5-2016

Opinion formation about childhood immunization and disease spread on networks

Shan Shan Zhao
The University of Texas Rio Grande Valley

OPINION FORMATION ABOUT CHILDHOOD IMMUNIZATION AND DISEASE SPREAD ON NETWORKS

A Thesis

by

SHAN SHAN ZHAO

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2016

Major Subject: Mathematics

OPINION FORMATION ABOUT CHILDHOOD IMMUNIZATION AND DISEASE SPREAD ON NETWORKS

A Thesis
by
SHAN SHAN ZHAO

COMMITTEE MEMBERS

Dr. Tamer Oraby
Chair of Committee

Dr. Santanu Chakraborty
Committee Member

Dr. Xiaohui Wang
Committee Member

Dr. George Yanev
Committee Member

May 2016

ABSTRACT

Zhao, Shan Shan, Opinion Formation about Childhood Immunization and Disease Spread on Networks. Master of Science (MS), May, 2016, 96 pages, 1 table, 82 figures, 32 references, 18 titles.

People are physically and socially connected with each other. Those connections between people represent two, probably overlapping, networks: biological networks, through which physical contacts occur, or social network, through which information diffuse. In my thesis research, I am trying to answer that question in the context of pediatric disease spread on the biological network between households as well as within them and its relationship with information sharing on the social network of households (parents in that case) via "Information Cascades." I mainly focus on the Erdos-Renyi network model. In particular, I use two different but overlapping Erdos-Renyi networks for the biological and social networks in the model. I am using agent-based stochastic simulations implemented in MatLab to study the modeling results.

DEDICATION

This work is dedicated to my family for their love, support and confidence in me.

ACKNOWLEDGMENTS

The work would not have been finished without the help of my advisor and mentor Dr. Tamer Oraby. His advice and guidance is valuable throughout the whole process. I would also like to thank my committee members. They provided a lot of ideas and opinions to make my work better.

TABLE OF CONTENTS

List of Figures

Page

CHAPTER I

INTRODUCTION TO NETWORKS AND AGENT-BASED MODEL

Human's lives consist of different networks such as the Internet network, social network and so on. Different people have different networks, for example, children have different networks from their parents. In this chapter, we first introduce some basic definitions of networks and then introduce some major types of networks.

1.1 Thesis Chapter Outline

Chapter 1:

In this chapter we briefly introduce basic definitions of the networks and agent-based model.

Chapter 2:

In this chapter we briefly introduce some major types of disease models.

Chapter 3:

We discuss our dual model of disease spread and vaccine opinion on social networks in detail. The dual model consists of two layers of networks: the children's disease network (biological network) and the parents' social network (social network). We will see how parent's vaccinating behaviors are affected by their friends and neighbors.

Chapter 4:

We discuss the results from the simulations. We will make inferences on the results.

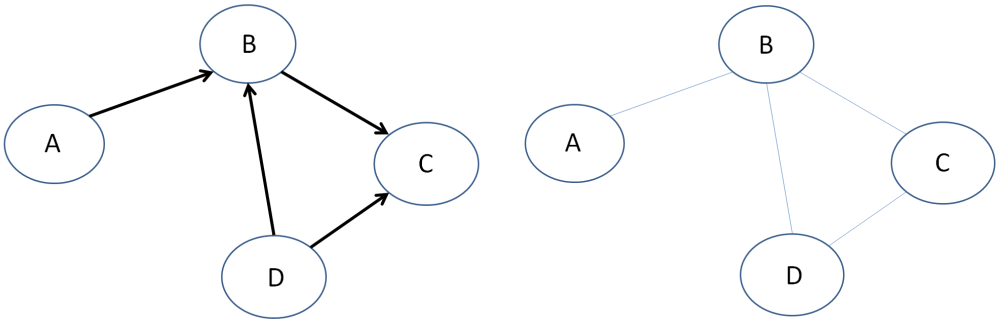


Figure 1.1: Directed and undirected graphs.

Chapter 5:

We discuss some works can be done in the future. They are related to our topics, however, we can’t finish all of them due to the limited time, and computer memory.

1.2 Introduction to Networks

1.2.1 Basic Definitions

Graphs of networks. We denote *graph* as relationships among a set of objects, known as *nodes*, which are connected by links called *edges*. If two nodes are connected with an edge, then they are said to be adjacent. [7]

There are two types of networks. (1) *Directed graph* is a set of nodes connected with a set of *directed edges*, i.e. edges with directions. See Figure 1.1 (a), with edges represented by arrows. (2) *Undirected graph* is a graph without direction, as shown in Figure 1.1 (b).

Paths. In networks, objects and information move from one node to the other through the edges. A *path* in a graph is a sequence of edges that connects two nodes.

Cycle. A *cycle* is a particular kind of path, which looks like a "ring", as shown in Figure 1.2.

Geodesic. The shortest path between two nodes, as shown in Figure 1.3.

Connectivity. A graph is *connected* if every pair of nodes is connected with a path. If a

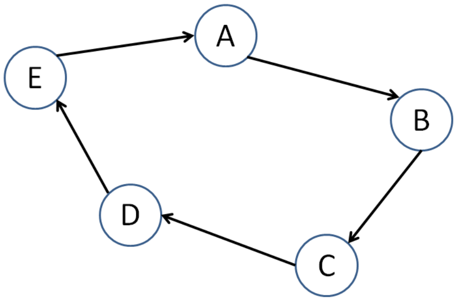


Figure 1.2: A cycle.

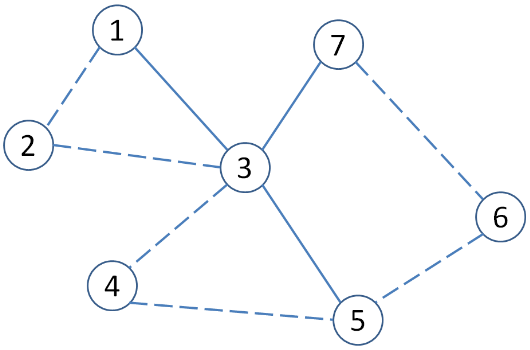


Figure 1.3: An example of geodesic.

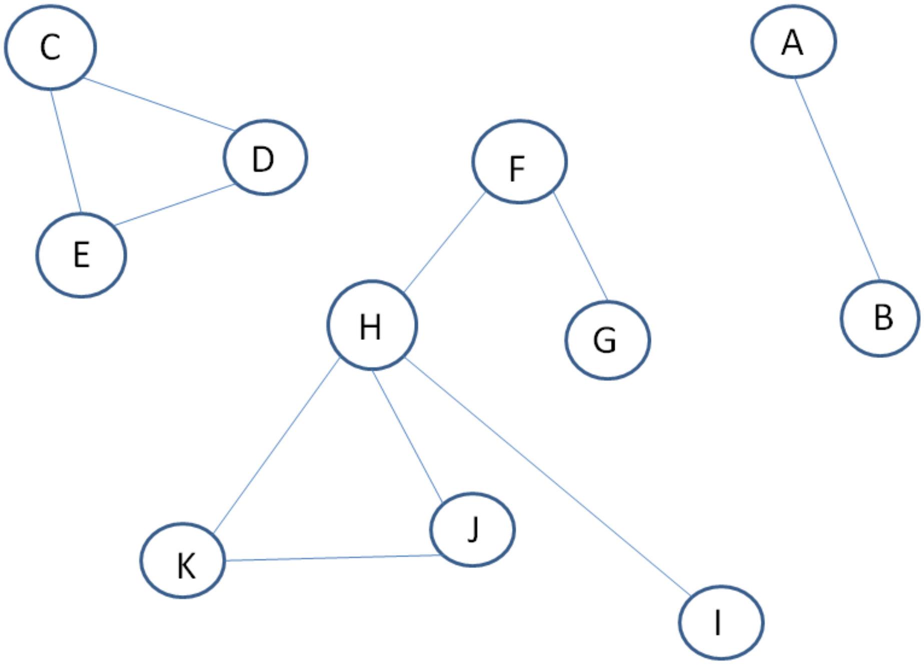


Figure 1.4: An example of three components.

graph is not connected as a whole, it must break into several connected *components*, each node is connected within the group.

In Figure 1.4, we see that the graph consists of three components: one consists of nodes *A* and *B*, another consists of nodes *C,D* and *E*, and the third one consists of three rest of the nodes.

Degree. *Degree* captures the connectedness of a node, i.e. how "connected" is a node. Figure 1.5 shows an example of different degrees in a network.

Closeness. One way to see the realtive distance from one node to others is to use *Closeness*. In other words, the *Closeness* tells us how far the given node is away from the central node, as shown in Figure 1.6. The formula to calculate the closeness centrality is:

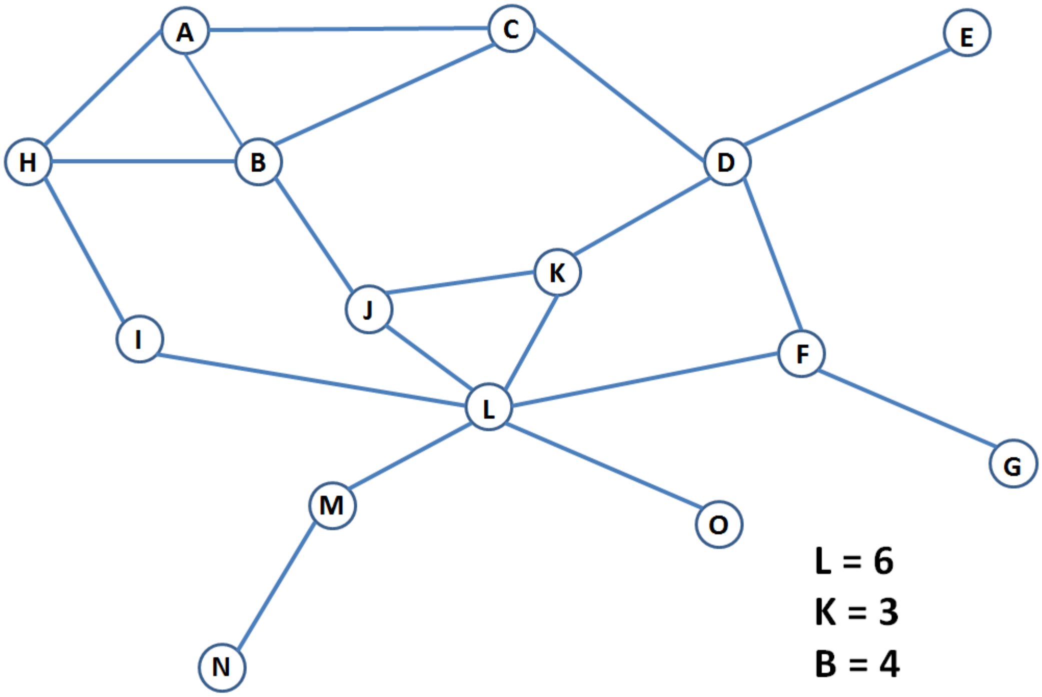


Figure 1.5: An example of degree centrality.

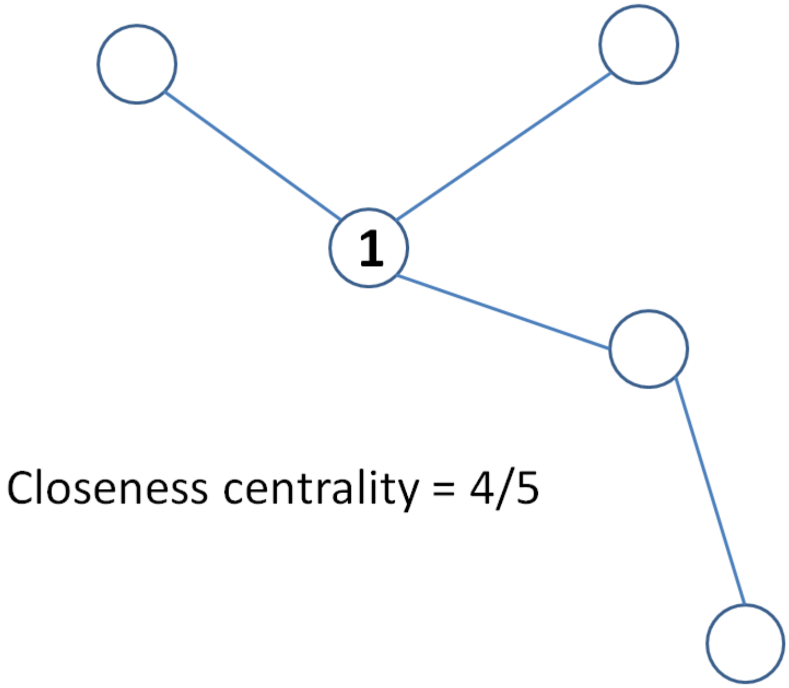


Figure 1.6: An example of closeness centrality.

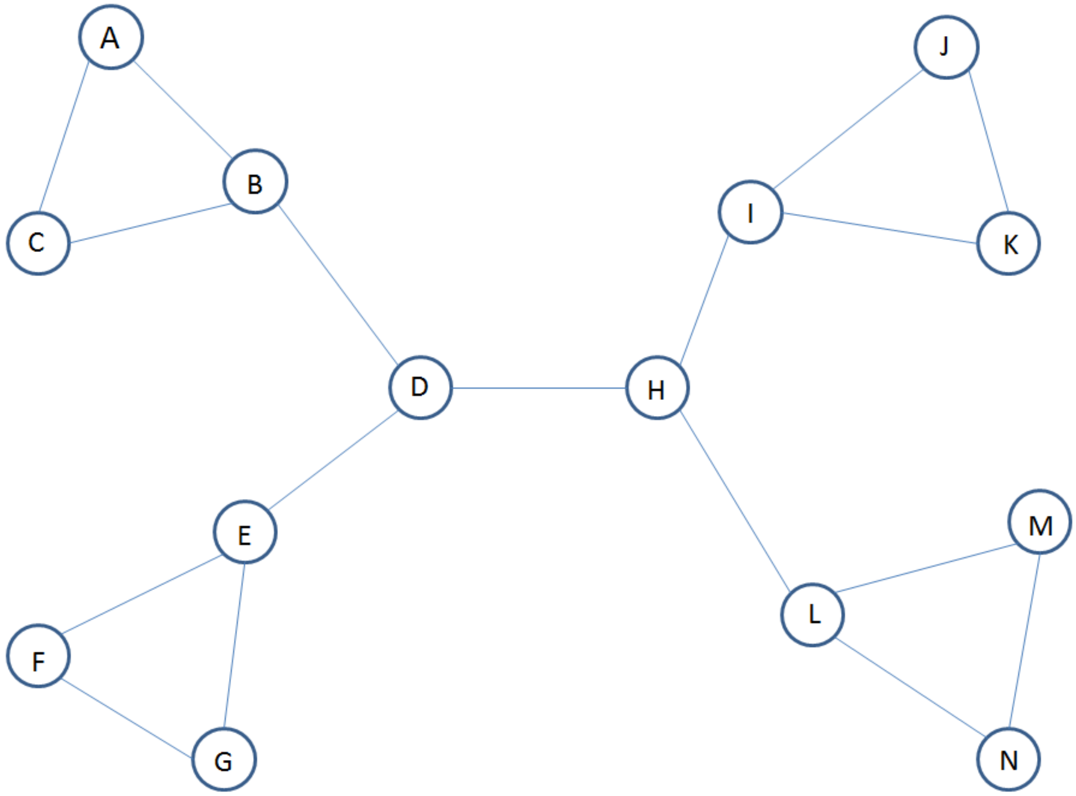


Figure 1.7: A sample network.

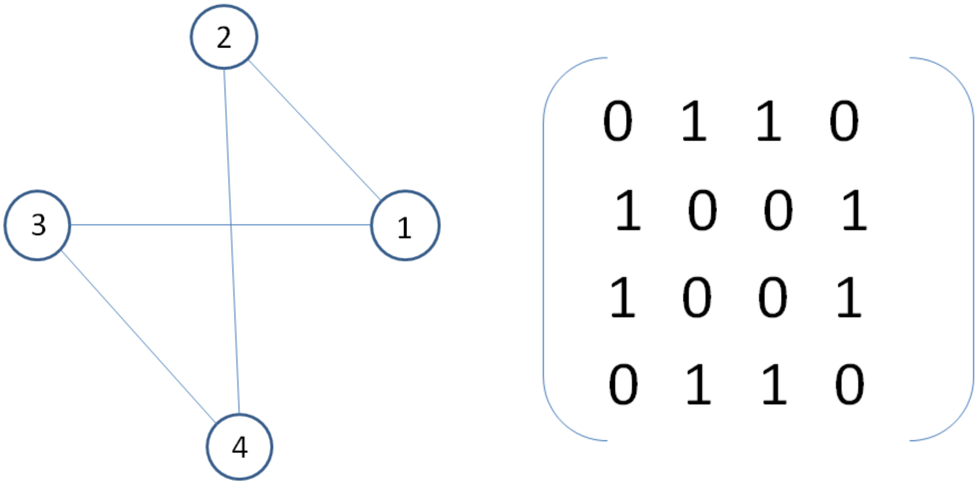
$$\frac{(n-1)}{\sum_j I(i,j)},$$

where I(i,j) is the length of the shortest path (geodesic) between the two nodes i and j.

Closure. If every node in a network is connected to another node, then we call the network a *closure*.

Betweenness. We define the *betweenness* of a network to be the total amount of flow the edge carries. We can consider the *betweenness* as an intermediary or a connector. For the network depicted in Figure 1.7, one can determine the betweenness of each edge as follows:

1. Consider D-H edge. For each node in the left half of the graph, and each node in the right half of the graph, their flow passes through the D-H edge. On the other hand, no flow



$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Figure 1.8: Adjacency matrix example.

passing between pairs of nodes that both lies in the same half uses this edge. As a result, the betweenness of the D-H edge is $7 \times 7 = 49$.

2. Consider B-D edge carries a unit of flow from each node among A,B, and C to each node among E-N. Thus, the betweenness of this edge is $3 \times 11 = 33$.

Adjacency Matrix. The adjacency matrix is a square matrix. The entries of the matrix indicate if pairs of nodes are adjacent or not in the network. In such a case, the adjacency matrix is a 0-1 matrix with zeros on the diagonal since we assume there is no self-loops in our model and is symmetric since we assume it is undirected network.

Figure 1.8 shows a simple example of using an adjacency matrix to represent a network.

We have gone through basic notations of networks. Now let's look at some major types of network models.

1.3 Major Types of Network Models

1.3.1 Erd$\ddot{o}$s - Renyi Model

Definition: We denote *G(n,p)* as a Erd$\ddot{o}$s - Renyi distribution, i.e. all the *undirected* networks on the set of nodes n. Every pair of nodes has a probability $p \in (0,1)$ to be connected.

[1, 9]

The number of edges in a *G(n,p)* graph is a random variable with expected value $\binom{n}{2}p$.

Let $I_{ij} \in 0,1$ be a Bernoulli random variable indicating the presence of edge i,j. For the Erd$\ddot{o}$s - Renyi model, random variables I_ij are independent and

$$I_{ij} = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p \end{cases}$$

Then $\mathbb{E}$[number of edges] = $\mathbb{E}[\sum I_{ij}]$ = $\frac{n(n-1)}{2}p$.

Moreover, using weak law of large numbers, we have $\forall \; \alpha > 0$,

$$\mathbb{P}\left(\left|\sum I_{ij} - \frac{n(n-1)}{2}p\right| \geq \alpha \frac{n(n-1)}{2}p\right) \rightarrow 0,$$

as $n \rightarrow \infty$. Hence, the number of edges is a random variable, but it is tightly concentrated around its mean for large n.

Let D be a random variable that represents the degree of a node. Then:

- D is a binomial random variable with $\mathbb{E}$ [D] = (n-1)p, i.e.

$$\mathbb{P}(D=d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}.$$

- Keeping the expected degree constant as $n \rightarrow \infty$, D can be approximated with a Poisson random variable with $\lambda = (n-1)p$, i.e.

$$\mathbb{P}(D=d) = \frac{e^{-\lambda}\lambda^d}{d!}.$$

That's why the Erd$\ddot{o}$s - Renyi model can be approximated to a Poisson distribution network when n is very large and $0 < p << 1$.

Our research mostly based on the Erd$\ddot{o}$s - Renyi Model. In the following sections, we will generate random networks to simulate the data we want.

Properties:

- Notice that the degree distribution falls off faster than an exponential in d, hence it is **not** a power-law distribution.
- For a given property A (e.g. connectivity), we define a threshold function t(n) as a function that satisfies:

$$\mathbb{P}(A) = \begin{cases} 0 & \text{if } \frac{p(n)}{t(n)} \to 0, \\ 1 & \text{if } \frac{p(n)}{t(n)} \to \infty \end{cases}$$

This makes sense for "monotone or increasing properties", i.e. properties such that if a given network satisfies it, any network satisfies it. When such a threshold function exists, we say that a phase transition occurs at that threshold.

We will prove the threshold function for connectivity.

Theorem: (Erd$\ddot{o}$s and Renyi 1961) A threshold function for the connectivity of the Erd$\ddot{o}$s - Renyi model is

$$t(n) = \frac{\log(n)}{n}.$$

To prove this, it is sufficient to show that when $p(n) = \lambda(n)\frac{\log(n)}{n}$ with $\lambda \to 0$, we have $\mathbb{P}(\text{connectivity}) \to 0$ (and the converse).

However, we will show a stronger result: let $p(n) = \lambda(n)\frac{\log(n)}{n}$.

$$\mathbb{P}(\text{connectivity}) \to 0 \quad \text{if } \lambda < 1, \tag{1.1}$$

$$\mathbb{P}(\text{connectivity}) \to 1 \quad \text{if } \lambda > 1 \tag{1.2}$$

Proof: We first claim (1.1). To show disconnectedness, it is sufficient to show that the probability that there exists at least one isolated node goes to 1.

Let I_{ij} be a Bernoulli random variable defined as

$$I_{ij} = \begin{cases} 1 & \text{if node } i \text{ is isolated,} \\ 0 & \text{otherwise} \end{cases}$$

We can write the probability that an individual node is isolated as

$$q = \mathbb{P}(I_i = 1) = (1-p)^{n-1} \approx e^{-pn} = e^{-\lambda \log(n)} = n^{-\lambda} \tag{1.3}$$

,

where we use $\lim_{n\to\infty}\left(1-\frac{a}{n}\right)^n = e^{-a}$ to get the approximation.

Let $X = \sum_{i=1}^{n} I_i$ denote the total number of isolated nodes. Then, we have

$$\mathbb{E}[X] = n \cdot n^{-\lambda}. \tag{1.4}$$

For $\lambda < 1$, we have $\mathbb{E}[X] \to \infty$. We want to show that this implies $\mathbb{P}(X=0) \to 0$. In general, it is not true. We can't use Poisson approximation since the random variables I_i are dependent. We then shoe that the variance of X is of the same order as its mean.

$$\begin{aligned} var(X) &= \sum_i var(I_i) + \sum_i \sum_{j\neq i} cov(I_i, I_j) \\ &= nvar(I_1) + n(n-1)cov(I_1, I_2) \\ &= n1(1-q) + n(n-1)(\mathbb{E}[I_1 I_2] - \mathbb{E}[I_1]\mathbb{E}[I_2]), \end{aligned}$$

where the second and third equations follow since the I_i are identically distributed Bernoulli random variables with parameter q (dependent).

Now we have

$$
\begin{aligned}
\mathbb{E}[I_1 I_2] &= \mathbb{P}(I_1 = 1, I_2 = 1) = \mathbb{P}(\text{both 1 and 2 are isolated}) \\
&= (1-p)^{2n-3} \\
&= \frac{q^2}{1-p}.
\end{aligned}
$$

Combining the preceding two relations, we obtain

$$
\begin{aligned}
var(X) &= nq(1-q) + n(n-1) + n(n-1)\left[\frac{q^2}{1-p} - q^2\right] \\
&= nq(1-q) + n(n-1) + n(n-1)\frac{q^2 p}{1-p}
\end{aligned}
$$

For large n, we have $q \to 0$, or $1-q \to 1$. Also, $p \to 0$. Hence,

$$
\begin{aligned}
var(X) &\sim nq + n^2 q^2 \frac{p}{1-p} \\
&\sim nq + n^2 q^2 p \\
&= nn^{-\lambda} + \lambda n \log(n) n^{-2\lambda} \\
&\sim nn^{-\lambda} \\
&= \mathbb{E}[X],
\end{aligned}
$$

where $a(n) \sim b(n)$ denotes $\frac{a(n)}{b(n)} \to 1$ as $n \to \infty$.

