.002	1.3 - 1.45	273	3	11	1000
0.003	1.8 - 2	368	5	32	1000
0.005	2.7 - 3.1	240	20	70	885
0.012	4.6 - 4.8	107	50	90	622

Table 8: Small World Infection Rates Results

λ	$\mathcal{R}_0$	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	$R(\infty)$	RD
1	2.2 - 2.4	6.8 - 7.2	0.46 - 0.49	2.9 - 3.1	330	10	50	1089
2.5	2.7 - 3.1	8 - 9	0.33 - 0.38	2.5 - 2.7	240	20	70	885
5	3.4 - 3.7	10 - 11	0.25 - 0.27	2.2 - 2.3	176	30	79	743

Table 9: Small World λ Results

d	$\mathcal{R}_0$	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	$R(\infty)$	RD
1	1.3 - 1.5	4 - 4.5	0.02 - 0.04	3.1 - 3.5	294	5	21	993
3	2.7 - 3.1	8 - 9	0.33 - 0.38	2.5 - 2.7	240	20	70	885
4	3.3 - 3.5	10 - 10.5	0.41 - 0.45	2.4 - 2.5	218	27	80	787
6	4.6 - 4.8	13.9 - 14.3	0.5 - 0.53	2.2 - 2.3	148	42	83	714

Table 10: Small World d Results

Inf Rate	$\mathcal{R}_0$	Time of Rumor Peak	% of I(t) at peak	Total Removed	Rumor Duration
0.0025	0.8	130	2	6	577
0.003	0.9	197	3	8	768
0.004	1.14	173	3	12	815
0.01	2	186	17	53	755

Table 11: Preferential Attachment Infection Rates Results

λ	$\mathcal{R}_0$	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	$R(\infty)$	RD
2	1.31	3.94	0.11 - 0.2	2.9 - 3.1	276	5	19	809
8	1.42	4.24	0.2 - 0.3	2.8 - 2.9	201	7	25	766
14	1.76	5.26	0.34 - 0.44	2.5 - 2.7	147	10	31	676
25	3	9	0.63 - 0.77	2.4	129	16	44	722

Table 12: Preferential Attachment λ Results

d	$\mathcal{R}_0$	$\langle k \rangle$	$CC(G)$	$L(G)$	TRP	% at peak	$R(\infty)$	RD
1	0.7	2.1	0.02	3.6 - 4.6	148	2	5	535
4	2.6	7.8	0.16 - 0.21	2.3	198	20	54	831
5	3.2	9.7	0.18 - 0.22	2.2	167	27	68	773

