

The basic reproductive number (R_0) of SIS epidemics on networks

Haoyuan Wang



Delft University of Technology

The basic reproductive number (R_0) of SIS epidemics on networks

by

Haoyuan Wang

To obtain the degree of Master of Science

In Electrical Engineering

Track Wireless Communication and Sensing

at the Delft University of Technology,

to be defended publicly on Tuesday November 21st, 2023 at 13:30 PM.

Student number: 5278163

Project duration: March 2023 – November 2023

Thesis committee: Prof. dr. ir. P.F.A. Van Mieghem, TU Delft, chair
Dr.ir. G.F. Nane, TU Delft

Supervisors: Prof. dr. ir. P.F.A. Van Mieghem, TU Delft, thesis advisor
Dr. Fredrick Scott Dahlgren, TU Delft, daily supervisor

An electronic version of this thesis is available at <http://repository.tudelft.nl/>.



Preface

This thesis, “The basic reproductive number (R_0) of SIS epidemics on networks,” represents the end of my study at Delft University of Technology as a master's student in Electrical Engineering and the track of Wireless Communication and Sensing. This thesis project has been conducted at the Network Architectures and Services (NAS) group, where I received a lot of help and warmth. The past nine months have been one of the most precious experiences in my life.

First, I express my sincere gratitude to Professor Piet Van Mieghem for accepting me to work with the NAS group and providing all the timely support I need. I also want to give my great appreciation to Dr. Scott Dahlgren, my daily supervisor, for his kind and detailed guidance, patience, and encouragement throughout the thesis. It has been a great honor and fortune for me to work with him. Additionally, I would like to thank Professor Tina Nane for being on my thesis committee.

During my graduate study, I encountered many challenges and upset moments. Thus, I want to thank everyone who has helped me, encouraged me, and accompanied me. Especially my family and friends, who give me unconditional love and care. Without your support, I cannot accomplish what I have done.

Haoyuan Wang
Den Haag, November 2023

Abstract

The basic reproductive number (R_0) of an infectious disease is the expected number of secondary cases infected by the first case in an otherwise susceptible population. In this thesis, based on this fundamental concept, we propose a new method: the IIN-TN method that could calculate R_0 of a network from the local structure of the network under the scenario of the SIS model. The analytical solution of this method is given in this thesis, and simulations are performed to verify the result. After the code implementation in MATLAB, the method is applied to a human contact network in reality. The results from the real-world network show that the R_0 calculated by the IIN-TN method will be smaller than the value given by the traditional definition. Apart from that, we also find that the IIN-TN method is more effective for infectious diseases, which have a relatively larger infection probability of link and a higher effective infection rate.

Contents

Preface	2
Abstract	3
Contents	4
1. Introduction	6
1.1 Motivations	6
1.2 Objectives and Challenges	6
1.3 Thesis outline	7
2. Background and Basic Epidemiology	9
2.1 SIS(Susceptible-Infected-Susceptible) Model	9
2.2 Basic Epidemiology	10
3. Graph Theory	13
3.1 First approach: Holistic method	13
3.1.1 Possible epidemics on P3	13
3.1.2 Possible epidemics on K3(ignore cycles) and K3	16
3.1.3 Possible epidemics on K4	20
3.1.4 Disadvantage of holistic method	23
3.2 Second approach: IIN-TN method	24
3.2.1 Introduction of K2, PA + e and revisit of K3	25
3.2.2 Analytical solution of K2, PA + e and revisit of K4	29
4. Verification of Graph Theory by Simulation	39
4.1 Verification of K3(7)	39
4.2 Verification of three cases in K4	41
4.3 Verification of Pr(IIN → TN GI) in K2, PA + e	42
5. Application and Results	45
5.1 Code implementation of IIN-TN method	45
5.2 Application in human-contact network	48
5.2.1 Dataset	48
5.2.2 Infectious Disease	49
5.3 Results	50
6. Conclusion and Future Research	57
6.1 Conclusion	57
6.2 Future research	58
Bibliograph	60
Appendix	63
Appendix A: Derivation	63
Appendix B: Nomenclature	70
Appendix C: Codes	71

1. Introduction

1.1 Motivations

The basic reproductive number (R_0) of an infectious disease is the expected number of secondary cases infected by the first case in an otherwise susceptible population [1]. Modelers estimate the basic reproductive number by fitting an exponential curve to a plot of new cases versus time from a statistical perspective [2, 3]. However, by combining the SIS epidemic model with further research on network topology, there might be an alternative approach to directly derive and calculate the basic reproductive number of a given network.