This implies that

$$
\mathbb{E}[X] \sim var(X) \geq (0 - \mathbb{E}[X])^2 \mathbb{P}(X = 0),
$$

and therefore,

$$
\mathbb{P}(X = 0) \leq \frac{\mathbb{E}[X]}{\mathbb{E}[X]^2} = \frac{1}{\mathbb{E}[X]} \to 0.
$$

It follows that $\mathbb{P}$(at least one isolated node) $\rightarrow 1$ and therefore, $\mathbb{P}$(disconnected) $\rightarrow 1$ as $n \rightarrow \infty$, completing the proof.

Next, let's look at the converse. We next show claim (1.2), i.e. if $p(n) = \lambda(n)\frac{\log(n)}{n}$ with $\lambda > 1$, then $\mathbb{P}$(connectivity) $\rightarrow 1$, or equivalently $\mathbb{P}$(disconnectivity) $\rightarrow 0$.

From Eq. (1.4)m we have $\mathbb{E}[X] = n \cdot n^{-\lambda} \rightarrow 0$ for $\lambda > 1$.

This implies probability of isolated nodes goes to 0. However, we need more to establish connectivity. "Graph is disconnected" is equivalent to the existence of k nodes without an edge to the remaining nodes, for some $k \leq n/2$. Then we have

$$\mathbb{P}(\{1,2,...,k\} \text{ not connected to the rest}) = (1-p)^{k(n-k)},$$

and therefore,

$$\mathbb{P}(\exists \text{ k nodes not connected to the rest}) = \binom{n}{k}(1-p)^{k(n-k)}.$$

Using the union bound, i.e. $\mathbb{P}(\cup_i A_i) \leq \sum_i \mathbb{P}(A_i)$. Then we obtain

$$\mathbb{P}(\text{disconnected graph}) \leq \sum_{k=1}^{n/2} \binom{n}{k}(1-p)^{k(n-k)}.$$

Using the Stirling's formula $k! \sim \left(\frac{k}{e}\right)^k$, which implies

$$\binom{n}{k} \leq \frac{n^k}{\left(\frac{k}{e}\right)^k}$$

in the preceding relation and some "algebra", we obtain

$$\mathbb{P}(\text{disconnected}) \rightarrow 0,$$

Thus, we complete the proof. [1]

1.3.2 Exponential Random Graph Models (ERGM)

The general form of ERGM. An ERGM is a member of the Exponential family, which is a broad family of models that convey many types of data. [16]

Let a random graph consists of n nodes and m edges $\{Y_{ij} : i = 1,...,n; j = 1,...,n\}$ where $Y_{ij} = 1$ if the nodes (i, j) are connected and $Y_{ij} = 0$ otherwise. An ERGM is generated by the probability distribution function:

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa}\right)\exp\{\sum_{A} \eta_A g_A(\mathbf{y})\} \tag{1.5}$$

where (i) $\sum_A$ is all the configurations of A; (ii) η_A is the parameter corresponding to A; (iii) $g_A(\mathbf{y}) = \prod_{y_{ij}\in A} y_{ij}$ is the *network statistic* corresponding to A; $g_A(\mathbf{y}) = 1$ if the configuration is observed in the network $\mathbf{y}$, and is 0 otherwise; (iv) κ is a normalizing constant which ensures Eq. (1.1) is a proper probability distribution. θ is a vector of model parameters associated with $s(g)$ (any statistics of the observed network) and $c(\theta)$ is a normalizing constant.

All exponential random graph models are in the form of Eq.(1.1), which describes a general probability distribution of graphs of n nodes. The probability of observing any particular graph $\mathbf{y}$ in this distribution is given by the equation, ans this probability is dependent both on the statistics $g_A(\mathbf{y})$ in the network $\mathbf{y}$ and on the various non-zero parameters η_A for all configurations A in the model. Configurations might include reciprocate triads and so on, so the model enables us to examine a variety of possible structural regularities.

We have to talk about the dependence assumptions before we move on. The reason that the dependence assumptions are critical is they have the consequence of picking out different types of configurations as relevant to the model. Note from (ii) above, parameters are zero whenever variables in a configuration are conditionally independent of each other. In other words,the only configurations that are relevant to the model are those in which all possible ties in the configuration are mutually contingent on each other. [16]

We take Bernoulli graphs, which is the simplest dependence assumption, as an example.

Bernoulli random graph distributions are generated when we assume that edges are independent, for instance of they occur randomly according to a fixed probability α (see Erd$\ddot{o}$s and Renyi, 1959; Frank and Nowicki, 1993). The dependence assumption is simple in this case: all possible distinct ties are independent of one another. We noted that the only configurations relevant to the model are those in which all possible ties in the configuration are conditionally dependent on each other. When all possible ties are independent, the only possible configurations relate to single edges $\{Y_{ij}\}$. So from Eq. (1.1) the general model is:

$$\mathbb{P}(\mathbf{Y}=\mathbf{y}) = \left(\frac{1}{\kappa}\right)\exp\left(\sum_{i,j}\eta_{ij}y_{ij}\right)$$

Note that compared to Eq.(1.1) every set A comprising a single possible edge $\{Y_{ij}\}$ is a configuration in this model, and there is a parameter η_{ij} for each of these configurations. The network statistic $g_A(\mathbf{y}) = g_{ij}(\mathbf{y}) = y_{ij}$ tells us whether that configuration is observed or not. If we impose a homogeneity assumption so that the effect for each tie is identical we equate parameters such that $\eta_{ij} = \theta$ for all i and j, hence,

$$\mathbb{P}(\mathbf{Y}=\mathbf{y}) = \left(\frac{1}{\kappa}\right)\exp(\theta L(\mathbf{y}))$$

where $L(\mathbf{y}) = \sum_{i,j} y_{ij}$ is the number of arcs in the graph $\mathbf{y}$ and the parameter θ is related to the probability of a tie being observed. The parameter θ is usually called the *edge* or *density* parameter. [2, 18]

A theorem by Hammersly and Clifford claims that any network model can be expressed in the exponential family with counts of graph statistics. Take the Erd$\ddot{o}$s - Renyi model as an example. Let p be the probability of a link and L be the number of links in the network g. The probability of g in the Erd$\ddot{o}$s - Renyi model network is:

$$\begin{aligned}\mathbb{P}[(g)] &= p^{L(g)}(1-p)^{\frac{n(n-1)}{2}-L(g)}\\ &= \left[\frac{p}{(1-p)}\right]^{L(g)}(1-p)^{\frac{n(n-1)}{2}}\\ &= \exp\left[\log(\frac{p}{(1-p)})L(g)-\log(\frac{1}{(1-p)})\frac{n(n-1)}{2}\right]\\ &= \exp[\beta_1 s_1(g)-c]\end{aligned}$$

where $s_1(g)$ is the statistics of the graph, i.e. the number of links in the graph.

1.3.3 Power Laws

An Example. When researchers studied the distribution of links on the Web, they found something unexpected. In studies over many different Web snapshots, taken at different times in the Web's history, the similar phenomenon being formed for many times is that the fraction of Web pages that have n in-links is approximately proportional to $1/n^2$. (More precisely, the exponent on n is generally a number of slightly larger than 2.)

The finding is not what researchers expected: a normal distribution. Why is this so different? The crucial point is that $1/n^2$ decreases much more faster as n increases, so pages with very large number of in-links are much more common than we'd expect with a normal distribution. For example, $1/n^2$ is only one in a million for $n = 1000$. A function that decreases as n to some fixed power, such as $1/n^2$ in the present case, is called a *power law*. When measure the fraction of items having value n, it's possible to see very large values of n.

This provides a quantitative form for one of the points we made initially: popularity seems to exhibit extreme imbalances, with very large values likely to arise. This is what we expect with our intuition about the Web, where there are certainly a reasonable large number of extremely popular pages. One may see similar power laws when measuring popularity in many other domains as well: for example, the fraction of telephone numbers that receive m calls per day is roughly proportional to $1/m^2$; the fraction of books that are bought by v people is roughly proportional to $1/v^3$; the fraction of scientific papers that receive w citations in total is roughly proportional to

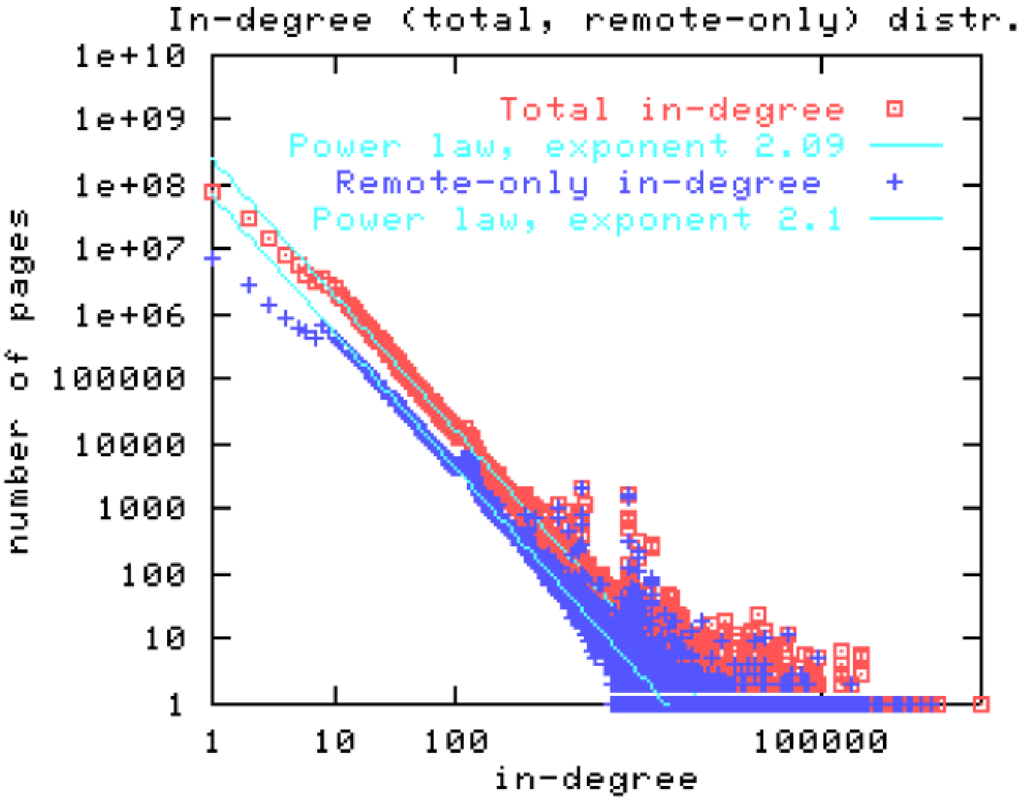


Figure 1.9: A power law distribution (such as this one for the number of Web page in-links, from Broder et al.) shows up as a straight line on a log-log plot.[7].

$1/w^3$.

Indeed, just as the normal distribution is widespread in the natural sciences, power laws seem to dominate in cases where the quantity being measured can be viewed as a type of popularity. Hence, if you are handling data of this sort, one of the first things that's worth doing is to test whether it's approximately a power law $1/n^c$ for some c, and if so, to estimate the exponent c.

There's a simple way that provides a quick test for whether a dataset exhibits a power-law: Suppose you want to know whether the equation $f(k) = a/k^c$ approximately holds, for some exponent c and constant of proportionality a and we denote $f(k)$ be the fraction of items that have value k. Then, if we write this as $f(k) = ak^{-c}$ and take the logarithms of both sides, we get:

$$\log f(k) = \log a - c \log k$$

That means if we have a power-law relationship, and we plot log $f(k)$ as a function of log k, then we should see a straight line: $-c$ will be the slope. Such a "log-log" plot thus provides a quick way to see if one's data exhibits an approximate power-law because it is easy to see if one has an approximately straight line. For example, Figure 1.9 does this for the fraction of Web pages with k in-links.

Rich-Get-Richer Models. Just as normal distributions arise from many independent random distribution averages, the power laws arise from the combination of information cascades and correlated decisions among population.

It is actually a research question that provides a satisfactory model of power laws starting from simple models of individual decision-making (just like information cascades). We will build our model from the observable consequences of the decisions made by individuals in the process of cascades. Here, we will assume that people have a tendency to copy the decisions of people who act before them.

Based on this idea, we create a simple model of links among Web pages.

(1) Pages are created in order, and named 1,2,3,...,n.

(2) When page j is created, it produces a link to an earlier Web page according to the following probabilistic rule (controlled by a probability p between 0 and 1):

(a) With probability p, page j chooses a page i uniformly at random from earlier pages, and creates a link to this page i.

(b) With probability $1-p$, page j chooses a page i uniformly at random from earlier pages, and creates a link to the page that i points to.

(c) This creates a single link from page j; one can repeat this process to create multiple, independently generated links from page j. (To keep things simple, we assume that each page creates just one link.)

Part (2b) of this process is the key: after finding a random earlier page i in the population, the author of page j does not link to i, but instead copies the decision made by the author of page i — linking to the same page that i did.

If we run it for quite a long period of time, the fraction of pages with k in-links will be distributed approximately according to a power law $1/k^c$, where the value of c depends on the choice of p. The process goes in an intuitive direction: as p gets smaller, so that copying becomes more frequent, the exponent c gets smaller as well, making one more likely to see extremely popular pages.

Since the copying is an implementation of the following "rich-get-richer" dynamics, i.e. when copying the decision of a random earlier page, the probability that you end up linking to some page ℓ is directly proportional to the total number of pages that currently link to ℓ. Thus, an equivalent way to write phrase (2b) is:

With probability $1 - p$, page j chooses a page ℓ with probability proportional to ℓ's current number of in-links, and creates a link to ℓ.

Why do we call this a "rich-get-richer" model? Because the probability that page ℓ experiences an increase in popularity is directly proportional to ℓ's current popularity. This phenomenon is also known as *preferential attachment*, in the sense that links are formed "preferentially" to pages that already have high popularity. Essentially, the more well-known someone is, the more likely you are to hear their name, and consequently the more likely you are knowing about them.

The rich-get-richer also predicts that a page's popularity grows at a rate proportional to its current value, and exponentially with time. Indeed, rich-get-richer models suggests a basis for power laws in a wide range of settings. For example,if we assume that cities are formed at different times, and, once formed, a city grows in proportion to its current size simply as a result of people having children, then we have almost precisely the same rich-get-richer model. [7]

$$\begin{array}{cc|c|c|}
 & & \multicolumn{2}{c}{w} \\
 & & A & B \\
\hline
\multirow{2}{*}{v} & A & a, a & 0, 0 \\
\hline
 & B & 0, 0 & b, b \\
\hline
\end{array}$$

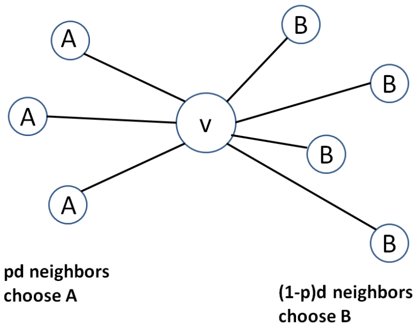


Figure 1.10: The cascading example.

1.4 Information Cascades

1.4.1 Introduction

Diffusion Model of a Network Before moving to information cascades, we first look at a simplified diffusion model. In a social network, we have a situation in which each individual has a choice between two options, *A* and *B*. If two individuals *v* and *w* are linked by an edge, then there is a possibility that their choices match. We represent this by a game in which *v* and *w* are the players and *A* and *B* are the options. The payoffs are as follows:

1. if *v* and *w* both choose *A*, they each get a payoff of $a > 0$.
2. if they both choose *B*, they each get a payoff of $b > 0$.
3. if they choose opposite options, they each get a payoff of 0.

We use a matrix in Figure 1.10 to show the situation above. The point is each *v* is copying choices of its neighbors. The basic question faced by *v* is : since some of its neighbors choose A, and some choose B, what should *v* choose to get maximum payoff? Clearly, this depends on the number of each of the two choices, and payoff values *a* and *b*. Suppose *p* fraction of *v*'s neighbors choose A, and $(1-p)$ fraction choose B, i.e. if *v* has total *d* neighbors, then *pd* neighbors will choose A and $(1-p)d$ neighbors will choose B, as shown in Figure 1.9.

If v chooses A, it gets payoff pda; if chooses B, it gets payoff $(1-p)db$. As a consequent, A is the better choice if

$$pda \geq (1-p)db$$

or we can re-write it as

$$p \geq \frac{b}{a+b}$$

This intuitively makes sense: when $p \geq \frac{b}{a+b}$ is small, then choice A is more enticing; on the other hand, if $p \geq \frac{b}{a+b}$ is large, the opposite holds: choice B is more attractive.

Within social network, people are connected with each other. It is easy for them to influence each other's decisions and behaviors. There are so many scenarios in which people are influenced by others: color preference, the political opinions, the opinions for some social issues such as vaccination. What we're interested in is to explore how the influences happen. We show such a situation by an simple example. [7]

An Example Experiment. Let's consider an experiment take place in a room with a large group of randomly selected people, A box was put in front of the room with three balls in it. The people are told that there is a 50% chance the box contains two red balls and one blue ball, and a 50% chance the box contains two blue balls and one red ball. We call the first case "majority-red", and the second case "majority-blue".

Then the group of people start drawing a ball from the box without showing it to the rest of the people. Each individual guesses whether the box is majority-red or majority-blue and tells the guess to the others. For those people who have not yet had their turn to draw, they do have to hear the guesses.

We want to see what will happen with each individual's decisions. Obviously, the first person just tell exactly what he draws from the box: if he draws a red ball, it is reasonable to guess that the box is majority-red; and if he draws a blue ball, it is more reasonable to guess that the box is majority-blue. As for the second person, if he/she sees the same color that the first person

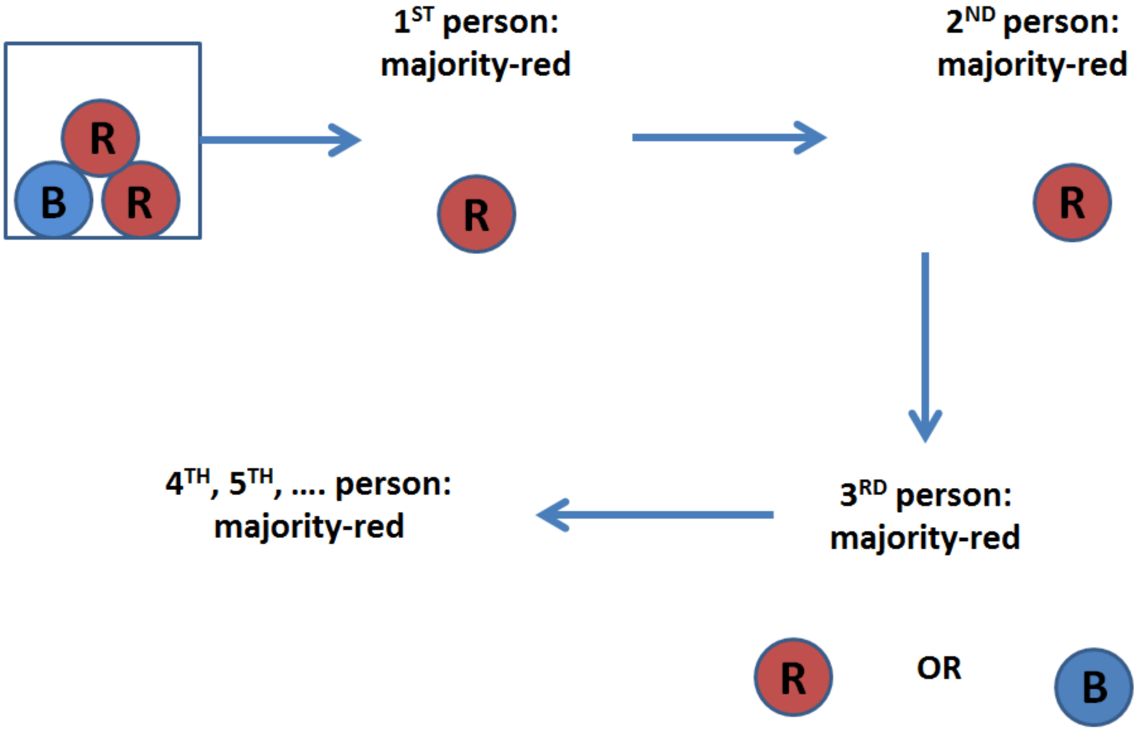


Figure 1.11: First case of the example experiment.

told, then he/she should guess the same color as well; on the other hand, if he/she sees the opposite color, we assume that he/she will break the tie by guessing the color he/she saw. Consequently, no matter what color the second person draws, the guess also conveys perfect information.

The decisions for the first two students are straightforward. Things become more interesting from the third person. If the first two guesses have been the same – blue, for example – and the third person draws red. Since we assume the first two guesses convey perfect information, then the third person should guess that the box is majority-blue, no matter what is his own information. The point is the third person should guess the same color as the first two people do, regardless of which color he actually draws from the box. Figure 1.11 shows how this information cascades phenomenon happens.

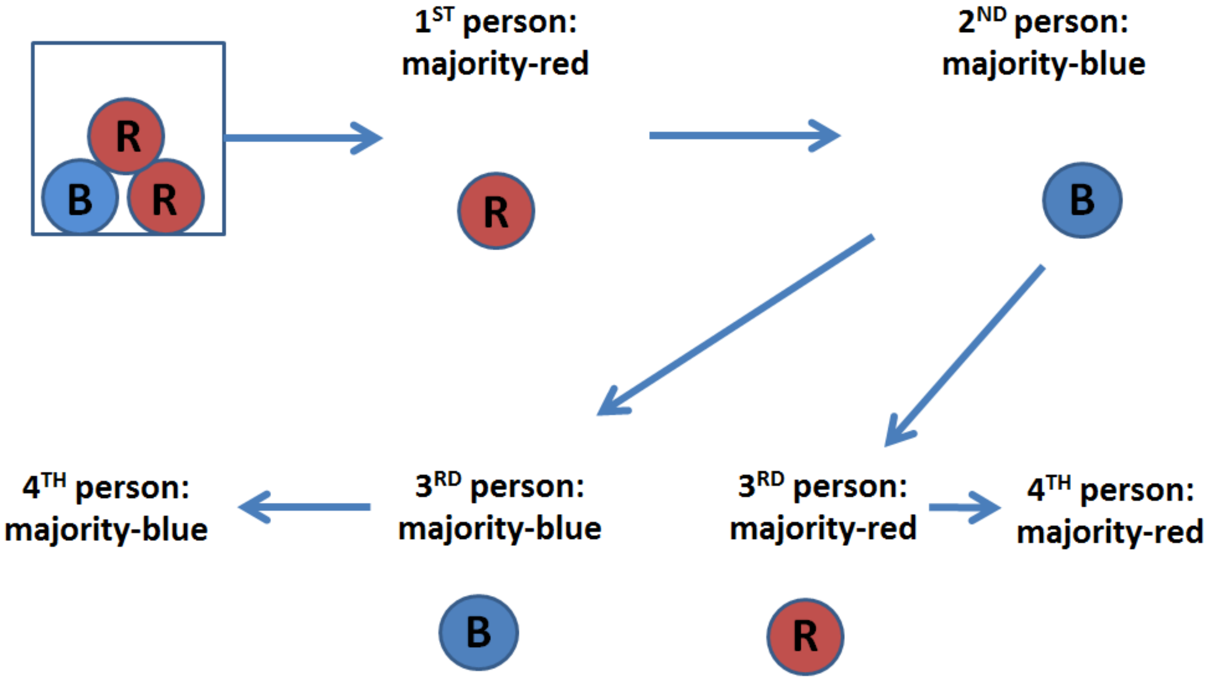


Figure 1.12: Second case of the example experiment.