Table 13: Preferential Attachment d Results

B Mathematica Code

Listing 1: Mathematica code for Simple Rumor Model and Rumor Spreading Model

```
1
2  Basic Rumor Model in a Closed Population
3
4  Clear[ solution ]
5
6  Here are your set of ordinary differential equations:
7  seqn[\[Lambda]_, \[Sigma]_, \[Delta]_] := s'[t] == (-\[Lambda] i[t] s[t]);
8  ieqn[\[Lambda]_, \[Sigma]_, \[Delta]_] :=
9    i'[t] == (\[Lambda] i[t] s[t] - \[Sigma] i[t] (i[t] + r[t]) - \[Delta] i[
10         t]);
11 reqn[\[Lambda]_, \[Sigma]_, \[Delta]_] :=
12   r'[t] == (\[Sigma] i[t] (i[t] + r[t]) + \[Delta] i[t]);
13
14
15 This command clears your solutions from previous iterations and prepares to solve your system of equations:
16 Clear[\[Lambda], \[Sigma], \[Delta]];
17 solution[{\[Lambda]_, \[Sigma]_, \[Delta]_, initials_, initiali_, initialr_,
18       tfinal_}] :=
19   NDSolve[{seqn[\[Lambda], \[Sigma], \[Delta]],
20       ieqn[\[Lambda], \[Sigma], \[Delta]], reqn[\[Lambda], \[Sigma], \[Delta]],
21       s[0] == initials, i[0] == initiali, r[0] == initialr}, {s, i, r}, {t, 0,
22       tfinal}] // Flatten;
23
24 This command inputs initial conditions for each value within the ODEs and allows you to control each value's input with a
       slider to get a sense for how each parameter affects your model:
25 Manipulate[
26  Plot[Evaluate[{s[t], i[t], r[t]} /.
27     solution[{\[Lambda], \[Sigma], \[Delta], initials, initiali, initialr,
28       tfinal}]], {t, 0, tfinal}], {{\[Lambda], 0.005}, 0,
29   1}, {{\[Sigma], 0.01}, 0, 1}, {{\[Delta], 0.05}, 0, 1}, {{ initials, 1000},
30   0, 10000}, {{ initiali, 1}, 0, 10}, {{ initialr, 0}, 0, 10}, {{ tfinal, 10},
31   10, 1000}]
32
33 ------------------------------------------------------------------------------------------------------------------------

34 Complex Rumor Model
35
36  Installing the NetLogo-Mathematica Link
37
38 To install the NetLogo-Mathematica link, go to the menu bar in Mathematica, click on File and select Install ... In the
       Install Mathematica Item dialog, select Package for Type of item to install, click Source, and select From file ...
      In the file browser, go to the location of your NetLogo installation, click on the Mathematica Link subfolder, and
      select NetLogo.m. For Install Name, enter NetLogo. You can either install the NetLogo link in your user base
      directory or in the system-wide directory. If the NetLogo link is installed in the user base directory, other users
      on the system must also go through the NetLogo-Mathematica link installation process to use it. This option might be
       preferable if you do not have permission to modify files outside of your home directory. Otherwise, you can
       install NetLogo-Mathematica link in the system-wide Mathematica base directory.
39
40 Starting NetLogo
41
42 Once installed, the NetLogo package can be loaded at any time with the following command:
43
44 << NetLogo`
45
```

```
To start NetLogo simply type the following command, and use the file browser to locate the NetLogo parent directory.

NLStart[]

Choose netlogo 5.2 file within the IONTW file on your desktop to activate


I will give you an example of one experiment:

Section 1: Simulations on Preferential Attachment
N = 100
Discrete, 1000+ time steps per simulation
time step size = 1
Run 250 simulations per experiment (about 30 minutes for each)

This command loads the netlogo file for Preferential Attachment:
NLLoadModel[ToFileName[{$NLHome, "models", "Sample Models"}, "pa.nlogo"]];



Experiment 1: Vary for different lambda
inf rate = .005
end inf rate = .01
\mathcal{R}_0 = 1.31
d = 2
<k> = 3.94
CC(G)= .11 − .2
L(G) = 2.9 − 3.1
lambda = 2

This command counts the number of infectious at each discrete time step for 1000 time steps. Then it tells it to repeat
      this process 250 times:
gi1 = Table[
  x = NLCommand["new−network",
    "ask n−of 1 turtles [become−infectious]"];
  NLDoReport["go", "(count turtles with[ infectious ?])", 1000], {x,
   250}]

This command plots out all the values for each of the 250 simulations generated above:
ListLinePlot [gi1]

This command takes the average of all the values from the 250 simulations to make one line:
hi1 = Mean[gi1]

This command counts the number of susceptible at each discrete time step:
gs1 = Table[
  x = NLCommand["new−network",
    "ask n−of 1 turtles [become−infectious]"];
  NLDoReport["go", "(count turtles with[ susceptible ?])", 1000], {x,
   250}]
ListLinePlot [gs1]
hs1 = Mean[gs1]

This command counts the number of removed at each discrete time step:
gr1 = Table[
  x = NLCommand["new−network",
    "ask n−of 1 turtles [become−infectious]"];
  NLDoReport["go", "(count turtles with[removed?])", 1000], {x, 250}]

   ListLinePlot [gr1]

  hr1 = Mean[gr1]

 This command combines the means of each simulation into one line plot:
   ListLinePlot [{hs1, hi1, hr1}]

   Results:
Infection peak: t = 250, 5% of population
```

113 # of removed: 19
114 Infection **Length**: t = 800