This is a typical example to show how information cascades happen in a social network. As a matter of fact, we can consider the event above as a conditional probability problem. To quantify the probability of the event, let A be the event that the box in the experiment is majority-blue, and B be the event that the ball drawn is blue. Then the conditional probability of A given B is:

$$Pr[A|B] = \frac{Pr[A \cap B]}{Pr[B]}.$$

Since $A \cap B$ and $B \cap A$ are the same, the conditional probability of B given A is

$$\begin{aligned} Pr[B|A] &= \frac{Pr[B \cap A]}{Pr[A]} \\ &= \frac{Pr[A \cap B]}{Pr[A]} \end{aligned}$$

By simple substitution, we have

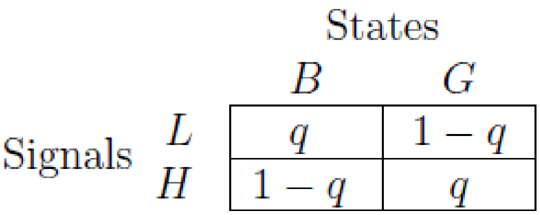
$$\begin{aligned} Pr[A|B] \cdot Pr[B] &= Pr[A \cap B] \\ &= Pr[B|A] \cdot Pr[A] \end{aligned}$$

Hence,

$$Pr[A|B] = \frac{Pr[A] \cdot Pr[B|A]}{Pr[B]}. \tag{1.6}$$

Equation (1.1) can also be called as the *Bayes' Rule*.

When generalizing the case, we need to formulate a model that represents all the general situations. Consider a group of people (numbered 1,2,3,...) who will sequentially make decisions. We describe the decision as a choice between *accepting* or *rejecting* some option: whether to adopt a new technology device, eat in a new restaurant, or vote for a particular political candidate.



		States	
		B	G
Signals	L	q	$1-q$
	H	$1-q$	q

Figure 1.13: The probability of receiving a low or high signal, as a function of the two possible states of the world (*G* or *B*).

We denote the two possible states as *G*, representing the state in which the option is a good idea, and *B*, representing the state in which the option is a bad idea. Suppose each individual knows that the initial random event that placed the world into state *G* or *B* placed it into state *G* with probability *p*, and into state *B* with probability *1-p*. This will serve as the prior probabilities of *G* and *B*; in other words, $Pr[G] = p$, and hence $Pr[B] = 1 - Pr[G] = 1 - p$.

We assume that before any decisions are made, each individual gets a *private signal* about whether accepting is a good idea or a bad idea. The private signal represents the information the person happens to know, beyond just the prior probability *p* that accepting the option is a good idea. There are two possible signals: a *high signal* (denoted *H*), suggesting that accepting is a good idea; and a *low signal* (denoted *L*), suggesting that accepting is a bad idea. If accepting is a good idea, then high signals are more frequent than low signals: $Pr[H|G] = q > \frac{1}{2}$, while $Pr[L|G] = 1 - q < \frac{1}{2}$. Similarly, if accepting the option is a bad idea, then low signals are more frequent: $Pr[L|B] = q$ and $Pr[H|B] = 1 - q$, for this same value of $q > \frac{1}{2}$. The idea above is summarized in the table in Figure 1.12.

Using Bayes' Rule, it's easy to reason directly about an individual's decision when they get a sequence *S* independently generated signals consisting of *a* high signals and *b* low signals. We can derive the following facts:

(i) the posterior probability Pr[G|S] is greater than the prior Pr[G] when $a < b$;

(ii) the posterior pr[G|S] is less than the prior Pr[G] when $a > b$; and

(iii) the two probabilities Pr[G|S] and Pr[G] are equal when $a = b$.

As a result, individuals should accept the option when they get more high signals than low signals, and reject it when they get more low signals than high signals; they are indifferent when they get the same number of each.

To apply Bayes' Rule, we write

$$Pr[G|S] = \frac{Pr[G] \cdot Pr[S|G]}{Pr[S]}, \tag{1.7}$$

where S is a sequence with a high signals and b low signals. To compute $Pr[S|G]$ in the numerator, we note that since the signals are generated independently, we can simply multiply their probabilities: this gives us a factor of q and b factor of $(1-q)$, and so $Pr[S|G] = q^a(1-q)^b$.

To compute $Pr[S]$, we consider that S can arise if the option is a good idea or a bad idea, so

$$Pr[S] = Pr[G] \cdot Pr[S|G] + Pr[B] \cdot Pr[S|B] \tag{1.8}$$

$$= pq^a(1-q)^b + (1-p)(1-p)^a q^b. \tag{1.9}$$

Plugging this back into the Equation 1.2, we get

$$Pr[G|S] = \frac{pq^a(1-q)^b}{pq^a(1-q)^b + (1-p)(1-q)^a q^b}. \tag{1.10}$$

This is the general equation of information cascades and it is the basic idea of our mode when we try to find the probability of parents who will vaccinate their children. [7]

1.4.2 Information Cascades on Social Networks

The social network analysis explains why information of neighbors about the vaccinating behaviors within the social network is critical to the decision-making behaviors of parents, both for vaccine acceptance and for vaccine refusal. The logic behind it is simple: if an unvaccinated individual is more likely to be in contact with other unvaccinated individuals than with vaccinated individuals, then clusters of susceptible individuals will form a subpopulation in which the disease can spread and cause local outbreaks.

Many reasons can cause vaccination exemption such as worries about the safety and usefulness of vaccines, religious beliefs and so on. What is necessary for such a cluster of susceptible individuals to form is a process that leads to clustering of individuals who share negative opinions about vaccination. Individuals with a negative opinion towards vaccination will be more likely to be in contact with other individuals holding the same opinion.

That's what we called "information cascades" in the epidemic disease network. Information cascades happen when parents of the households want to decide whether to vaccinate their children. It's reasonable to speculate that parents from one household willing to vaccinate their children are more likely to cluster with other parents holding the same opinion than to cluster with parents that are unwilling to vaccinate their children. Through social contact network, children are connected with others. As a consequence, their parents are connected with each other through a new social network generated by the network of children's. When parents from one household meet and talk with each other, or learn from the Internet, other parents' vaccinating behavior will affect their children. Here, the information cascades start to affect parents' behavior.

1.5 The Agent-Based Model

1.5.1 Introduction

Agent-based model is a powerful simulation model that has become more and more popular in the last several years. In the agent-based model, one system is modeled as a group of autonomous decision-making nodes called agents. Each agent can individually assesses its situation and makes its own decisions on the basis of some rules. Agents may present various behaviors. At the simplest level, an agent-based model consists of a group of agents and the relationship between them. But even a simple agent-based model can exhibit complex patterns of behaviors and provide valuable information that enable people to know more about the dynamics of the real world.

The benefits of agent-based model can be concluded in three aspects:

Agent-based model can capture emergent phenomena. Emergent phenomena come up from the interactions of individuals. An emergent phenomenon is very difficult to understand and predict, sometimes it is counterintuitive. A simple example can show the counterintuitive characteristic: a game that is easy to play with a group of 10-40 people. One asks each member of the audience to randomly select 2 individuals, person A and person B. One then asks the two people to move so they always keep A between them and B, so A is their protector from B. Every audience in the room will walk around in a random way and will soon begin to ask why they are doing this. On e then asks them to move so that they keep themselves in between A and B. The results are striking: almost instantaneously the whole room will implode, with everyone clustering in a tight knot. This example shows how small changes in some simple individual rules can have a dramatic impact on the group behavior, and how intuition can be a very poor guide to the outcomes.

That's why we usually use agent-based model when individual behaviors are nonlinear and are difficult to be presented with differential equations; or when individual behavior exhibits memory, non-markovian behavior, or temporal correlations.

Agent-based model provides a natural description of the real world. Agent-based model is most natural for describing and simulating a system composed of "behavioral" individuals. For example, it is more natural to describe how consumers move in a supermarket than to come up with the differential equations to show the dynamics of the consumers' behaviors.

One can use agent-based model when the individual behavior is complex. Of course, almost everything can be shown with equations, but the complexity of differential equations increase nearly exponentially as the complexity of human's behaviors increase. Agent-based model is often the most appropriate way of describing what is actually happening in the real world. The stochasticity applied to the agents' behaviors, with agent-based model, the randomness can be applied to the right places like an error term to the equation.

Agent-based model is flexible. Agent-based model not only can be observed by multiple dimensions, but also can change levels of description and aggregation, i.e. it is easy to work with single agents or aggregate agents.

1.5.2 Applications

Agent-based model has many applications in real world, such as in the social, political, economics, and epidemic diseases. It has four major areas of applications: the flow simulation, mainly dealing with traffic and customer flow management. The market simulation, mainly works for the stock market for the strategic simulations. The organization simulation, used to predict and evaluate operational risk and organizational design. The diffusion simulation, used to study the dynamic of the epidemic diseases.

The diffusion simulation is most attractive to us. We first use a simple example to illustrate the value of the agent-based model in the social network analysis. Assume a new product's value V depends on the number of its users N in a total population of N_T potential buyers, according to the following function

$$V(N) = V(\rho) = \frac{(1+\theta^d)\rho^d}{\rho^d+\theta^d}, \tag{1.11}$$

where ρ is the fraction of the population that has purchased the product, θ is a characteristic value (here we fix $\theta = 0.4$), and d id an exponent that determines the steepness of the function (here we fix $d = 4$). $V(N) = 0$ when there is no buyers at all and is maximum ($V(N) = 1$) when all the population has purchased the product. θ acts as a threshold: when the buyer base approaches 40% of the population, the value curve takes off. Customers may not know the exact number of people who have already bought the product in the population, but they can estimate the fraction of buyers in their neighborhood. If we assume that each person is connected to n people in the population, we can define person k's estimate of the fraction of buyers in the total population as

$$\hat{p}_k = n_k/n$$

, where n_k is the number of k's neighbors who have bought the product. Then the estimated value $\hat{V}_k$ of the product is given by

$$\hat{V}_k = V(\hat{\rho}_k) = \frac{(1+\theta^d)\hat{\rho}_k^d}{\hat{\rho}_k^d + \theta^d}. \tag{1.12}$$

If person k is connected to everyone else, $\hat{V}_k$ is identical to V. However, that is highly unlikely. With the average fraction of buyers $\rho = N/N_T$, the perceived value is:

$$V(\rho) = \frac{(1+\theta^d)\rho^d}{\rho^d + \theta^d}. \tag{1.13}$$

The resulting differential equation is

$$\partial_t N = V(\rho)(N_T - N), \tag{1.14}$$

which is equivalent to

$$\partial_t \rho = V(\rho)(1-\rho). \tag{1.15}$$

Now we try to solve the same problem using agent-based approach. The first transformation is from Eq. 1.11 to individual transition probabilities, where each agent has a transition probability given by the rate of Eq. 1.11. That is to say, for each agent who is not a buyer, the probability of becoming one is equal to $V(\rho)$ per time unit. The meaning of this model is that each agent acts individually but has perfect knowledge of how many buyers there are in the total population.

Figure 1.13 shows how the fraction of buyers increases in time for a population of 100 agents. This curve is almost indistinguishable on average from that obtained with the system dynamics approach, except when the initial population of buyers is very low, in which case the takeoff can be significantly slower in the agent-based description in some simulations because of significant fluctuations in the early part of the simulation. These fluctuations reflect the individual decision-making by agents as opposed to an average global flow. Yet, on average, one obtains the same dynamics as the flow model. Things become quite different, however, as soon as one starts assuming that the agents estimate the fraction of users from the fraction of their neighbors who are

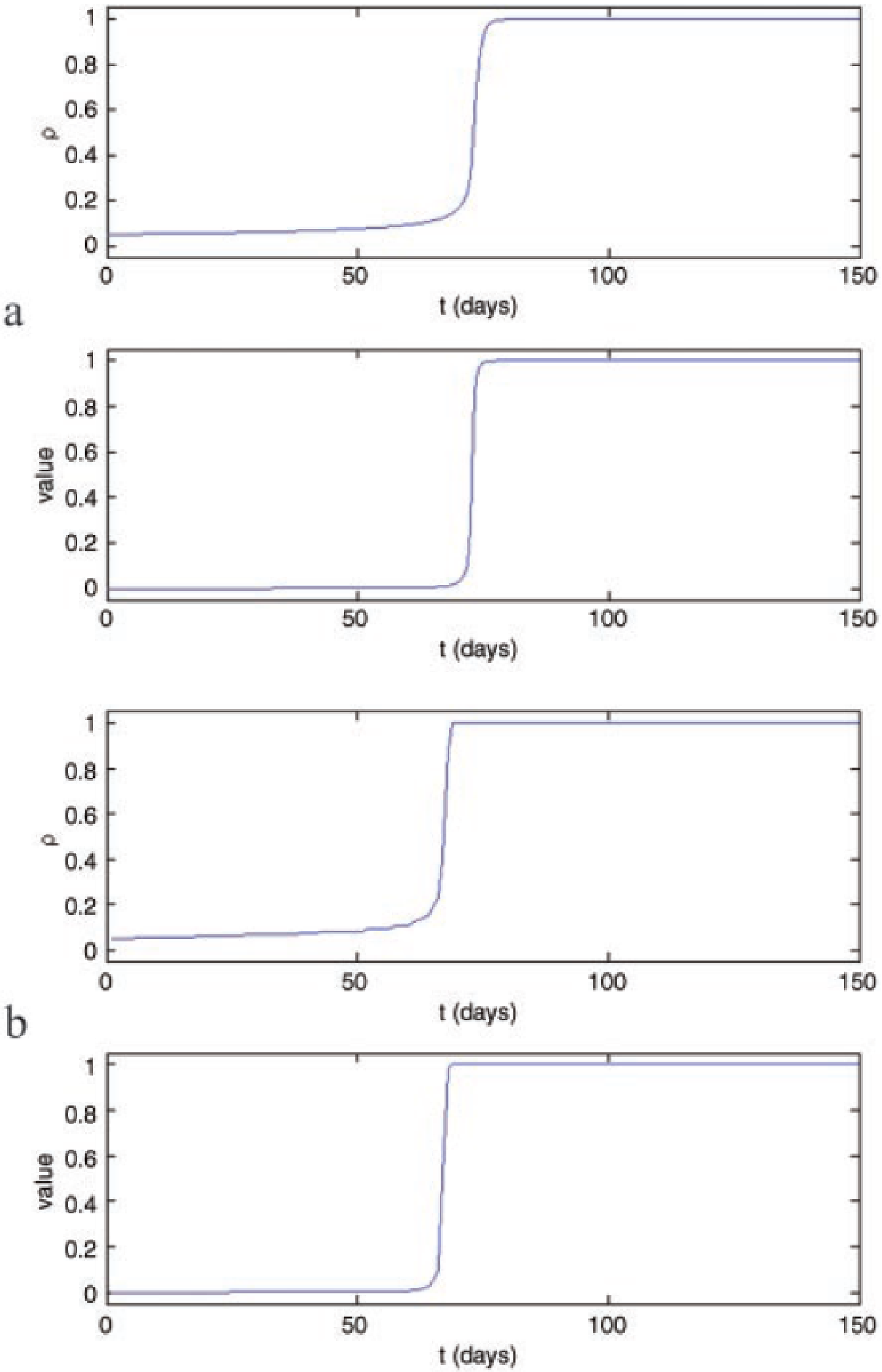


Figure 1.14: Agent-based model example-1.

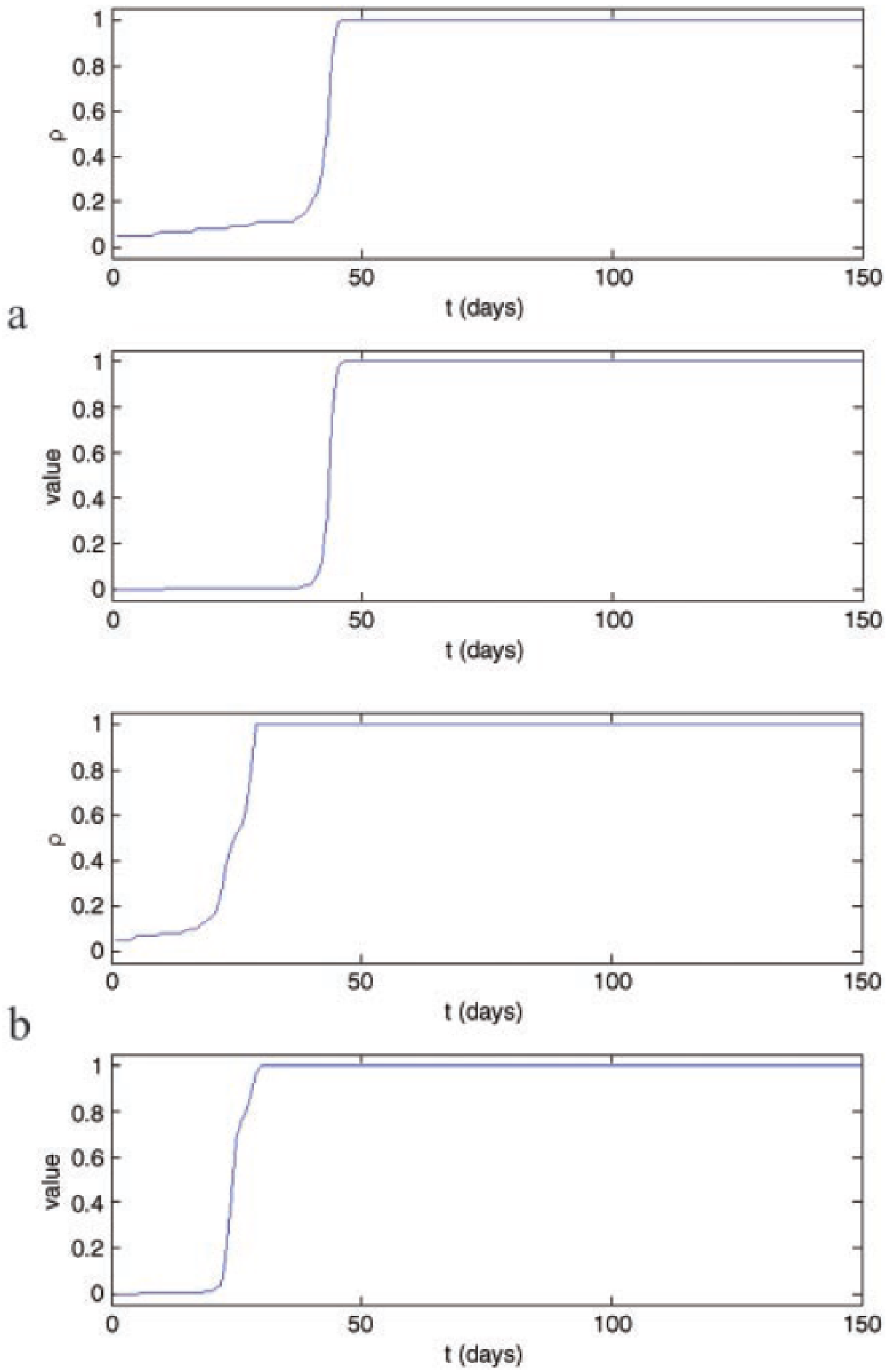


Figure 1.15: Agent-based model example-2.

buyers. Let us assume that each individual in the population has exactly $n = 30$ neighbors. Let's us consider two cases:

1. Those 30 neighbors are selected randomly in the population.

2. There is clustering in the structure of social interactions in that a neighbor of a neighbor is likely to be a neighbor. Assume that the population s divided into two subpopulations of equal size. The probability that two individuals from the same subpopulation are neighbors is equal to $P = 0.5$, and the probability that two individuals from different subpopulations are neighbors is equal to 0.1. In a population of 100 agents, the average total number of neighbors of any given node is $0.5 \cdot 50 + 0.1 \cdot 50 = 30$. We assume the initial 5% of buyers is within one of the subpopulations.

The second case introduces localization in the dynamics: a person interacts only with her neighbors and there are few long-range interactions and little global mixing. In the first case, one might expect to observe a dynamics similar to the system dynamics model, whereas the dynamics in the second case could be quite different. It appears that even in the first case the resulting dynamics is different from the mean-filed dynamics (Figure 1.14a), but the second case leads to potentially dramatically different results, as can be seen in Figure 1.14b. product purchase is a lot faster with clustering, even when the initial buyer population is located entirely within one cluster.

This simple example shoes not only how useful agent-based model is when dealing with inhomogeneous populations and interaction networks but also how to go from a differential equation model to an agent-based model– usually it is the opposite transformation that is used, where the differential equation model is the analytically tractable (but deceivingly so) mean-field version of the agent-based model. What is useful about this "reverse" transformation is that it clearly show that an agent-based model is increasingly necessary as the degree of inhomogeneity increases in the model. [5]

CHAPTER II

INTRODUCTION TO DISEASES MODELS

2.1 Introduction

The study of epidemic disease has always been a topic where biological networks mix with social ones. Epidemics can pass explosively though a social network and persist for a long period of time. They can also experience wave-like cyclic patterns of increasing and decreasing prevalence or skyrocket in a short time.

The patterns by which epidemic spread through people is determined by both the properties of the pathogen carrying it and the network structures. The possibilities for a disease to spread are determined by a *contact network*, where a node represents a person, and an edge for two people is the contact with each other such that the disease transmit from one to the other.

To understand how an epidemic spreads through populations, we need to accurately model their contact network. In this chapter, we will discuss three major kinds of disease models.

2.2 The SIR Epidemic Model

In the SIR epidemic model, there are three potential stages:

- *Susceptible (**S**)*: The status before an individual was infected by the disease is denoted as susceptible stage. During this stage, an individual can be infected by others.

- *Infectious (**I**)*: Once the individual has caught the disease, he/she is infectious or infected. During this stage, it is possible for them to infect others.

- *Removed (**R**)*: After a particular period of time the node has experienced the infection, the individual is recovered and is considered removed from the epidemic model permanently,

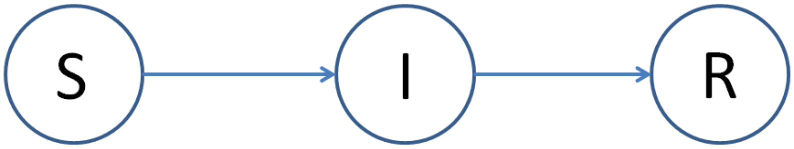


Figure 2.1: A simple graph of the SIR epidemic model.

since we assume that he/she will gain life-time immunity once he/she recovers from the infection.

The three-stage "life cycle" helps us define an epidemic model. Figure 2.1 shows a simple example of the $S\ I\ R$ disease model. Obviously, the contacts between people are symmetric, i.e. suppose an edge pointing from A to B in the graph, if A becomes infected at some point, the disease has the potential to spread directly to B and vice versa.

Now, each individual has the potential to go through the Susceptible-Infectious-Removed cycle, where we abbreviate these three states to $S\ I$, and R. The progress of the epidemic is determined by the rate of contact and by two other variables: p (the probability of transmission upon contact) and t_I (the length of the infection, or the incubation period).

- Initially, some individuals are in the I state and all others are in the S state.

- Each individual v that enters the I state remains infectious for a fixed exponentially distributed length of time.

- During that t_I time, v has a probability p of passing the disease to each of its susceptible encounters.

- After t_I, individual v is no longer infectious or susceptible; we denote it as *removed* (R),

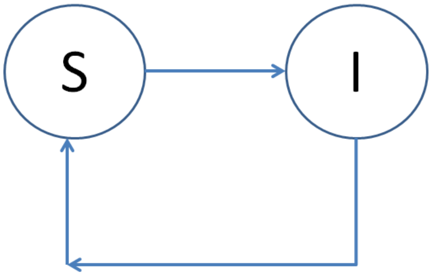


Figure 2.2: A simple graph of the SIS epidemic model.

since we assume that the individual will gain life-time immunity and can no longer contract the disease.

This describes the full epidemic disease model; we refer to it as the *SIR model*. [7]

2.3 The SIS Epidemic Model

The SIS model is used for studying some recurring diseases. One of the assumptions of the SIR model is that each individual gets infected by the disease only once. For some diseases, like flu, an individual can contract the disease more than once.

The SIS epidemic model is actually an extension of the SIR model. The nodes is in the SIS model are either susceptible or infected. To show such a model, we just have to slightly change the SIR model. This time there is no *Removed* state; instead, after an individual is recovered from the Infectious stage, he/she goes back to the *Susceptible* state and can be infected by the disease again. We refer to the model as the *SIS model*, as shown in Figure 2.2.

To find out the probability of an individual to get infected, let ρ be the percent infected at any point of time and δ be the rate that the node will be recover at certain period of time. We assume every node in the network has the equal probability to get infected and they randomly meet each other. Since the SIS model can be considered as a large Markov chain. the equilibrium state of the model is

$$\frac{d\rho}{dt} = 0$$

.

Then we have the differential equation of the SIS model:

$$\frac{d\rho}{dt} = (1-\rho)(v\rho+\varepsilon) - \rho\delta = 0, \tag{2.1}$$

where v is the probability that one get infected to the number of infected neighbors, ε is the spontaneous error, and $(1-\rho)(v\rho+\varepsilon)$ are those who are infected, and $\rho\delta$ are those who recover from the infection [10].

By solving Equation (2.1), we have

$$\rho = \frac{[(v-\delta-\varepsilon)+((v-\delta-\varepsilon)^2+4\varepsilon v)^{\frac{1}{2}}]}{2v} \tag{2.2}$$

If we let $\varepsilon = 0$, Equation (2.1) becomes

$$\frac{d\rho}{dt} = (1-\rho)v\rho - \rho\delta = 0 \tag{2.3}$$

Then we have two solutions:

$$\rho = 1 - \frac{\delta}{v}$$

or

$$\rho = 0$$

If we just look at $\rho = 1 - \frac{\delta}{v}$, we can imply that if $\delta > v$, then the individual will recover faster than get infected, which means no infection stays; otherwise, infection will stay at some level and for low recovery rates, an outbreak of the infection is on the way [7].

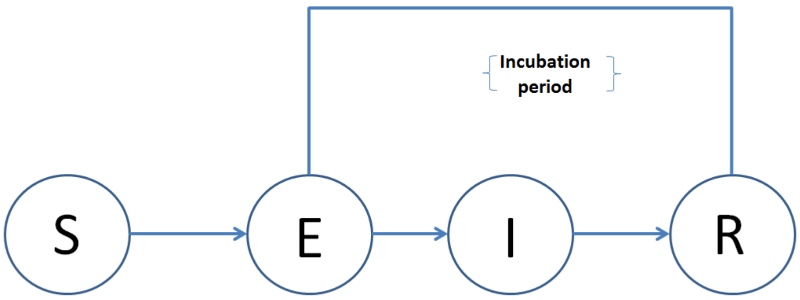


Figure 2.3: A simple graph of the SEIR epidemic model.

2.4 The SEIR Epidemic Model

In addition to the existing SIR model, one can add another state called *Exposed* to represent the diseases that are not contagious once one contacts them, as intermediate state between the *Susceptible* and *Infectious* states. Figure 2.3 shows a simple graph of the SEIR disease model [15].

2.5 SIR Model of Vaccinator with Social Norms

In real social life, people can be affected by others' opinions, just like we discussed in the information cascades section. Consider such a scenario, a woman in a family just gave birth to a baby, the next decision the parents have to make is whether or not to vaccinate their baby to prevent from some potential pediatric diseases. If the community or even the country, the family live in, requires a compulsory vaccination to new-born, then the family may vaccinate their baby under the social pressure imposed on them; If, however, the community or the country requires only the voluntary vaccination, then the probability of the family to vaccinate their baby is not 100%. Also, if the family live in a community where most of the people are vaccine refusers, then the parents of the new born will probably not vaccinate their baby. Then just like we discussed in the previous "information cascades" section, parents will be affected be others.

Moreover, the authors find out that if we include injunctive social norms to the disease model, both the pertussis vaccine uptake behavior and the disease dynamics can be better explained. Furthermore, they find out that social norms can strongly suppress vaccine uptake even if frequent outbreaks happen. As a whole, social norms can either support or impede immunization goals. Their study of the transmission dynamics obtained from a susceptible-infected-recovered compartmental disease model can be represented by the full model equations

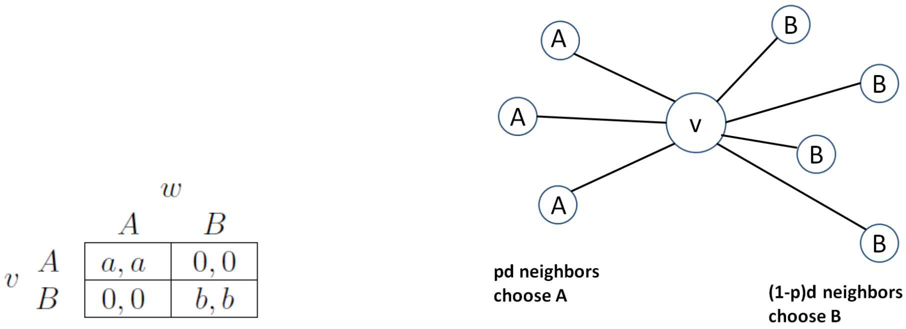
$$\left.\begin{aligned} \frac{dS}{dt} &= \mu(1-x) - \beta SI - \mu s, \\ \frac{dI}{dt} &= \beta SI - (\mu+\gamma)I, \\ \frac{dx}{dt} &= \kappa x(1-x)(-\omega + I + \delta(2x-1)), \end{aligned}\right\} \tag{2.4}$$

where S, I and R are the proportion of susceptible, infected, and recovered individuals in the population, respectively, at time t, μ is the per capita birth/death rate, β is the transmission rate and γ is the recovery rate, κ and δ are the rescaled sampling rate and effect of injunctive social norms, respectively [14].

Now we can generalize the scenario to a diffusion model. In a social network, we have a situation in which each individual has a choice between two options: A or B. If two individuals v and w are linked by an edge, then there is a possibility that their choices match. We represent this by a game in which v and w are the players and A and B are the options. The payoffs are as follows:

1. if v and w both choose A, they each get a payoff of $a > 0$,
2. if they both choose B, they each get a payoff of $b > 0$,
3. if they choose opposite options, they each get a payoff of 0.

We use a matrix in Figure 2.4 to show the situation above. The point is each v is copying choices of its neighbors. The basic question faced by v is : since some of its neighbors choose A, and some choose B, what should v choose to get maximum payoff? Clearly, this depends on the number of each of the two choices, and payoff values a and b. Suppose p fraction of v's neighbors choose A, and $(1-p)$ fraction choose B, i.e. if v has total d neighbors, then pd neighbors will choose A and $(1-p)d$ neighbors will choose B, as shown in Figure 2.4.



$$\begin{array}{cc|c|c|} & & \multicolumn{2}{c}{w} \\ & & A & B \\ \hline v & A & a,a & 0,0 \\ \hline & B & 0,0 & b,b \\ \hline \end{array}$$

Figure 2.4: The cascading example.

If v chooses A, it gets payoff pda; if chooses B, it gets payoff $(1-p)db$. As a consequent, A is the better choice if

$$pda \geq (1-p)db$$

or we can re-write it as

$$p \geq \frac{b}{a+b}$$

This intuitively makes sense: when $p \geq \frac{b}{a+b}$ is small, then choice A is more enticing; on the other hand, if $p \geq \frac{b}{a+b}$ is large, the opposite holds: choice B is more attractive.

Consequently, people with more neighbors with payoffs of pda will be more likely to option A since they have more pressure come from choosing option A. On the other hand, people with more neighbors with payoffs of $pd(1-p)$ will be more likely to option B since they have more pressure come from choosing option B[7].

2.6 Why Do We Need to Model Social Networks?

Why do we need to use model to describe social networks? First and foremost, many techniques used to measure properties of a network are valuable in describing and understanding network features. Besides, there are some other reasons lead us to use social networks in modeling:

(1) Social behavior is complex, capturing the characteristics of a network and at the same

time recognizing the variability that we can't model in detail. Moreover, social behavior is very stochastic. The randomness we add to a regular process will dramatically change outcomes of the process.

(2) We make inferences about the statistical models to see if they have any properties observed more than our expectations. Then we can construct hypotheses about social networks that will produce those structural properties for our future study.

(3) In order to differentiate similar predictions from different social processes, we need to quantify different social network structures carefully to evaluate the differences in the predictions.

(4) Although there are many approaches to analyze social networks, many of them are not appropriate. With more complex network structures, it's more difficult to achieve efficient outcomes. As a consequence, well-established models are of great values.

(5) Questions such as how social processes and structures help to form network patterns and if those processes are sufficient to explain social network properties are of great importance in social network analysis. To fill the gap of investigating social network properties, construct models to simulate the social processes is a must [16].

2.7 Network Models of Diseases

In recent years, many researchers are studying the relationship between social contact networks and vaccines behaviors and influential factors, that will affect parents' decisions about vaccination. What we are doing is an extension of the previous works done by many researchers.

In [17], the authors try to find more reasonable explanations for large outbreaks of some vaccine-preventable diseases in some high-income countries. They found out that the effect of clustering on outbreak probabilities is strongest when the vaccination coverage is close to the level required to provide herd immunity under the assumption of random mixing.

Then further study shows that the real reason is the "belief systems," which include the reaction to efficacy of the vaccines, and religious considerations. Specifically, people with same negative opinions about the vaccination seem to form a susceptible clustering. Those clusterings form a subpopulation that impede the immunity of the community, and lead to outbreaks of the disease at some point in time.

In "Social contact networks and Disease eradicability under voluntary vaccination" [15], the authors find out that some vaccine-preventable diseases like smallpox have been globally eradicated by voluntary vaccination in many countries while other vaccine-preventable diseases such as MMR and pertussis are very difficult or impossible to eradicate under voluntary vaccination. They have seen through simulation and analysis from the SEIR model, individual will become less motivated to vaccination when the vaccine coverage is very high. But when the number of nonvaccinators in neighborhood becomes very large, it will reach a threshold and then go beyond that threshold, resulting in an epidemic. Then people will start seeking vaccination to prevent the disease and the number of people get vaccinated will increase again.

The social network is generated by the Erd$\ddot{o}$s - Renyi model, which means it is approximately Poisson-distributed. The total probability that the node becomes infected on a given day is

$$\lambda = 1-(1-\beta)^{n_{inf}}$$

where β is the probability of node-to node transmission and n_{inf} is the number of infectious neighbors.

Also, the perceived probability if an individual who has n_{inf} neighbors of being infected on any given day is

$$\lambda_{perc} = 1-(1-\beta_{perc})^{n_{inf}}.$$

where β_{perc} is the perceived probability per day that the individual is infected by a single given infectious neighbor.

When researchers dig deeper, they find out that cognitive effects also play a active role in affecting vaccinating behaviors [13]. First, they develop a "bounded rational" model by prospect theory to compare with the model lacks "bounded rationality." They find out that "bounded rationality" increases the dynamical richness of the model and at the same time makes eradication of a pediatric infectious disease even harder. Injunctive social norms can help prevent vaccine refusal even if the cost of vaccine is high and the vaccine is not efficacious. We can see cognitive processes have major impacts on predicting disease models.

Facing with the dilemma of how to achieve widespread immunity to infectious diseases by voluntary vaccination, the authors in [8] try to find the roles of individual imitation behavior and population structure in vaccination. They also present the way imitation of individuals shapes others' vaccination choices in social networks. They conclude that the vaccination strategy of an individual is influenced by his/her role in the population, i.e. those with many friends and neighbors are more likely than others to get infected by the disease [8].

Besides, in [3] the authors find out that network models become more intuitive and accurate to predict disease by heterogeneous population. Since many epidemiology models are based on a simplifying assumption that individuals in a social network have identical rates of contacts, which is not very realistic. Scientists try to quantify the extent that the real population contact patterns depart from the homogeneous population. Surprisingly, the patterns are not variable as they speculated [3].

In [11], researchers use network-based mathematical models to study the effects of both imitation behavior and contact heterogeneity on vaccination coverage and disease dynamics. The simulation results suggest that when the cost of vaccination is high, imitation behavior may decrease vaccination coverage. However, when the cost of vaccination is small relative to that of infection, imitation behavior increases vaccination coverage, but, surprisingly, also increases the magnitude of epidemics through the clustering of non-vaccinators within the network. The imitation behavior, in a whole, may impede the eradication of infectious diseases.

Social contacts can exhibit complicated dynamics, such as the threshold and cascading instabilities. Social contact can be conceptualized as a network, where each node is a person, and the network links are social contacts through which the information can spread (Figure 2.5, panel A). Infectious diseases can also exhibit complicated dynamics and be conceptualized as a network through which the biological contact spread (Figure 2.5, panel B). When a social contact is mixed

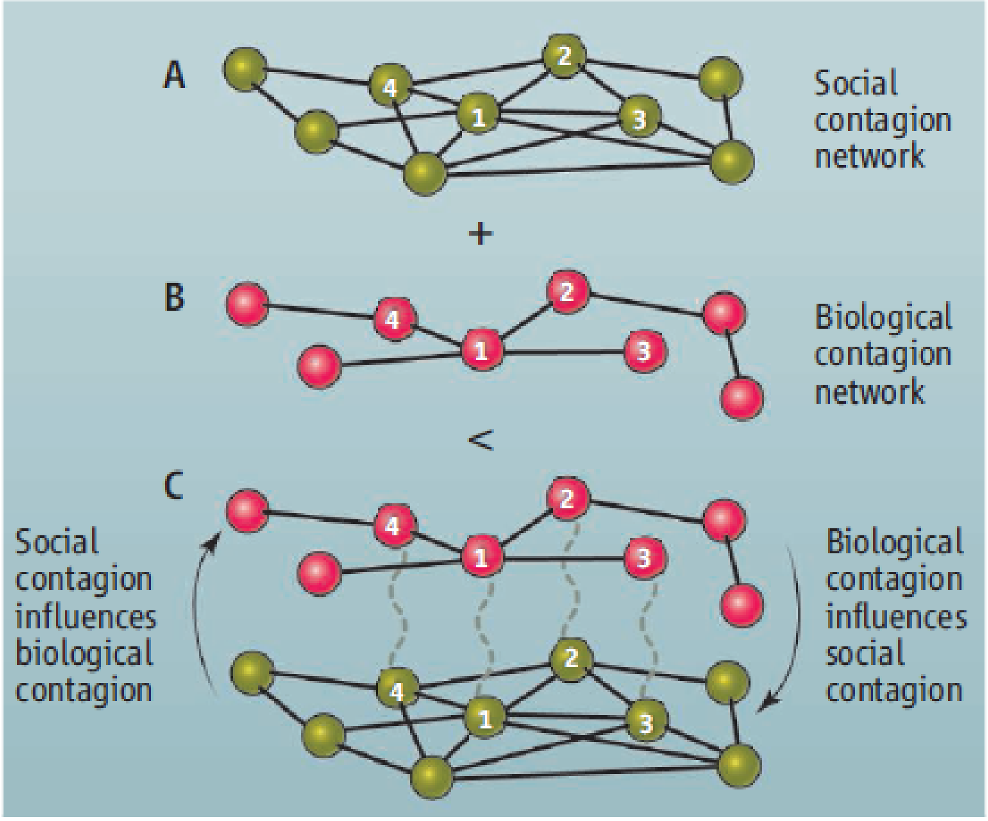

A
Social contagion network
+
B
Biological contagion network
<
C
Social contagion influences biological contagion
Biological contagion influences social contagion
1
2
3
4

Figure 2.5: More than the sum of its parts

[4]

with a biological one, the whole system will exhibit dynamics that do not occur when the two subsystems are isolated from one another. This illustrates that the whole is more than the sum of its parts (Figure 2.5, panel C).

For example, high levels of pediatric vaccine coverage can decrease disease incidence to low levels, reducing the perceived danger of infection and the urgency to get vaccinated. Subsequently, if highly connected nodes in the social network (such as celebrities) suggest that the vaccine carries risks, the perception of vaccine risks can propagate quickly through the social network, leading to a vaccine scare and a drop in vaccine coverage. In this case, biological contact influences the social one.

On the other hand, the drop in vaccine coverage allows the number of individuals who are susceptible to infection to increase. When the proportion of susceptible individuals exceeds a threshold, an outbreak of infectious disease occurs, which may motivate individuals to seek vaccination once again: social contact influences biological contact in this case (Figure 2.5, panel C). [4]

This is an perspective article from Science journal in 2013. It points out the direction we are working on in our research. Based on the perspective in the article, we build up our "double-layer" model to study its properties.

CHAPTER III

A DUAL MODEL OF DISEASE SPREAD AND VACCINE OPINION ON SOCIAL NETWORKS

Parents and children have different social contact networks. Children know their friends from day cares, schools or community activities. Parents are connected with each other through relatives, friends, co-workers and even through the social media on the Internet. Our network model consists of N vertices representing a population of N households (we use different values of N to be linked as an Erd*ö*s - Renyi network). Initially, the status of each household in our agent-based network is set to be susceptible except for a preselected number of children in households ($I_0 = 10$).

3.1 Model Description

1. Childrens' Contact Network Model Description (Biological Network)

We construct contact networks for both the children and the parents since it is reasonable to say that the parents and the children have different social networks. We assign random number of children to each household using a binomial distribution. They have contact with their siblings within the households and they can also contact their friends through day cares, schools and neighborhood activities. Assume the probability of children within and between the households have connection with each other is 0.002. We want to study how the disease spreads between and within the household through the biological network. We also want to study how long it takes for the disease to die out.

2. Parents' Social Network Description (Social Network)

Parents have their own social contact networks. Some parents are connected with each other through their children's contact network, while other parents may not even know each other

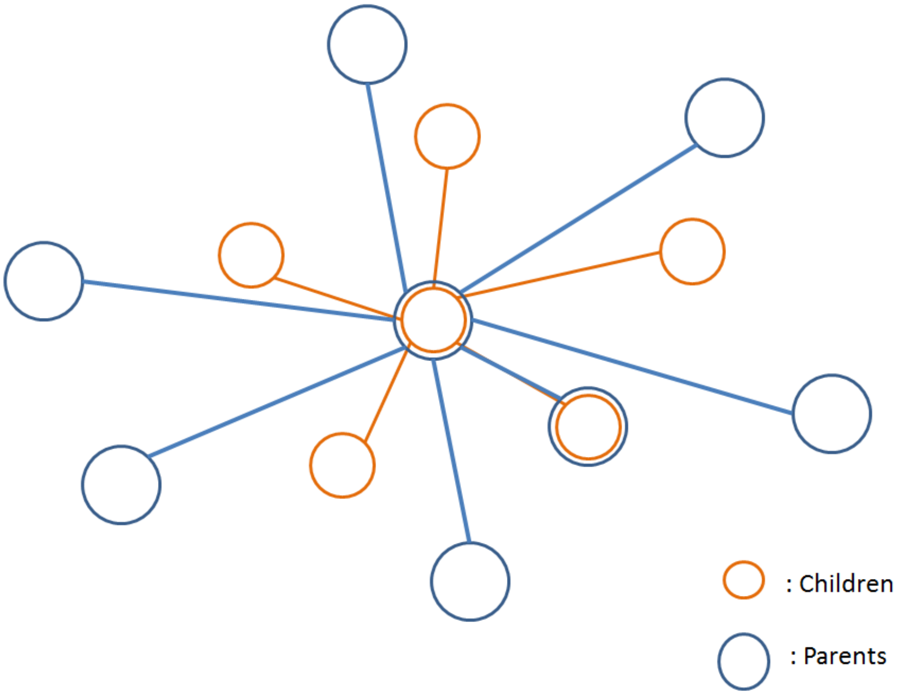


Figure 3.1: The dual-layer networks.

even if their children go to the same school or daycares. Consequently, the networks of parents and children partially overlap, as shown in Figure 3.1.

In our model, we randomly retain some links between parents from the children's network to represent the connection between them through their children. At the same time, we randomly add some links among parents to represent the parents connecting with each other through their social contact network instead of the children's. One of the goals of our research is to see how these two networks affect each other and, thus, how the social network will affect the parents' vaccinating behaviors.

3. Transmission of Disease Mechanism

There are three stages in the SIR model: the susceptible stage, the infected stage and the recovered stage (the time units are days). The model tracks the total number of days an infected child in each household will take to go through the incubation period. We assume once a family member recovers from the infection, he/she gains life-time immunity.

The initial status of each child (except 10 infected of them) is susceptible since they contact with each other through schools or daycare. At some point, a child can be infected by the disease through his/her contacts or siblings. We use pertussis as our prototype. The maximum incubation period of the disease is 28 days, meaning that it takes at most 28 days for an infectious child to become not infectious. We assume the mean incubation period of the disease is 22 days. We use truncated gamma distribution to model length of incubation period.

From the graph, we can see that it is nearly impossible for a child to recover within the first 9 days. Then from the 9th day on, the probability to recover starts to increase gradually peaking approximately at the 21st day. Consequently, there is a possibility that some children will recover without spending all 28 days since immune system varies in children. Children who have stronger immune systems may recover after a few days. The probability for children to recover on mth day is

$$Pr[m-1 < X \leq m] = \int_{m-1}^{m} f_{trunc.}(x)dx \tag{3.1}$$

and continue with probability of

$$Pr[X > m] = \int_{m}^{28} f_{trunc.}(x)dx. \tag{3.2}$$

There are two options for an infectious child to have on the next day: either stay infectious or recovered from the epidemic. The model checks their status daily to see if they are still infectious or recovered everyday and count how many of them are infected, specifically to the exact day of the incubation period, and how many are recovered from the disease, as shown in Figure 3.3.

The probability that a child within a household will contract the disease depends on the number of infected children in the neighboring households and the child's siblings. Each time, a susceptible household can become infected with the probability β, where β is the probability of transmission between households, and β_h is the probability one gets infected within the household. Then the probability of a child gets infected is:

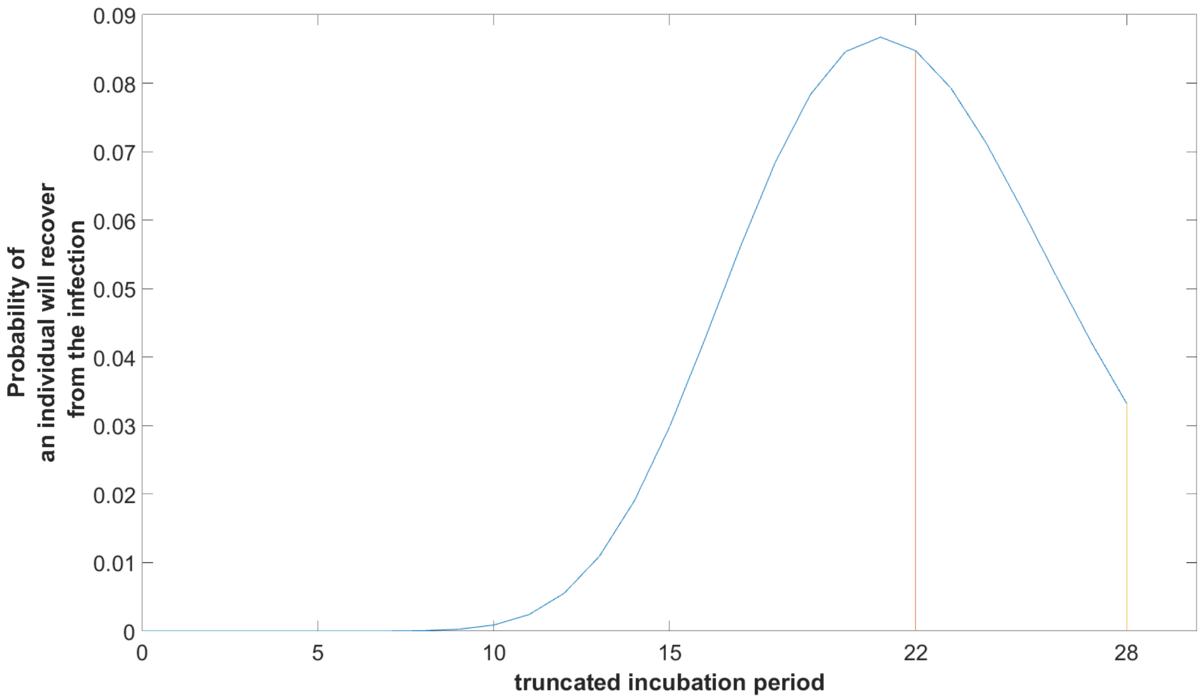


Figure 3.2: The truncated gamma probability distribution modeling the incubation period.

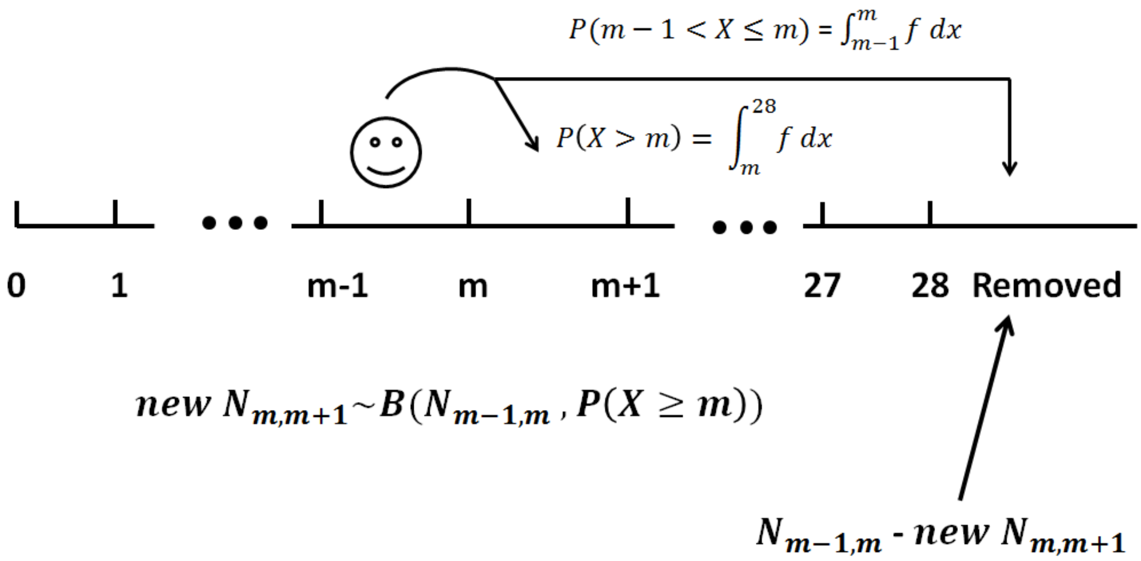

$P(m-1 < X \leq m) = \int_{m-1}^{m} f\,dx$
$P(X > m) = \int_{m}^{28} f\,dx$
0
1
m-1
m
m+1
27
28
Removed
$new\ N_{m,m+1} \sim B(N_{m-1,m}\,, P(X \geq m))$
$N_{m-1,m}$ - $new\ N_{m,m+1}$

Figure 3.3: The truncated gamma probability distribution modeling the incubation period.

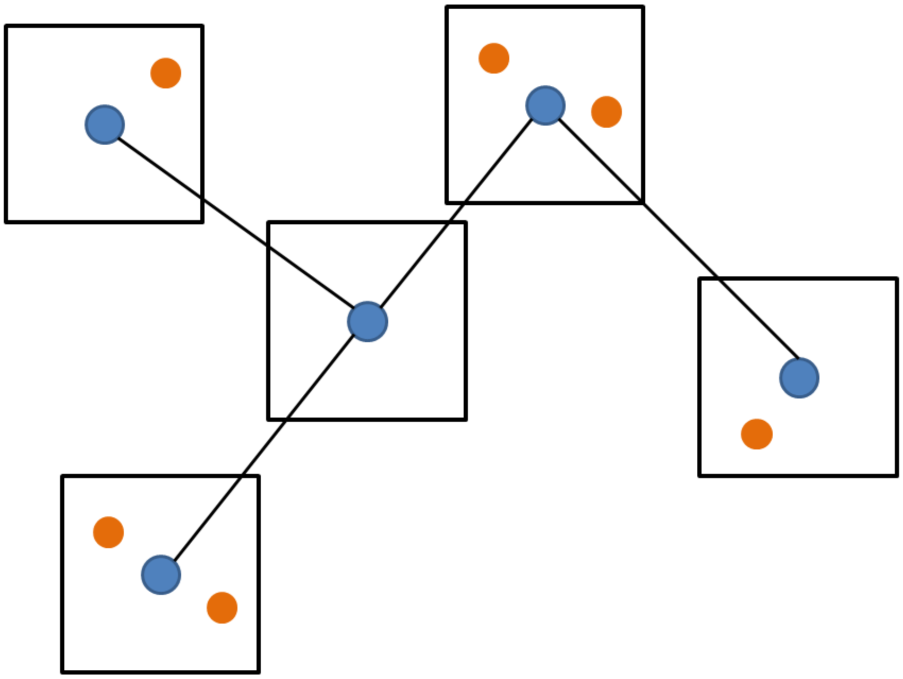


Figure 3.4: Binomial distribution of the transmission process.

$$Pr[a\ new\ infection\ in\ a\ household] = 1-(1-\beta_h)^X \cdot (1-\beta)^{sH} \tag{3.3}$$

where X is the number of infected children within the households and sH is the number of infected households of distance 1 in the biological networks of children.

The transmission process is not a Poisson process since there is zero, one or two children within the households (Figure 3.4). The number of susceptible children is not large enough to use Poisson approximation. Consequently, it's safe to say that the transmission process follows binomial distribution instead of Poisson distribution.

4. Vaccination Opinion Diffusion

The social network analysis helps us to explain why information of neighbors about the vaccine within the social network is critical to the decision-making behaviors of parents, both for vaccine acceptors and for vaccine refusers. The idea behind it is simple: if an unvaccinated individual is more likely to contact with other unvaccinated individuals, clusters of susceptible individuals will form and thus constitute small groups in which the disease can spread and cause outbreaks. One thing to be mentioned is that the vaccination opinion is affected also by how people feel about the disease and the adverse cases of the vaccine, i.e. all the possible complications and side effects of the disease and the vaccine. So we should weight those factors when we try to find how vaccination opinion diffuses around people in a social network.

We use "information cascades" to examine its effects established sociologically on vaccination. Information cascades happen when parents of the households want to decide whether to vaccinate their new-born. It's reasonable to speculate that parents from one household willing to vaccinate their children are more likely to cluster with other parents that are willing to do the same to their children than to cluster with parents that are unwilling to vaccinate their children. When parents from one household meet and talk with each other, or learn from the Internet that other parents vaccinate their children, they are more likely to vaccinate their children as well. Here, the information cascades start to affect parents' behavior.

We will see from our model how parents' vaccination decision behaviors are affected by their neighbors and friends and how much the effect of the opinions of others would have on the parents. First we calculate, in the social network, the number of people who intend to vaccinate their new-born baby and the number who do not within the social depth of 1, i.e. from a person who has a new-born in the social network to his/her friends. We have the objective probability, i.e. the probability of vaccination of a parent without any influences from others, that a parent decides to vaccinate their new-born

$$p = \frac{1}{1 + exp(-(\alpha I - \gamma advV))} \tag{3.4}$$

where advV is the adverse event of the vaccine, α is the proportionality of the infectious and γ is the proportionality of the adverse event of the vaccine.

Normally, γ is much greater than α due to the psychological behavior called "omission bias", which claims that people tend to avoid give harm to their children directly. If there is a

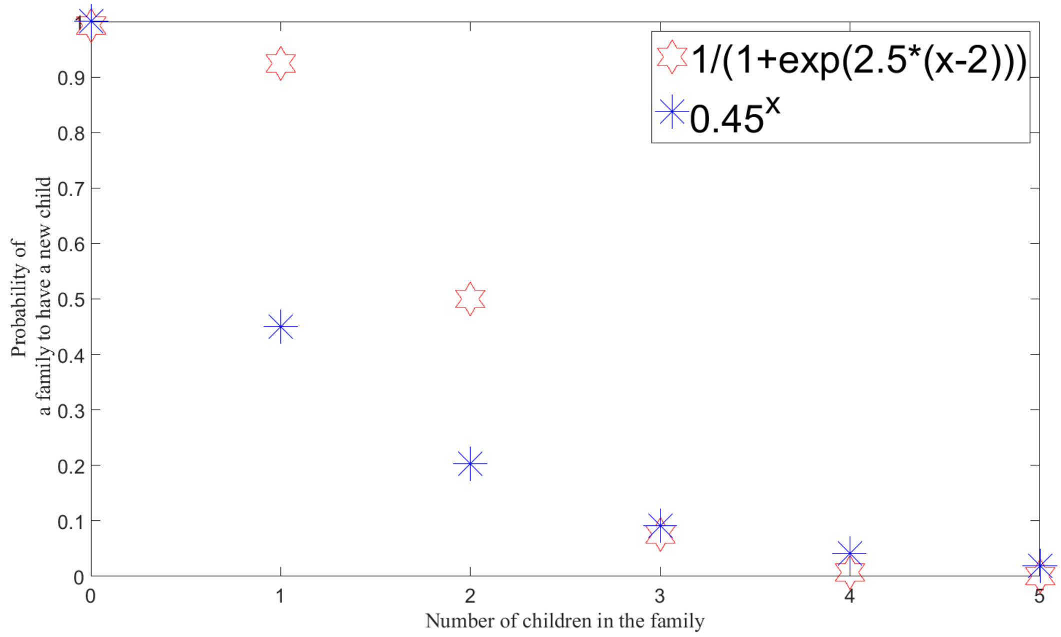


Figure 3.5: The probability of a household to decide whether to have a child based on family size.

possibility that the children can be infected by some epidemic and there is a possibility that the vaccine may have some side effects, the parents would rather to take the risk to let their children exposed to the disease other than vaccinate them because the parents do not want to regret for the possible side effects that may have on their children.

Then using the information cascade formula (Eq.(1.14)), [7]

$$Pr[V|S] = \frac{pq^a(1-q)^b}{pq^a(1-q)^b + (1-p)(1-p)^a q^b}. \tag{3.5}$$

where q is the probability to give a high signal for vaccinators and low signal for nonvaccinators. a and b are the numbers of people who intend to vaccinate or not vaccinate their children, respectively, within the social depth of 1.

Equation (3.5) gives us the updated probability of a parent to vaccinate a new-born.

5. Demographics

We use the birth process to introduce the new-born with the birth probability λ, modulating by the size of the family. It takes approximate 280 days for a pregnant woman to give birth to a

Table 3.1: Baseline parameters values for the model

Parameters	Description	Baseline
N	Number of households	5000
P	Probability of households connect with each other	0.002
i_p	Maximum length of incubation period	28
pf	Probability of having a initial child	0.5
β	Probability of network disease transmission	0.08 & 0.04
β_h	Probability of the disease transmitted within household	0.1 & 0.01
e	Vaccine efficacy	0.95
ReP	probability of retaining social network links	0.01
AdP	Probability of adding new social network links	Probability of adding new social network links
α	proportionality of the infectious	0.001
γ	proportionality of the adverse event of the vaccine	200
λ	proportionality of a family to give birth to a new-born	0.0001

baby. We assume each household in the social network consist of parents and an initial random number of children. The probability of a household to decide to have a child decreases with the increase of the number of children it already has, using probability functions depicted in Figure 3.3. The vaccine efficacy is assumed to be 0.95. Table 3.1 shows all the parameters used in our model.

CHAPTER IV

RESULTS AND INFERENCES

We have run simulations for long time; some went for 6000 simulation runs. We set $N = 5000$, that is, the community consists of 5000 households. We stop simulation when the disease dies out, that is, with the recovery of the last infectious child. Since we build our model based on the assumption of the SIR model, which claims that the child will gain life-time immunity after recovery. Different combinations of parameters' values give rise to different dynamic behaviors, which warrants a thorough exploration of the parameter space.

Figure 4.1 shows the mean incidence of the epidemic with different birth rates. From 0 to 1, the incidence of the epidemic shows a U shape, and reaches to the lowest point around the middle of the interval. The corresponding interval in the mean time length of the epidemic, shown in Figure 4.3, reaches its peak in the middle of the interval as well. The possible reason is that the population we use in our network model is small, which means the number of infected children is small, resulting in the insufficient number of susceptible population. With the decrease of disease severity and increase in population size, the time length of the epidemic is longer. Combining Figure 4.5, we can conclude that as the time length of the epidemic is longer, more and more people start to vaccinate children to protect them.

Besides, we noticed an interesting phenomenon from Figure 4.1: when people are frank about their true opinions about vaccination, i.e. if one prefer to vaccination, he/she will suggest his/her neighbors to do same thing and vice versa, the total the size of epidemic becomes very large. So sometimes it's better not share your true opinions regarding to vaccination. This is the counterintuitive property of the agent-based model. It is actually interesting and worth exploring

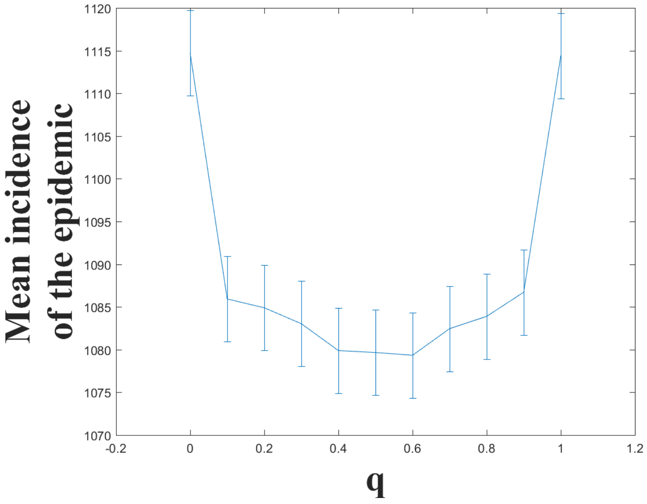


Figure 4.1: Mean incidence of the epidemic disease.

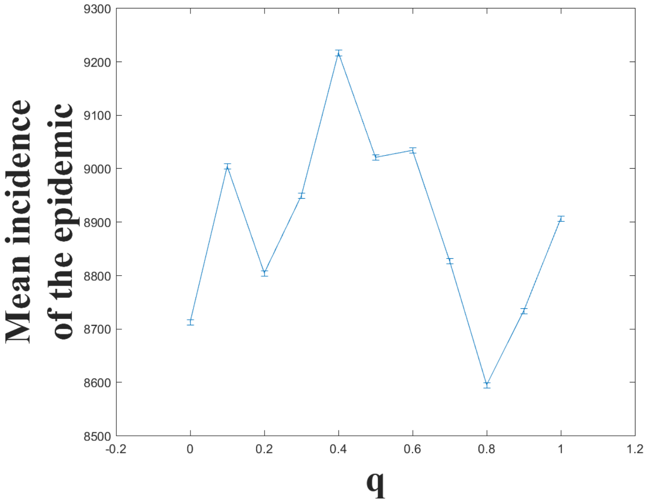


Figure 4.2: Mean incidence of the epidemic disease with larger birth rate.

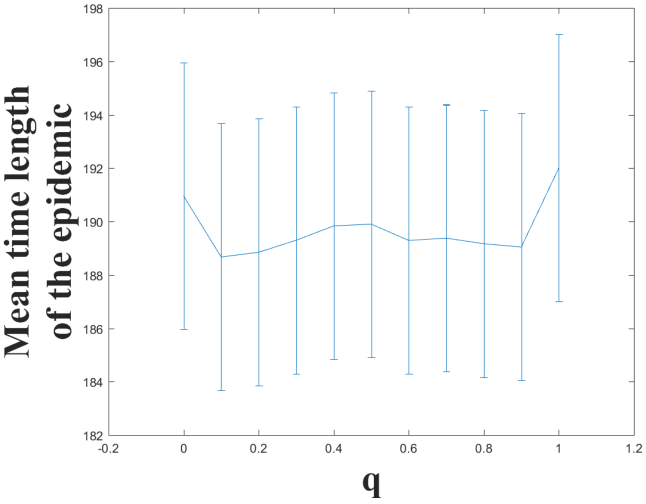


Figure 4.3: Mean length of the epidemic disease.

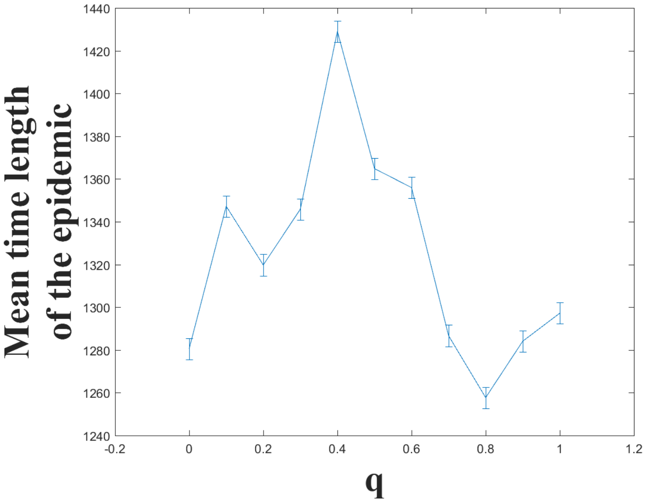


Figure 4.4: Mean length of the epidemic disease with larger birth rate.

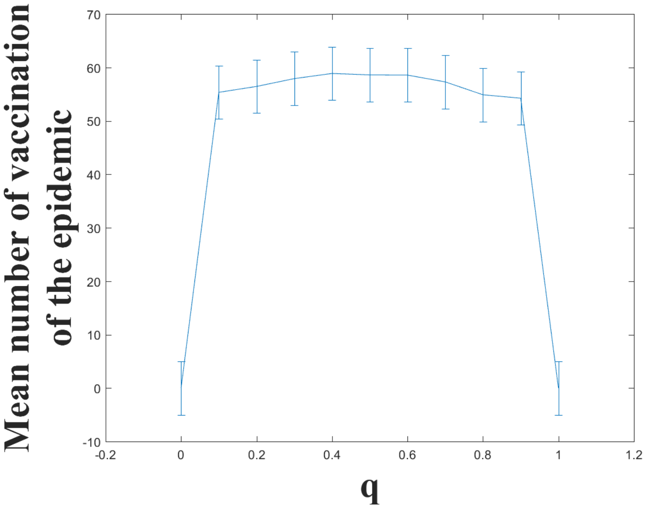


Figure 4.5: Mean number of vaccination of the epidemic disease.

in the future.

We are curious about what will happen if we increase the birth rate. Then according to Figure 4.2, 4.4 and 4.6, we increase birth rate by and decrease the transmission rate both within the household and among the network. Surprisingly, the incidence increases and reaches to its peak at the middle of the interval (Figure 4.2), and the number of incidence increases significantly. Although the reason for the oscillating behaviors of the incidence is unknown, the possible reason for the increasing number of incidence may be with larger birth rate, there is more chance for a family to have a new baby and in the long term, the total number of susceptible children will increase dramatically, leading to a higher probability that they become infected by the epidemic disease.

Similar oscillating behavior is also observed from the length of the epidemic disease, as shown in Figure 4.4. Undoubtedly, the more infected people within the network, the more chances for those susceptible to get infected again, resulting in longer transmission time period.

Some unexpected phenomenon shows in the number of vaccination (Figure 4.6). The number of vaccination decrease gradually after reach its peak at around 0.2. But since the range of the difference is just about 20, we consider the number of vaccination keep constant during the whole

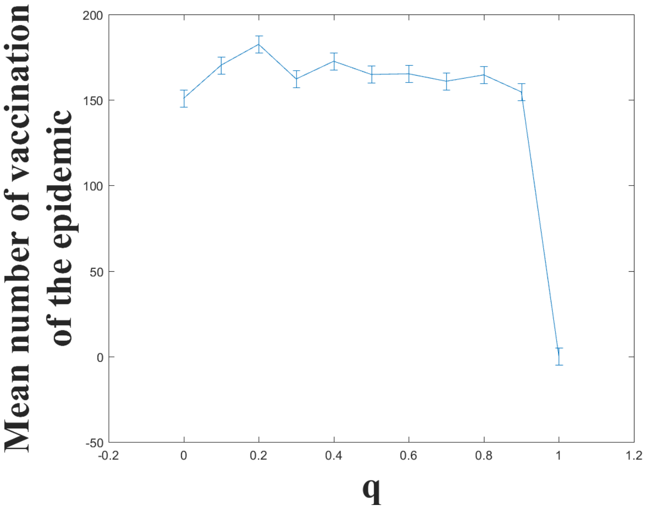


Figure 4.6: Mean number of vaccination of the epidemic disease with larger birth rate.

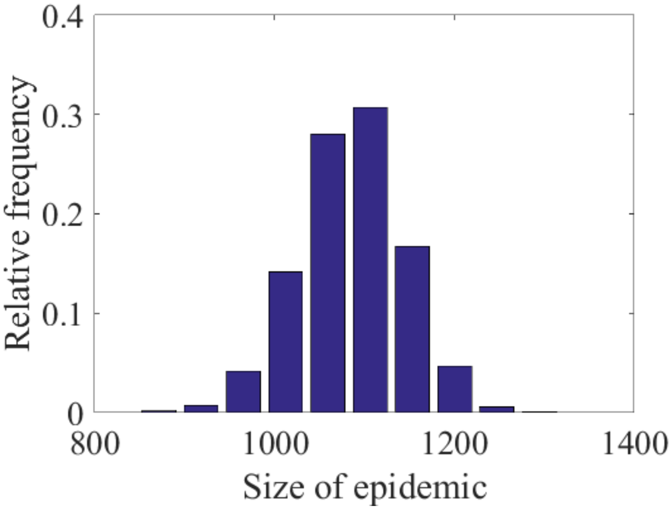


Figure 4.7: Number of epidemic incidence with P = 0.002 with q = 0.1.

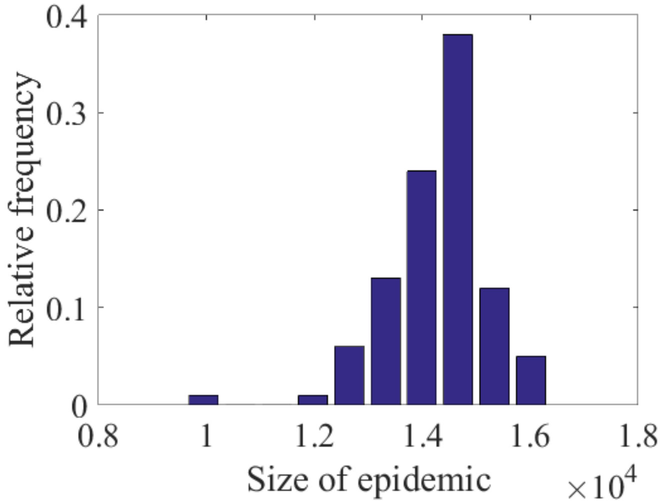


Figure 4.8: Number of epidemic incidence with P = 0.004 with q = 0.1.

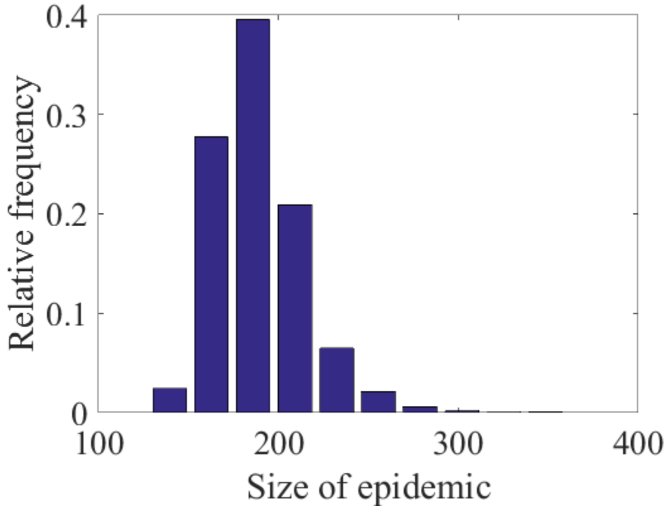


Figure 4.9: Time length of the epidemic with P = 0.002 with q = 0.1.

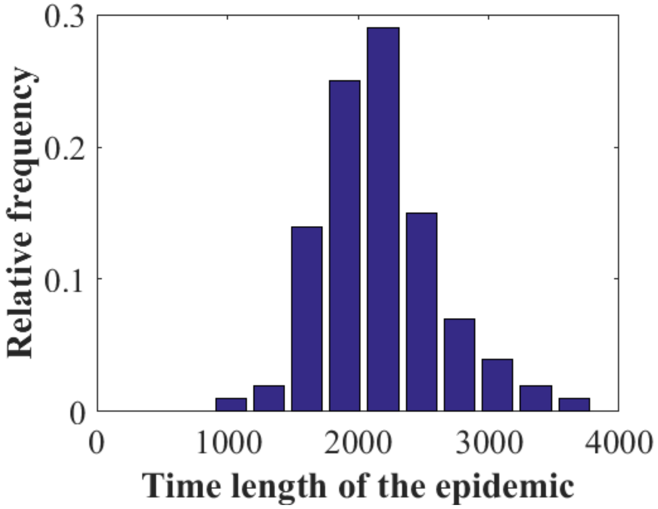


Figure 4.10: Time length of the epidemic with P = 0.004 with q = 0.1.

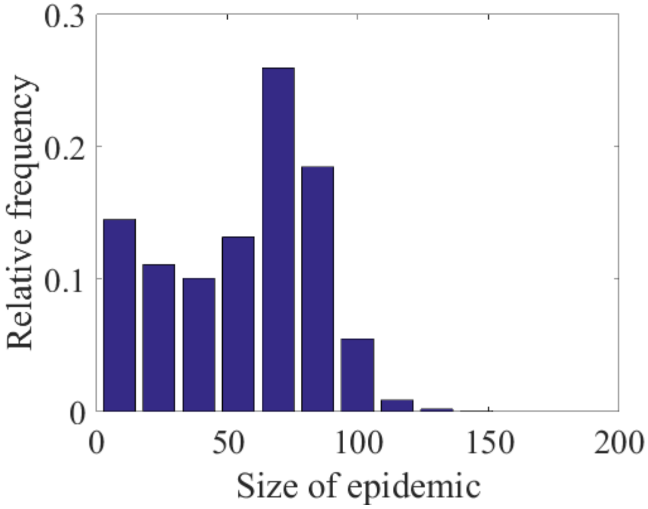


Figure 4.11: Number of vaccination with P = 0.002 with with q = 0.1.

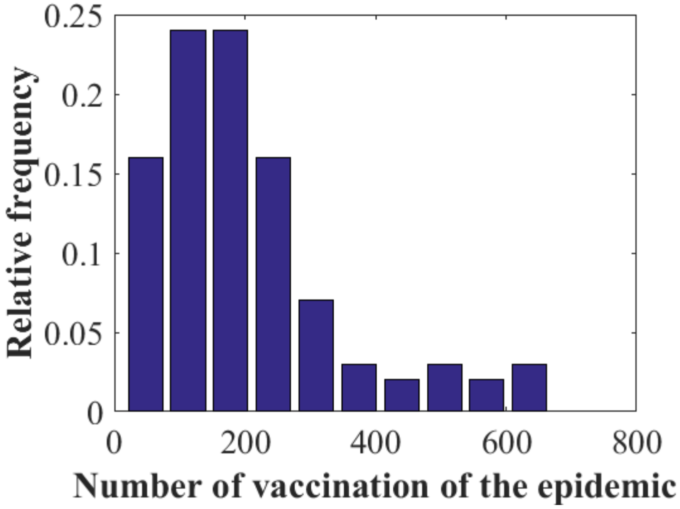


Figure 4.12: Number of vaccination with P = 0.004 with q = 0.1.

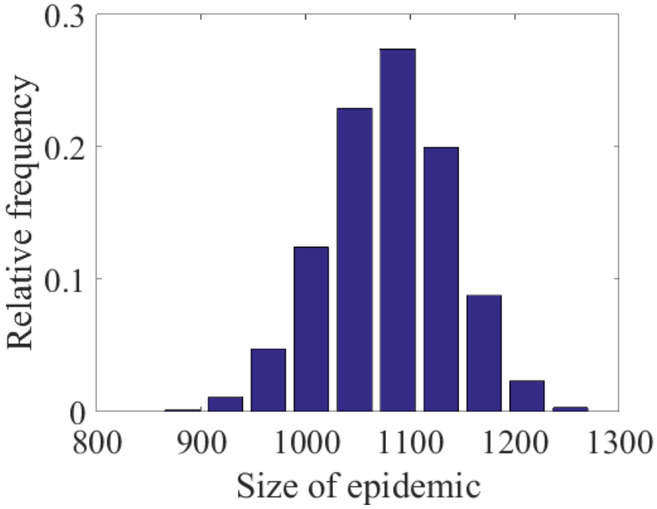


Figure 4.13: Number of epidemic incidence with P = 0.002 with q = 0.5.

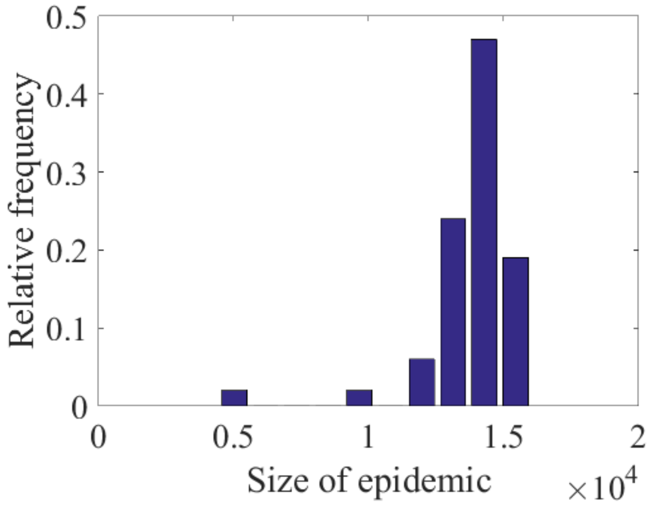


Figure 4.14: Number of epidemic incidence with P = 0.004 with q = 0.5.

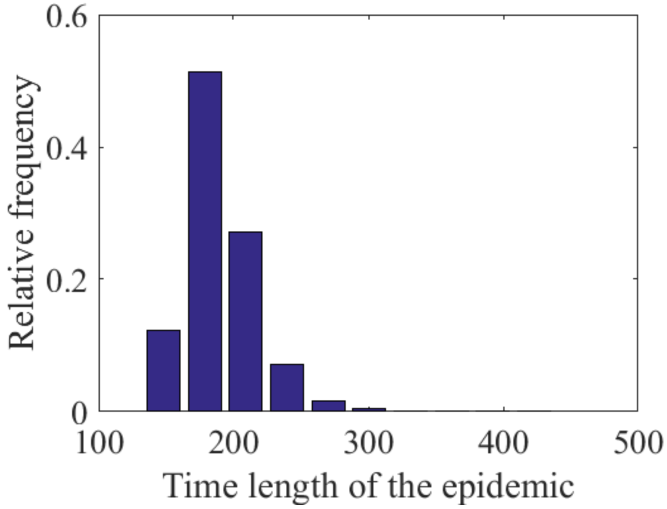


Figure 4.15: Time length of the epidemic with P = 0.002 with q = 0.5.

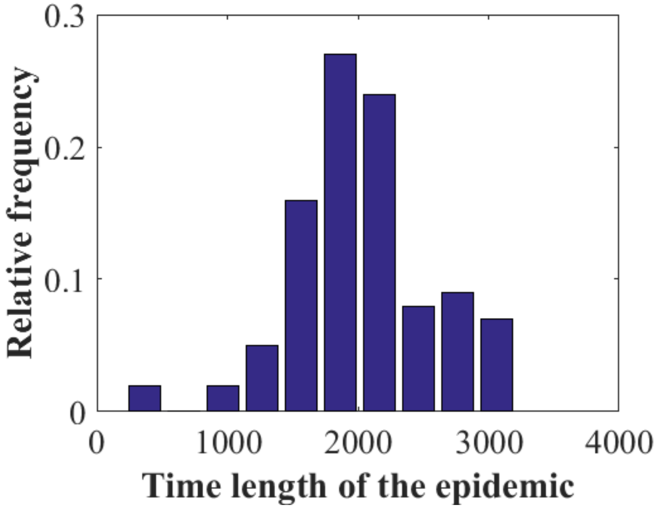


Figure 4.16: Time length of the epidemic with P = 0.004 with q = 0.5.

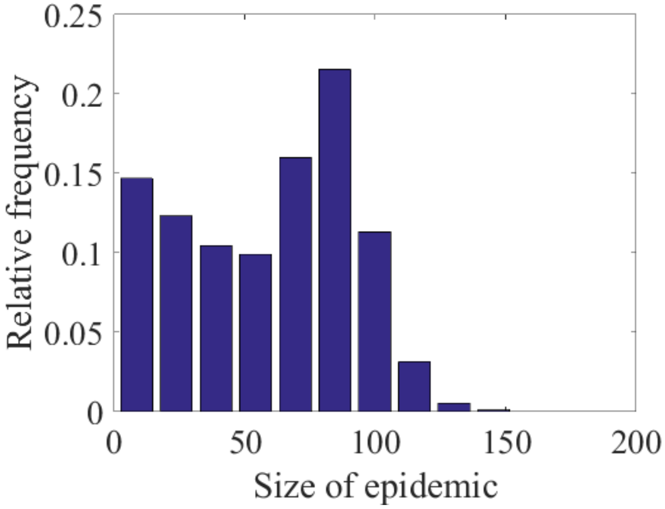


Figure 4.17: Number of vaccination with P = 0.002 with q = 0.5.

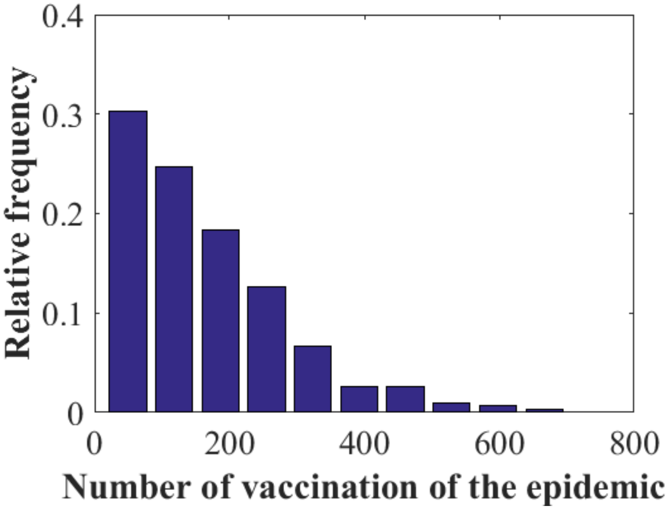


Figure 4.18: Number of vaccination with P = 0.004 with q = 0.5.

process. Perhaps we need more simulation results to find out the reason.

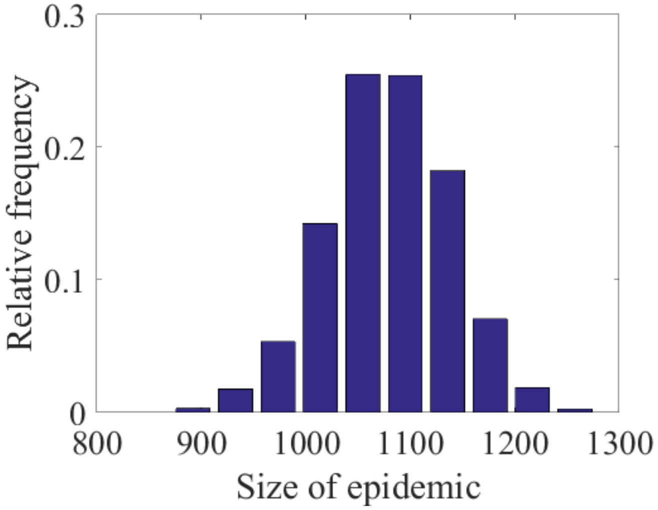


Figure 4.19: Number of epidemic incidence with P = 0.002 with q = 0.6.

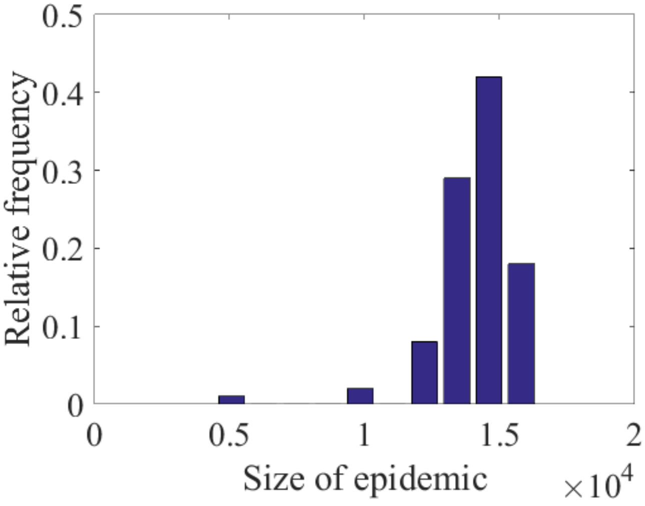


Figure 4.20: Number of epidemic incidence with P = 0.004 with q = 0.6.

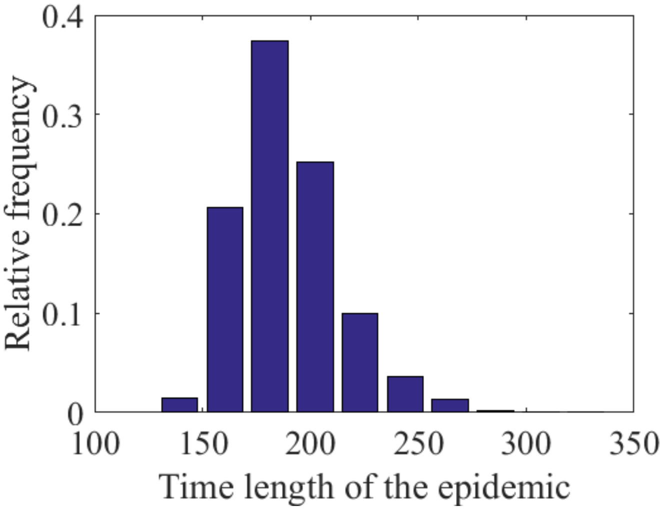


Figure 4.21: Time length of the epidemic with P = 0.002 with q = 0.6.

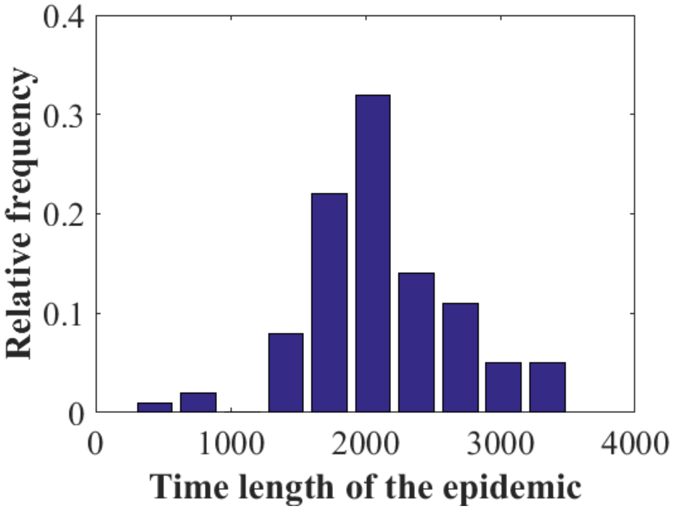


Figure 4.22: Time length of the epidemic with P = 0.004 with q = 0.6.

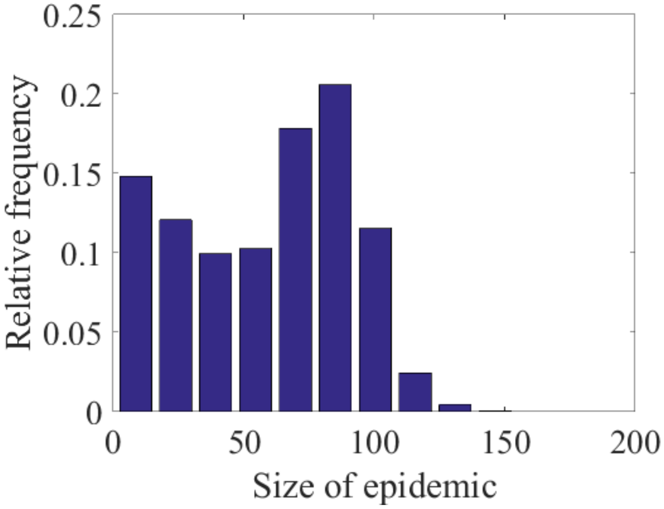


Figure 4.23: Number of vaccination with P = 0.002 with q = 0.6.

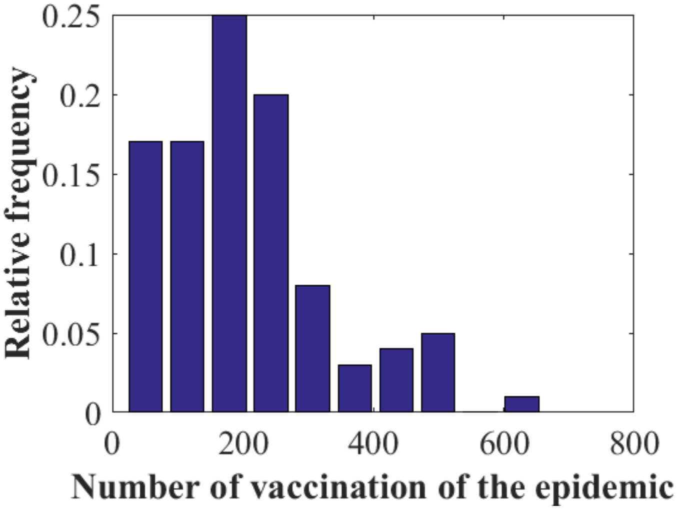


Figure 4.24: Number of vaccination with P = 0.004 with q = 0.6.

We are curious about the differences if we change some of the values of the parameters. We increase the probability that children connect with others from 0.002 to 0.004. Then the average degree increase from $5000 \cdot 0.002 = 10$ to $5000 \cdot 0.004 = 20$, meaning the children are now become more social and they contact with each other more frequently.

Figure 4.10-4.9 show the differences with different information cascades probability 'q' of 0.1, 0.5, 0.6, and 0.9, respectively. The reason of the differences are unknown for now. We need to do a lot more simulation runs to find out the real reason. But one thing we can tell directly from the histograms is that the size of epidemic is much larger after increasing the probability of connectivity. Clearly, more contacts between children increase the probability for those who are susceptible to get infected by the epidemic.

Figure 4.7-4.9 show the differences when increase β from 0.04 to 0.08, and decrease β_h from 0.02 to 0.01. The reason is unknown for now since this is just one of the many probabilities. Perhaps we will find the reason after we try different combinations of values of β and β_h. But, similarly, the size of epidemic is much larger.

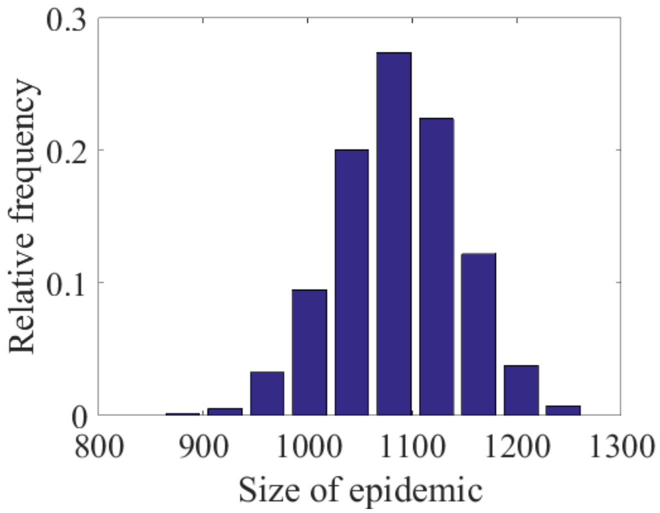


Figure 4.25: Number of epidemic incidence with P = 0.002 with q = 0.9.

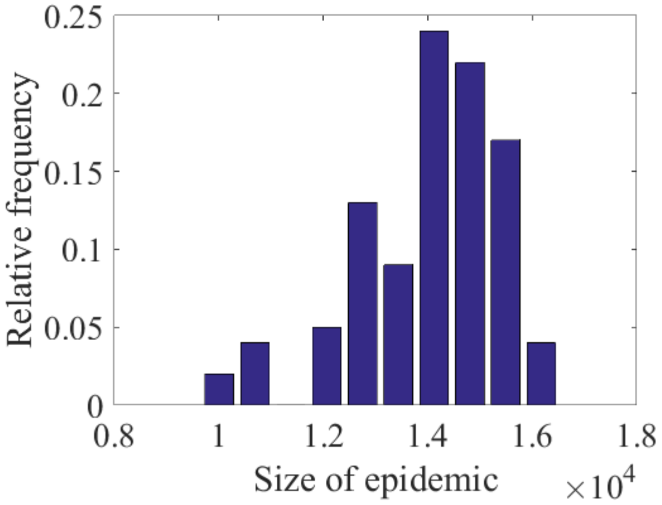


Figure 4.26: Number of epidemic incidence with P = 0.004 with q = 0.9.

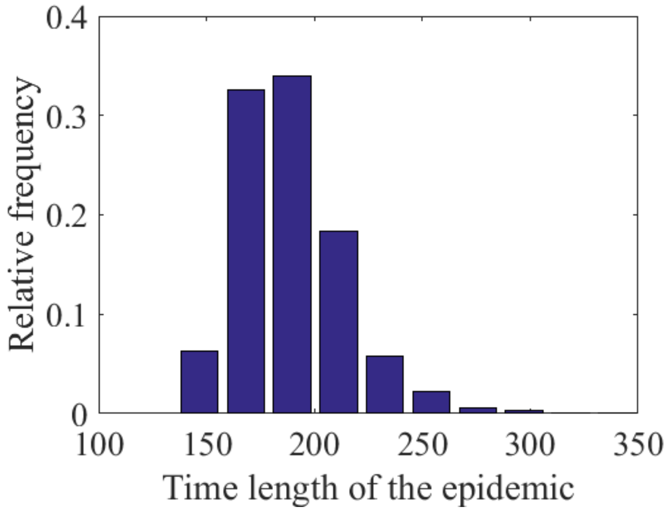


Figure 4.27: Time length of the epidemic with P = 0.002 with q = 0.9.

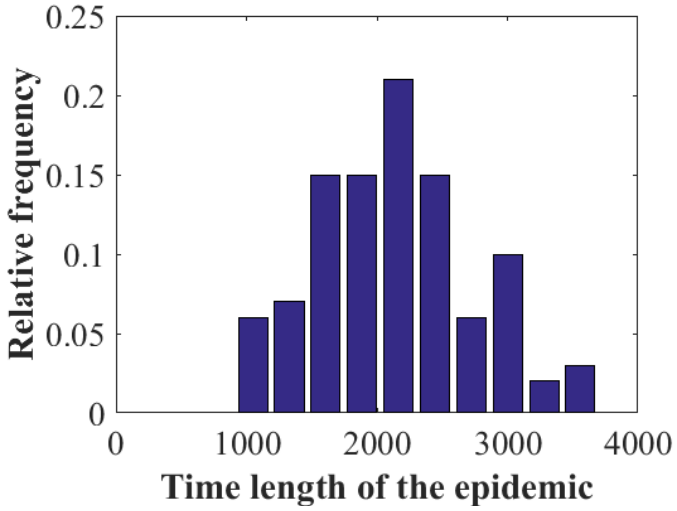


Figure 4.28: Time length of the epidemic with P = 0.004 with q = 0.9.

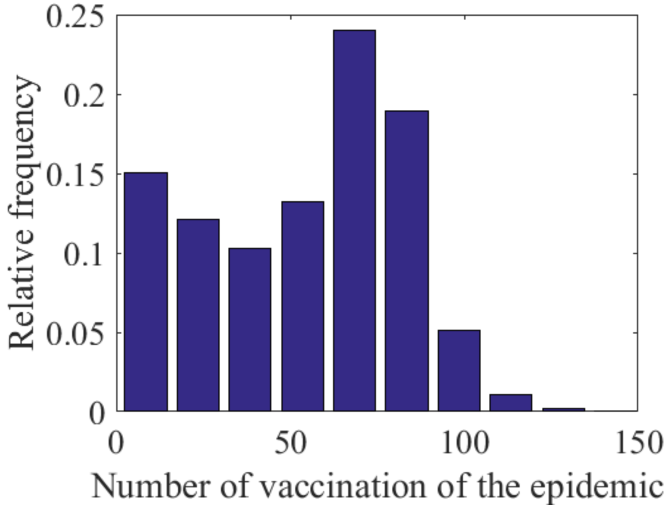


Figure 4.29: Number of vaccination with P = 0.002 with q = 0.9.

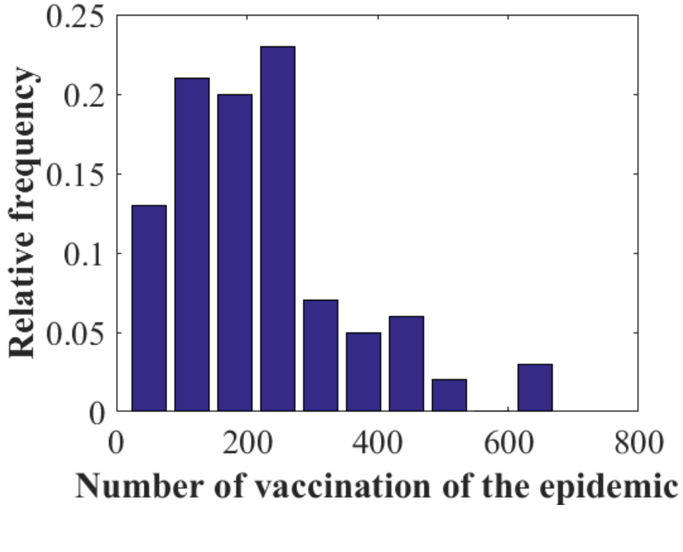


Figure 4.30: Number of vaccination with P = 0.004 with q = 0.9.

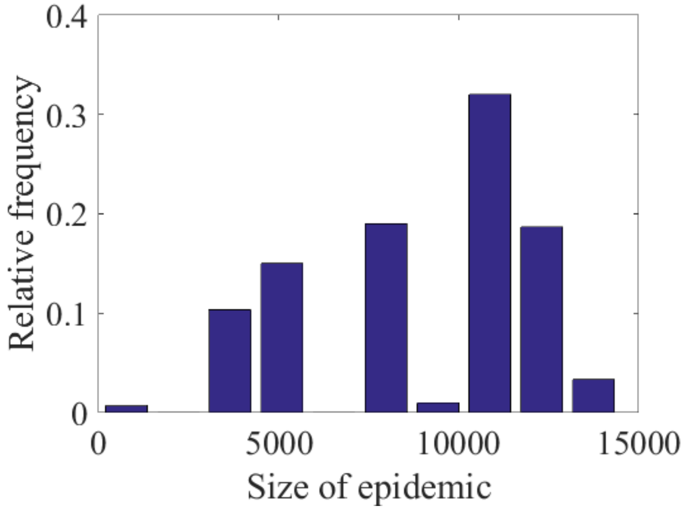


Figure 4.31: Number of epidemic incidence with P = 0.004 with q = 0.1.

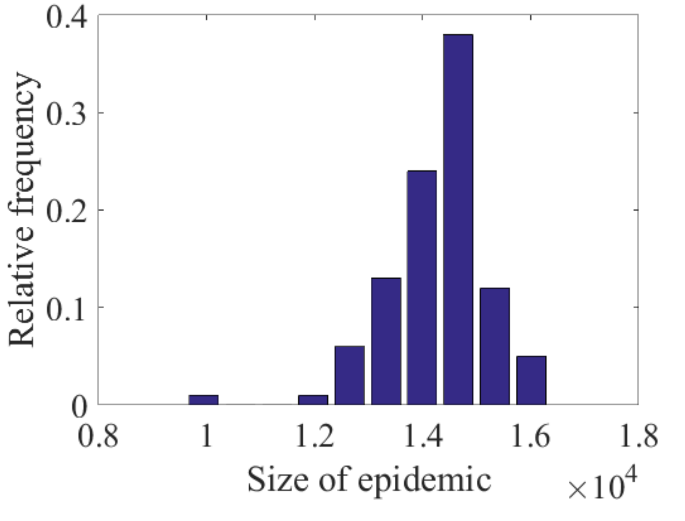


Figure 4.32: Number of epidemic incidence with P = 0.004 with q = 0.1.

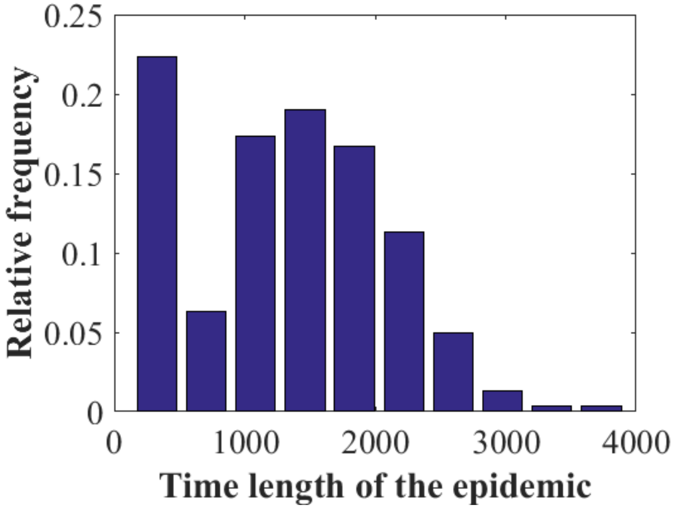


Figure 4.33: Time length of the epidemic with P = 0.002 with q = 0.1.

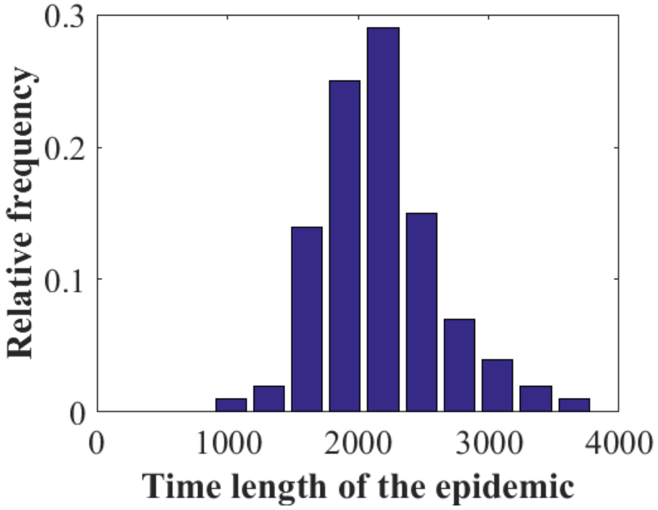


Figure 4.34: Time length of the epidemic with P = 0.004 with q = 0.1.

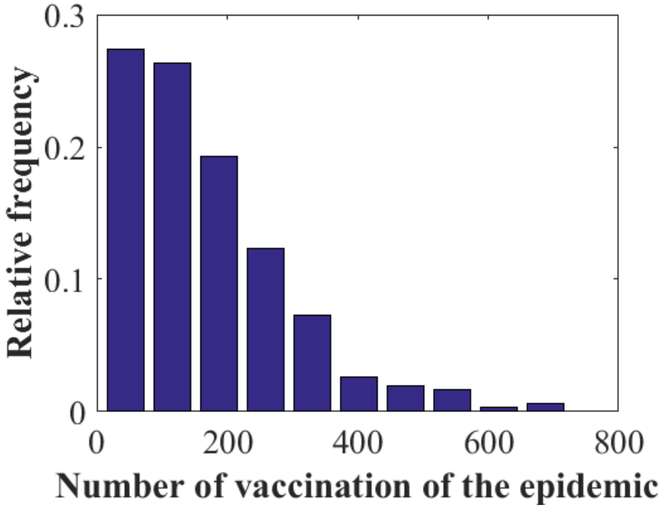


Figure 4.35: Number of vaccination with P = 0.002 with q = 0.1.

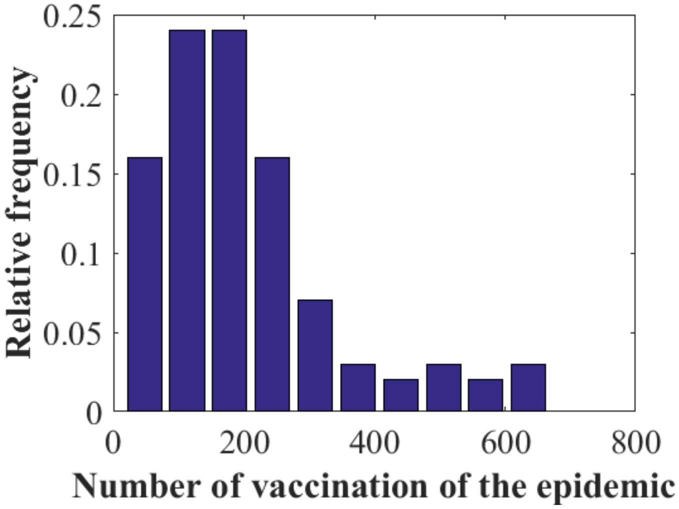


Figure 4.36: Number of vaccination with P = 0.004 with q = 0.1.

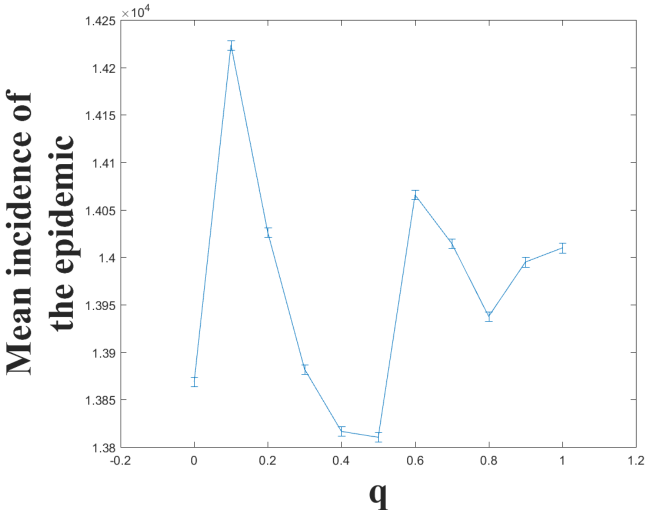


Figure 4.37: Mean number of vaccination of the epidemic disease with larger birth rate.

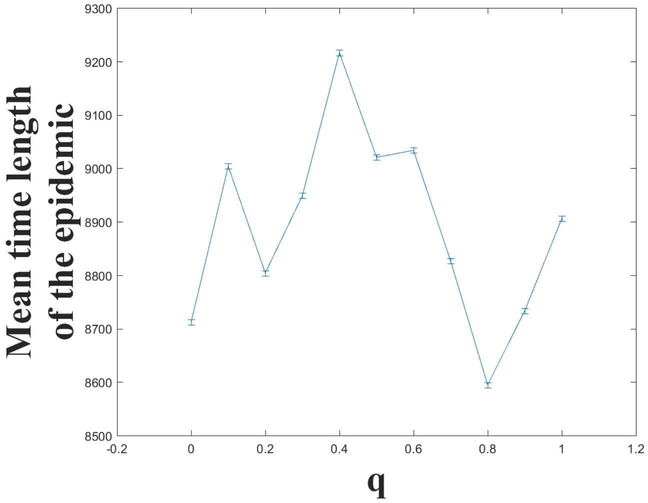


Figure 4.38: Mean number of vaccination of the epidemic disease with larger birth rate.

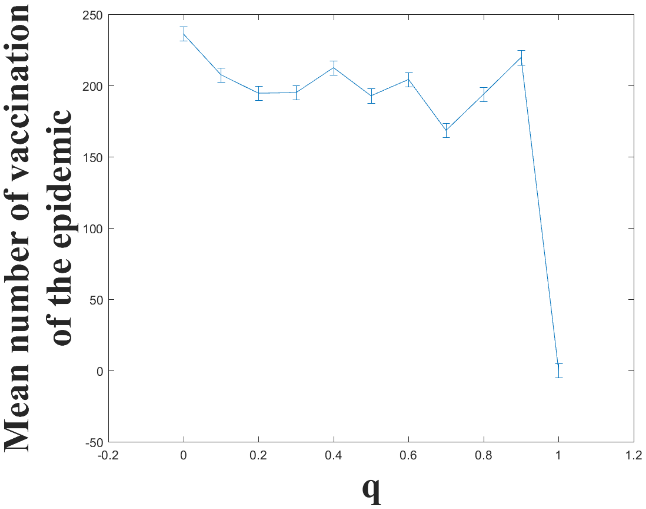


Figure 4.39: Mean number of vaccination of the epidemic disease with larger birth rate.

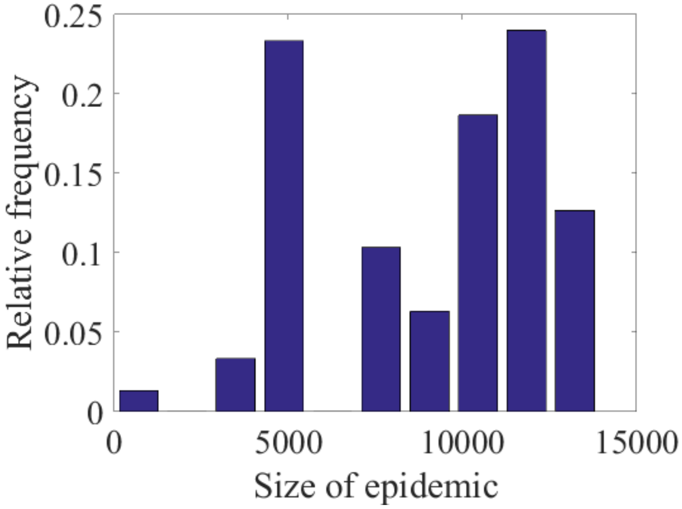


Figure 4.40: Number of epidemic incidence with P = 0.004 with q = 0.5.

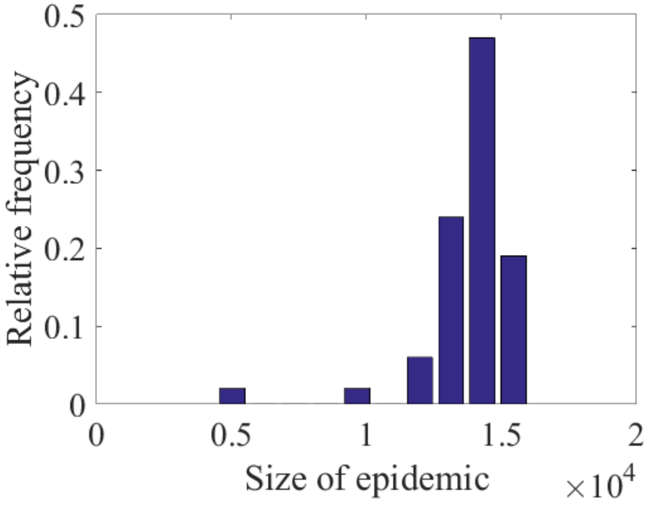


Figure 4.41: Number of epidemic incidence with P = 0.004 with q = 0.5.

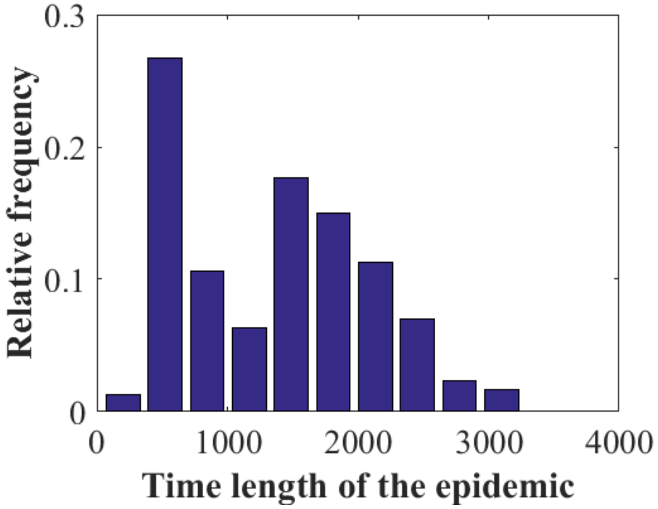


Figure 4.42: Time length of the epidemic with P = 0.002 with q = 0.5.

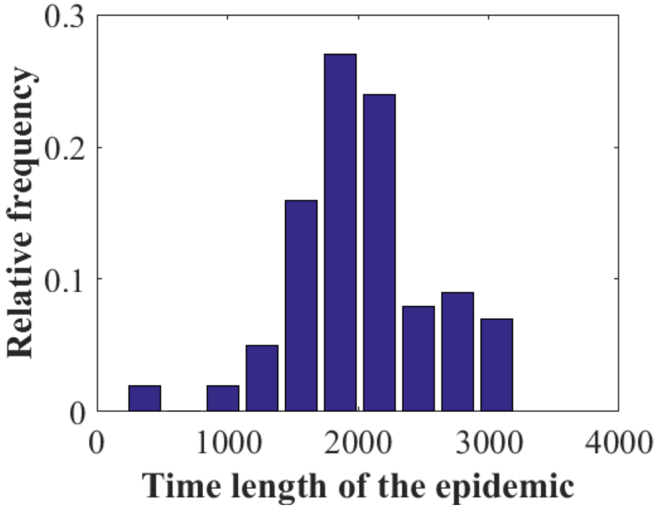


Figure 4.43: Time length of the epidemic with P = 0.004 with q = 0.5.

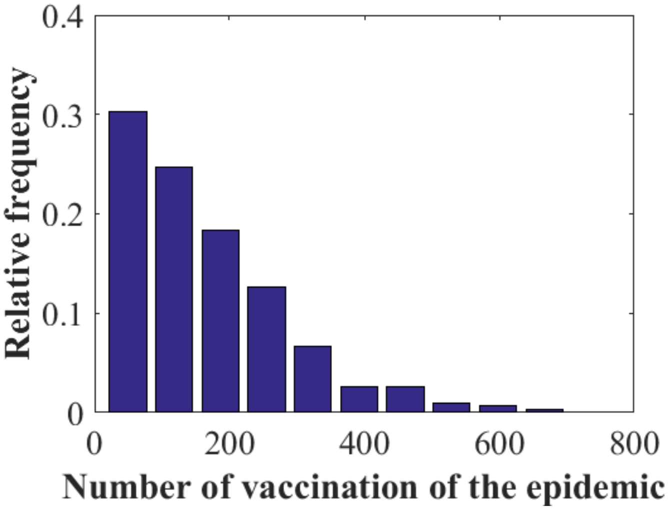


Figure 4.44: Number of vaccination with P = 0.002 with q = 0.5.

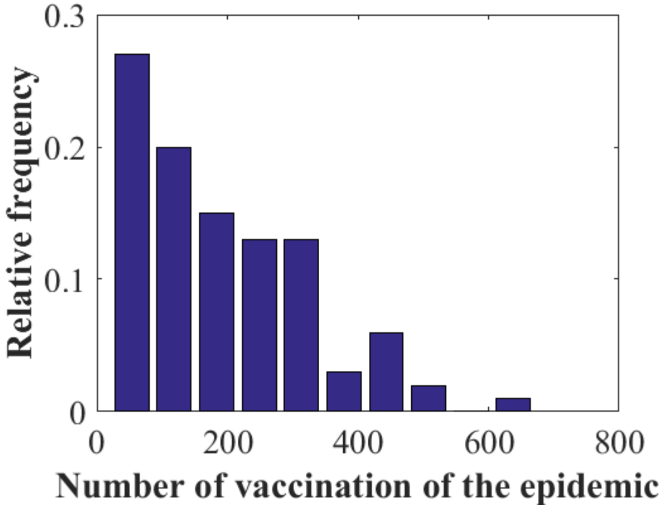


Figure 4.45: Number of vaccination with P = 0.004 with q = 0.5.

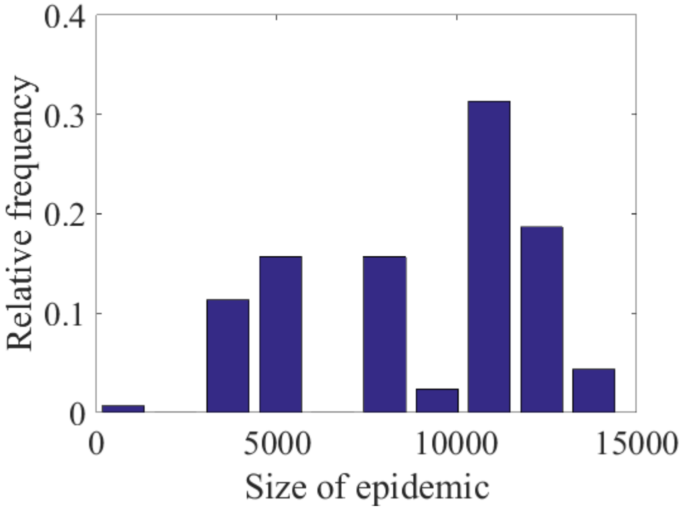


Figure 4.46: Number of epidemic incidence with P = 0.004 with q = 0.6.

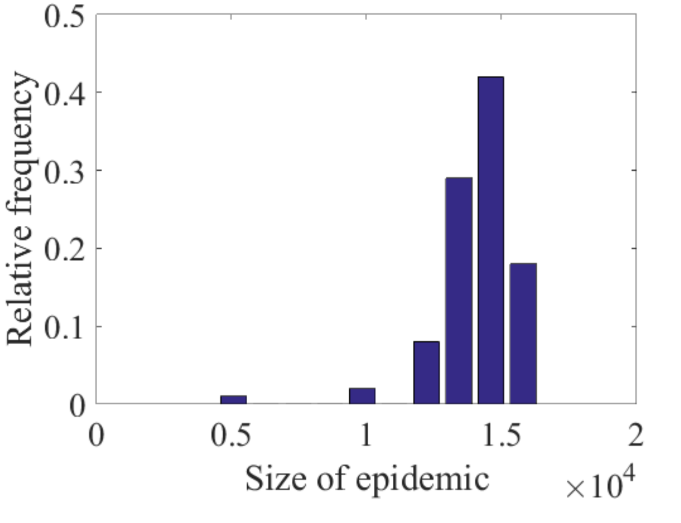


Figure 4.47: Number of epidemic incidence with P = 0.004 with q = 0.6.

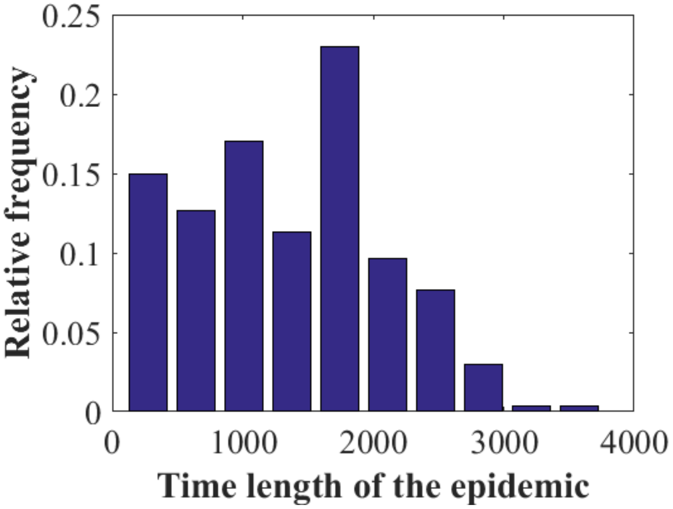


Figure 4.48: Time length of the epidemic with P = 0.002 with q = 0.6.

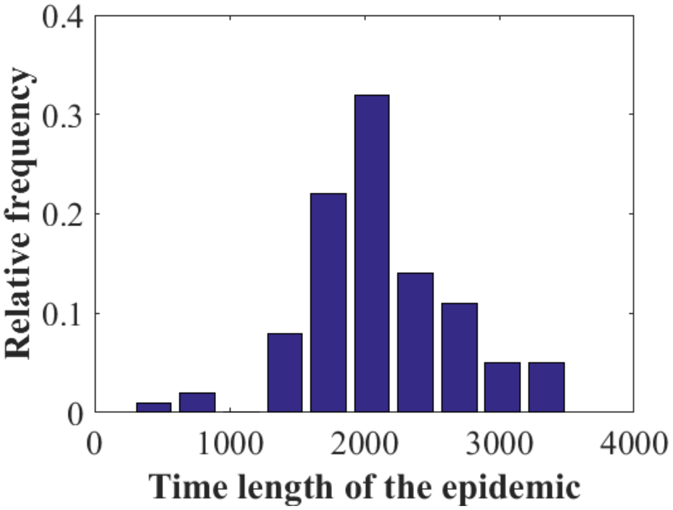


Figure 4.49: Time length of the epidemic with P = 0.004 with q = 0.6.

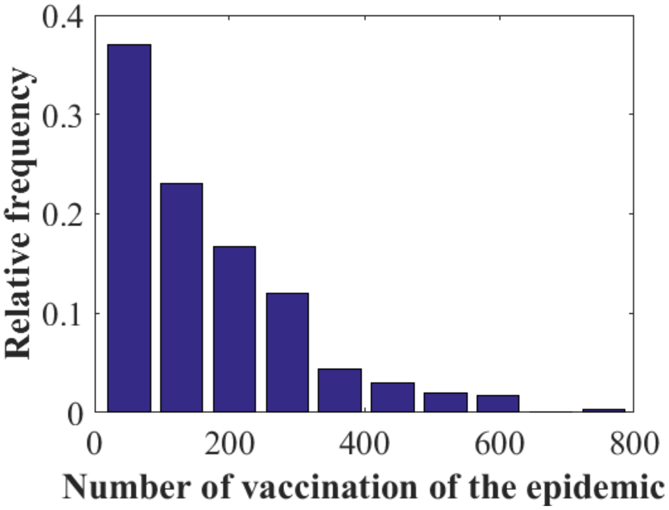


Figure 4.50: Number of vaccination with P = 0.002 with q = 0.6.

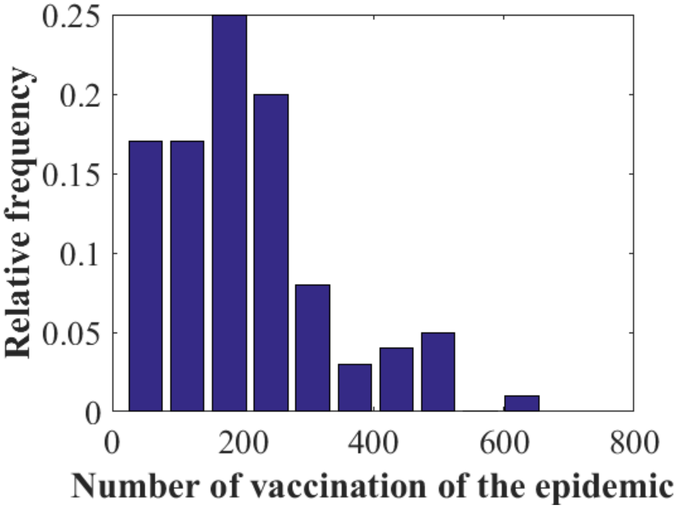


Figure 4.51: Number of vaccination with P = 0.004 with q = 0.6.

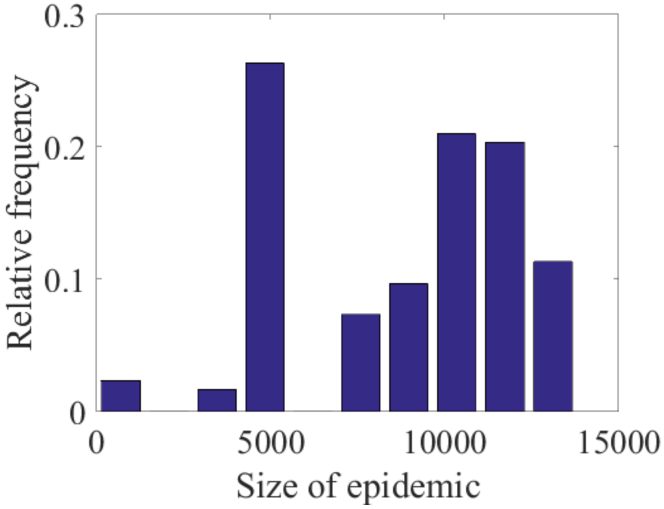


Figure 4.52: Number of epidemic incidence with P = 0.004 with q = 0.9.

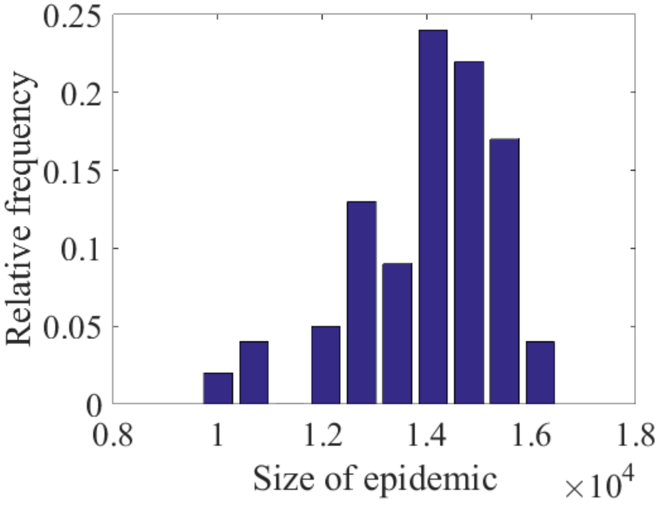


Figure 4.53: Number of epidemic incidence with P = 0.004 with q = 0.9.

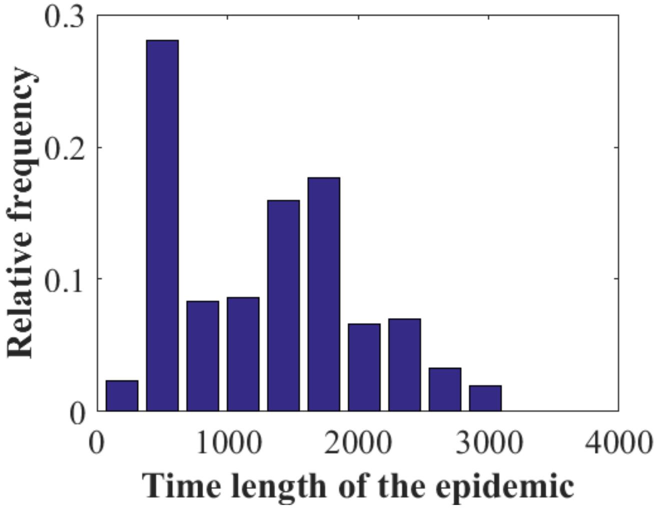


Figure 4.54: Number of epidemic incidence with P = 0.004 with q = 0.9.

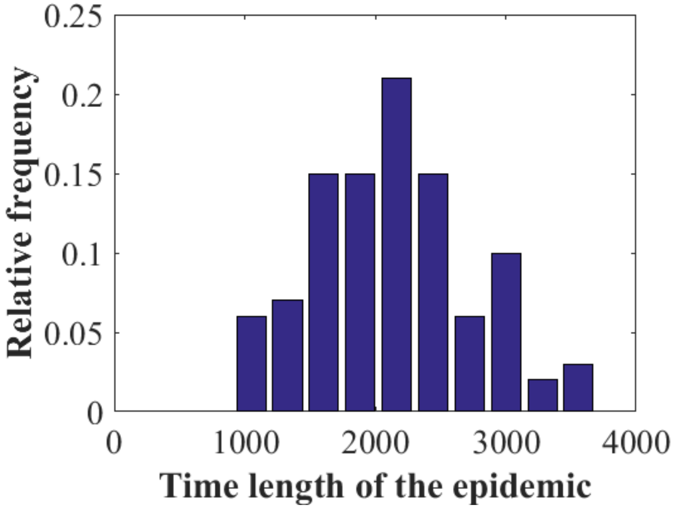


Figure 4.55: Time length of the epidemic with P = 0.004 with q = 0.9.

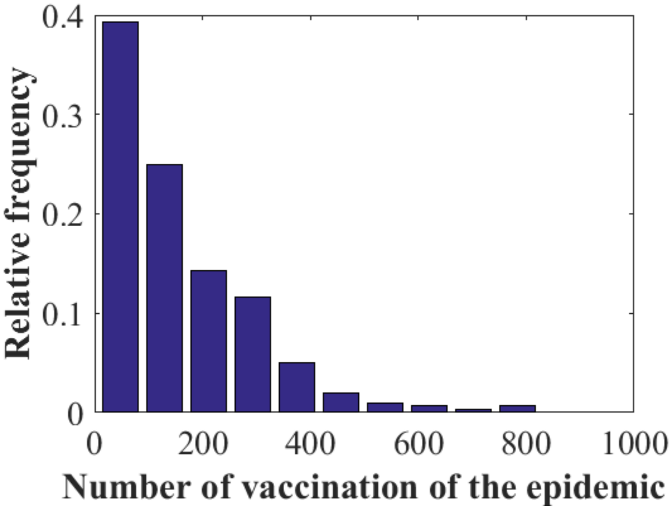


Figure 4.56: Number of vaccination with P = 0.002 with q = 0.9.

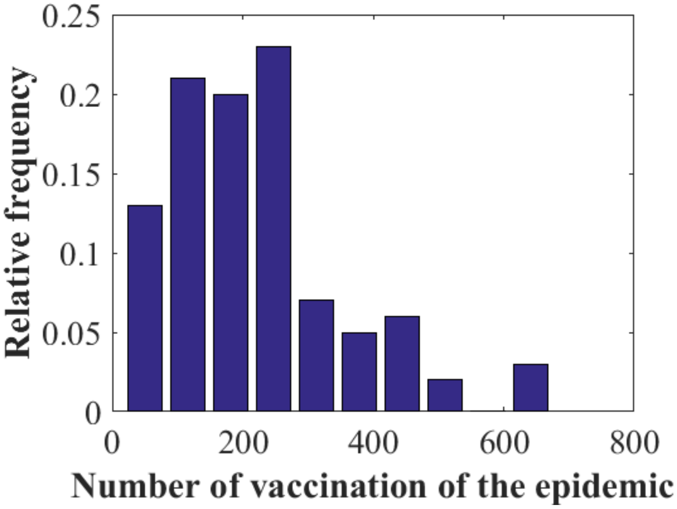


Figure 4.57: Number of vaccination with P = 0.004 with q = 0.9.

CHAPTER V

DISCUSSION AND FUTURE WORKS

Combining social network and the social network to study disease spread and vaccination opinion diffusion is a cutting-edge area. No one has actually studied ever before

Needless to say, what we have got is just part of the whole research. The results are preliminary and a lot of cases need to be investigated. Due to the limit time and computational requirement, we have to continue the work in our future research.

First, we mainly focus on the Erd*ö*s - Renyi network model since it is the most common type of the network. But there are many other types of networks such as Exponential Random Graph Model (ERGM) and the power law model. In the future, we want to compare the differences between different types of networks. For example, what if the children's biological network is Erd*ö*s - Renyi network model, the ERGM is for the parents' network.

In the future, we want to compare the differences between different types of networks. For example, the children's biological network we use the Erd*ö*s - Renyi network model, but use ERGM for the parents' networks. In addition, since we have a large number of parameters and since our agent-based model is difficult to predict the outcomes, we have to try different combination of parameters to get better results. Due to the limited time and computer memory, we are unable to try all the combinations. Maybe in the future, we can try the Latin-hypercube approach, which can provide us the best combination of parameters, to find the best result. Latin hypercube sampling [12] is inspired by the Latin square experiment design, which tried to eliminate the confounding effect of various experimental factors without increasing complexity of the experiment. The purpose of Latin hybercube sampling is to ensure that each value (or a range of values) of a variable is represented in the samples, no matter which value might turn out to be more important. [6]

We use pertussis as our epidemic disease to study in our dual model. Perhaps we can choose some more severe pediatric disease such as measles-mumps-rubella (MMR) to see if there are any differences in the results and to explore the reasons causing the potential differences.

There are a lot of different parameter values we can use for our model. In the future work, we can compare different results with different populations, or with different connection probabilities, i.e. to compare results in more sparse societies and more connected societies, etc.

Things become much more interesting when one considers the influence of information cascades. Since epidemic diseases with respect to people's opinions is a new field of modeling epidemics, a lot of work and research need to be done in order to thoroughly understand the model.

As concluded in the previous section, the characteristic of counterintuition is very interesting, but it is also a very complex phenomenon. This specific property of the agent-based model is actually a combination of psychological and sociological issues.

Bibliography

[1] D. Acemoglu and A. Ozdaglar, *Erdös-Renyi graphs and branching processes.*

[2] A. Annibale and O. T. Courtney, *The two-star model: exact solution in the sparse regime and condensation transition*, Journal of Physics A: Mathematical and Theoretical, 48 (2015), p. 365001.

[3] S. Bansal, B. T. Grenfell, and L. A. Meyers, *When individual behaviour matters: homogeneous and network models in epidemiology*, Journal of the Royal Society Interface, 4 (2007), pp. 879–891.

[4] C. T. Bauch and A. P. Galvani, *Epidemiology. Social factors in epidemiology*, Science (New York, NY), 342 (2013), pp. 47–49.

[5] E. Bonabeau, *Agent-based modeling: Methods and techniques for simulating human systems*, Proceedings of the National Academy of Sciences, 99 (2002), pp. 7280–7287.

[6] J. Cheng and M. J. Druzdzel, *Latin hypercube sampling in Bayesian networks*, in FLAIRS Conference, 2000, pp. 287–292.

[7] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*, Cambridge University Press, 2010.

[8] F. Fu, D. I. Rosenbloom, L. Wang, and M. A. Nowak, *Imitation dynamics of vaccination behaviour on social networks*, Proceedings of the Royal Society of London B: Biological Sciences, 278 (2011), pp. 42–49.

[9] Z. J and G. D, *Cs485 lecture 01: Large graphs.*

[10] M. O. Jackson, *Social and economic networks: Models and analysis.*

[11] M. L. N. Mbah, J. Liu, C. T. Bauch, Y. I. Tekel, J. Medlock, L. A. Meyers, and A. P. Galvani, *The impact of imitation on vaccination behavior in social contact networks*, PLoS Comput Biol, 8 (2012), p. e1002469.

[12] M. D. McKay, R. J. Beckman, and W. J. Conover, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 42 (2000), pp. 55–61.

[13] T. Oraby and C. T. Bauch, *Bounded rationality alters the dynamics of paediatric immunization acceptance*, Scientific reports, 5 (2015).

[14] T. ORABY, V. THAMPI, AND C. T. BAUCH, *The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases*, Proceedings of the Royal Society of London B: Biological Sciences, 281 (2014), p. 20133172.

[15] B. C. PERISIC A, *Social contact networks and disease eradicability under voluntary vaccination*, PLOS Computational Biology, 1 (1997), pp. 1–37.

[16] G. ROBINS, P. PATTISON, Y. KALISH, AND D. LUSHER, *An introduction to exponential random graph (p*) models for social networks*, Social networks, 29 (2007), pp. 173–191.

[17] M. SALATHÉ AND S. BONHOEFFER, *The effect of opinion clustering on disease outbreaks*, Journal of The Royal Society Interface, 5 (2008), pp. 1505–1508.

[18] T. A. SNIJDERS, P. E. PATTISON, G. L. ROBINS, AND M. S. HANDCOCK, *New specifications for exponential random graph models*, Sociological methodology, 36 (2006), pp. 99–153.

APPENDIX A

APPENDIX A

THE SIMULATION CODE

1.1 The MATLAB Code

DLayerN:

```
function [C,V]=DLayerN(q)
%% Physical (children contact) Network
n = 5000;
P = 0.002;
A=sparse(binornd(1,P,n,n));
A=triu(A)+tril(A')-2*diag(diag(A));

%% Household
pf=.3; % probability of having a child
ip=28; % maximum length incubation period (see below)
pnb=@(x)1./(1+exp(2.5*(x-2)));%probability of a new born (see below) reversed logistic or
%pnb=@(x)0.45.^x;%probability of a new born (see below) montonoically
%decreasing
nb=zeros(n,1);
ADvb=zeros(n,1); % number of vaccine adverse events
padv=.01; % probability of vaccine adverse event
alpha=.001; % household view of disease cases in its probability to vaccinate
gamma=200;% household view of adverse cases of vaccination in its probability to vaccinate
Initial=zeros(n,1);
Initial(datasample(1:n,10,'Replace',false))=1;
House=[binornd(1,pf,n,1) Initial zeros(n,1+ip)];%[S I1 ... Iip R V]
sH=sum(House(:,2:(1+ip)),2); % to see if houses are infectious in while statement

%% Disease
beta = 0.04; % network transmission
betah = 0.08; % within household transmission

pv=binornd(1,.5,n,1); % initial 0 or 1 status of a household being a vaccinator %
that is uniformally generated
```

```
P_INFOV=zeros(n,1); % probability to vaccinate due to social influence
%pv=binornd(1,.5,n,1); % initil vacinators is 50%
e=.95; % vaccine efficacy

%to build up transition probabilities for incubation period
shape = 22; % mean length of incubation period is shape*scale
scale = 1;
disc=0:ip;
cumprob=gamcdf(disc,shape,scale);
incubtransprob=diff(cumprob)./(cumprob(end)-cumprob(1:end-1)); % probability to move during incubation
to recovery=P(i<X<=i+1|i<X<=ip)

V=zeros(1,0); % vector of number of vaccinated
C=zeros(1,0); % vector of number of incidence
I=zeros(1,0); % vector of number of infected

%% Social (parents) Network
%we assume social network can overlap with physical network so we take A
%and clip parts of it then add other

ReP=.001; %link retain probability
AdP=.005; %add new link probability
B=sparse(binornd(A,ReP)+binornd(double(A==0),AdP));
B=triu(B)+tril(B')-2*diag(diag(B));

sodepth=1; %social influence depth

%% The process
SSnb=zeros(1,0);
day = 1;
while any(sH(:)>0);

    % Information cascade
    if(any(nb==280)>0)
        D=sparse(zeros(n,n));
        SCIV=pv;% to include household itself if vaccinates
        SCIN=1-pv; % to include household itself if don't vaccinate
% SCIV=zeros(n,1);
% SCIN=zeros(n,1);
        DD=1;
        s=1; % 0 for within household, 1 for close friends, 2 friends of friends
        while (s<=sodepth)&&(any(DD(:)>0))
            DD=(B^s-diag(diag(B^s))>0).*(D==0); % full of 1's for nodes of depth s
            D=D+DD;
            SCIV=SCIV+DD*double(pv); %socio-cultural influence upto distance sodepth between the agent
            % and counting vaccinators
            SCIN=SCIN+DD*(1-double(pv)); %socio-cultural influence upto distance sodepth between the agent
            % and counting non-vaccinators
            s=s+1;
        end

        ps = 1./(1+exp((gamma*sum(ADvb))-(alpha*sum(sum(I,1))))); % Looking through out the history of disease
        % and vaccination and have a personal prob to vaccinate
        P_INFOV = (ps.*(q.^(SCIV)).*((1-q).^(SCIN))) ./ (ps.*(q.^(SCIV)).*((1-q).^(SCIN)) +
        +(1-ps).*((1-q).^(SCIV)).*(q.^(SCIN)));
    end

    %disease spread process
    vb=binornd(single(nb==280),P_INFOV); % vaccinated newborn
    pv= (1-single(nb==280)).*pv+single(nb==280).*vb;% update status pv for only those had just a newbaby
    ADvb=ADvb+binornd(vb,padv);% adverse events due to vaccine
    V=[V;sum(vb)];
    House(:,ip+3)=House(:,ip+3)+vb; % update vaccinated
    House(:,1)=House(:,1)+binornd(vb,1-e)+(nb==280)-vb; % update susceptible with efficacy

    nb=(nb>0).*(nb<280).*(nb+1)+(nb==280).*0+binornd(single(nb==0),.0001*pnb(sum(House,2))); %
    pregnancy for 280 days, labor,
    and new carriage
    SSnb=[SSnb;sum((nb~=0))];
    if(day>280)
        torecover=binornd(House(:,2:(1+ip)),repmat(incubtransprob,n,1)); % generate who will recover
        recovery. % It may take longer to recover but we take only the infectious period into consideration.
        House(:,ip+2)= House(:,ip+2)+sum(torecover,2); % update recovered
        House(:,3:(1+ip))= House(:,2:ip)-torecover(:,1:end-1);% update incubating people

        X = A * any(sH>0,2);
        X=X./(House(:,1).*(House(:,1)>0)+(House(:,1)==0)); %adjust for personal network
        P_infection = 1- ((1-beta).^X).*((1-betah).^sH); %probability of infection from the
        network or withing household
        House(:,2) = binornd (single(House(:,1)>0),P_infection); %update incidence
        C=[C;sum(House(:,2))];
```

```
        House(:,1)=House(:,1)-House(:,2); %update susceptible
    end

    day = day+1;
    sH=sum(House(:,2:(1+ip)),2);
    I=[I;sum(sH)];
end
  I=I(281:end); V=V(281:end);
end
```

mainDLayerN:

```
tic
SC=zeros(1,0);
PC=zeros(1,0);
MC=zeros(1,0);
PV=zeros(1,0);
SV=zeros(1,0);
MV=zeros(1,0);
L=zeros(1,0);
PL=zeros(1,0);
ML=zeros(1,0);

for q=0:.1:1
    for i=1:100
        [C,V]=DLayerN(q);
        SC=[SC;sum(C)];
        SV=[SV,sum(V)];
        L=[L;length(C)];
    end
    PC=[PC;prctile(SC,[2.5 25 50 75 97.5],2)];
    MC=[MC;mean(SC)];
    PV=[PV;prctile(SV,[2.5 25 50 75 97.5],2)];
    MV=[MV;mean(SV)];
    PL=[PL;prctile(L,[2.5 25 50 75 97.5],2)];
    ML=[ML;mean(L)];
end
```

```
toc

time=toc;

save output.mat
```

BIOGRAPHICAL SKETCH

Shan Shan Zhao was born in Beijing, China on Jan. 13th, 1990.

She spent all her life in Beijing before coming to the United States. After graduating from the department of business at Beijing Wuzi University in 2012, she joined the master degree program in mathematics in the University of Texas Pan American in 2014. Her interest is in business analytics, which can also be called as the "Big Data."

From 2014-2016 she worked on her masters degree in mathematics (statistics concentration). She worked hard on building a solid mathematical and statistical foundation.

If you have any question about her work, you can reach Shan Shan by email at sszhao9001@gmail.com.