The SIS epidemic model allows for the introduction and spread of an infectious disease across a network. The expected number of secondary cases depends on the local structure of the network where the disease originates. Different starting points may result in varying patterns of epidemic spread across the network, and the basic reproductive number on a network represents the expected number of secondary cases averaged across all nodes.

The observed variation in the basic reproductive number during outbreaks of infectious diseases is often attributed to differences in population immunity, evolution of the infectious agent, and other significant factors. However, to our knowledge, the contribution of network topology to this variation remains unexplored. According to Socievole et al. [4], the R_0 cannot summarize and reflect the information from the underlying contact network because it does not do enough assessment of the threshold behavior of an outbreak of an infectious disease. Therefore, the objective of this thesis is to investigate how the local structure influences the R_0 on a network and whether this measure can capture the behavior of different epidemic outbreaks in reality.

1.2 Objectives and Challenges

Based on the motivation mentioned above, the objectives of our research are as follows:

1. Carefully define the basic reproductive number and explain the concept through some examples, such as P_3 , K_3 , and K_4 graph.
2. Derive an analytical solution for the basic reproductive number of networks.
3. Implement the algorithm in MATLAB and apply the algorithm to the networks from the real world.
4. Compare the results from different infectious diseases and further study the characteristics of the algorithm.

Based on the objective mentioned above, several challenges need to be solved:

1. Derive a general way that could compute the probability of getting certain G_I (Graph with infection process) in a complex network, especially if more paths exist at the same time.
2. Based on the first challenge, derive a general equation that could calculate the possibility of the direct infection from IIN (Initial infected node) to its NNs

- (neighboring nodes) with given G_I under the condition that:
- 2.1 The alternative path has two links.
 - 2.2 More than 2 paths exist.
 - 2.3 Paths may be dependent and share links.
 - 2.4 How to simplify the structure to make the calculation possible. It is nearly impossible to directly calculate the probability because an analytical solution requires knowledge of all paths.
3. After solving the first and second challenges, we need to derive an algorithm to calculate R_0 for any network. Then, we implement the algorithm in MATLAB and apply it to a network from the real world.

1.3 Thesis outline

The structure of this thesis is as follows: Chapter 2 introduces the background of the SIS epidemics model and the basic epidemiology. Some fundamental concepts in this field will be discussed. In Chapter 3, a detailed theory regarding how to calculate the R_0 will be given through several examples. Besides, a general expression will be derived for different networks. Then, in Chapter 4, simulations will be carried out to verify the analytical expression proposed in Chapter 3. After that, results from real-world networks will be obtained after implementing the algorithm on MATLAB in Chapter 5. Besides, the discussion of results and methods will also be given in Chapter 5. At last, a conclusion and possible scope for future work will be given in Chapter 6.

2. Background and Basic Epidemiology

2.1 SIS(Susceptible-Infected-Susceptible) Model

The SIS model is one of the most commonly used models to describe epidemics in networks nowadays. The node in a network only has two states in the SIS model: infected or susceptible. The SIS model allows nodes to change states continuously over time. Actually, the SIS model consists of two independent Poisson processes: the curing process with rate δ and the infection process with rate β [5]. This means that an infected node could cure back to susceptible again with rate δ , and a susceptible node could be infected by the connection with an infected node with rate β .

The graph with the infection process is G_I , which not only contains the topology information but also includes infected nodes and the link that spread the epidemic. Based on the two Poisson processes, we could present the probability that a susceptible node gets infected by the connection with an infected node as:

$$Pr(I < R) \quad (2.1)$$

Where I represents the amount of time before the infection event happens, and R represents the amount of time before the recovery event happens. Thus, Equation (2.1) represents the probability that an infected node will infect a susceptible node through the connected links, because the amount of time required by the infection event is shorter than the recovery event. Since both the infection and curing processes are Poisson process with rate β and δ separately and we focus on the amount of time that is required before the event happen. The time needed before the first event happening is described by the PDF of exponential distribution:

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x \leq 0 \end{cases} \quad (2.2)$$

Substitute $\lambda = \beta$ and $\lambda = \delta$ to Equation (2.2) and we could get the time before the first infection event happening and first curing event happening as:

$$f_I(t) = \beta e^{-\beta t}, \quad (2.3)$$

$$f_R(r) = \delta e^{-\delta r}. \quad (2.4)$$

Where t from Equation (2.3) is the continuous time variable for the infection process, and r from Equation (2.4) is the continuous time variable for the recovery(curing) process. Then, based on the two equations, we could find the probability by integration :

$$\begin{aligned} Pr(I < R) &= \int_{-\infty}^{\infty} Pr(I < R | R = r) f_R(r) dr \\ &= \int_{-\infty}^{\infty} Pr(I < r) f_R(r) dr \\ &= \int_{r=0}^{\infty} \int_{t=0}^r f_I(t) f_R(r) dt dr \end{aligned}$$

$$= \frac{\beta}{\beta + \delta} \quad (2.5)$$

$$\begin{aligned} Pr(I > R) &= 1 - Pr(I < R) \\ &= \frac{\delta}{\beta + \delta} \end{aligned} \quad (2.6)$$

Detailed integration process is given in Appendix A.

The results from Equation (2.5) and Equation (2.6) tell us that the susceptible node will be infected by an infected node with a probability of $\frac{\beta}{\beta + \delta}$ through the link with and will not be infected with a probability of $\frac{\delta}{\beta + \delta}$. The two probabilities enable us to find the probability of a graph with a complex infection process by multiplying the two probabilities according to the permutations of infection/non-infection events:

$$Pr(G_I) = \left(\frac{\beta}{\beta + \delta} \right)^m \left(\frac{\delta}{\beta + \delta} \right)^n \quad (2.7)$$

Where G_I represents the graph with the infection process. In a given G_I , based on the infection process on it, we could find the number of links that pass the disease, and which do not. Thus, m is the number of links that pass the disease (infection events happen), and n is the number of links that do not pass the disease (non-infection events happen).

2.2 Basic Epidemiology

In epidemiology, the R_0 of an infectious agent is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infections [1]. From the definition, we need to focus on several key concepts:

- i. The R_0 is an expected number which indicates the R_0 is an anticipated or average number from the statistical estimate.
- ii. The R_0 only counts the number of infected cases which are directly generated by the one case. This emphasizes that we will have one and only one initial infected case in the population. And only the number of secondary cases directly infected by the first case should be regarded as R_0 .
- iii. All other individuals in the population are initially susceptible which enables us to use the SIS model.

From the above definition, we can know that R_0 does not represent the total number of cases introduced by the first infected case within a specific time frame. Actually, if we want to measure R_0 in a given network, R_0 is related to the following factors:

- i. The number of neighboring nodes (NNs) of the initial infected node (IIN).
The maximum number of R_0 of a node is the number of NNs of that node if we ignore possible reinfections in the SIS model. The actual number depends on the infection rate β and curing rate δ , as well as the possible existence of cycles. Detail about cycles is introduced in Section 3.1.
- ii. The local structure of network around IIN.
The local structure is the basic condition for the infection process. The infection can only take the path given by the local structure.

- iii. The possible different infection processes.

Based on the local structure, the infection process may be different. We need to find the possibilities of each infection process together with the number of nodes directly infected by IIN.

Thus, based on the definition and the important factors, we propose our guiding concept to calculate the R_0 of a given network as follows:

- i. For a given IIN, the epidemic may propagate on different links, leading to different routes and infection processes. Thus, we need first to find all the possible infection processes. For each infection process, find the probability and the number of nodes directly infected by the IIN.
- ii. Calculate the R_0 of the given IIN by summing up the products of possibility and number of nodes for each possible infection process as calculated above.
- iii. Repeat step 1 and step 2, assign each node in the given network as the IIN separately, and find all the corresponding R_0 of node under the circumstances of different node as IIN.
- iv. Sum up all the R_0 of node and take the average of R_0 over the node set to derive the R_0 of the network.

Based on this guiding concept, we proposed two methods to find the R_0 of the network. The first one is the holistic method, which will give the infection process over the whole network by identifying the status of each link in the graph. The first method is presented and explained in Section 3.1. The second approach is the Initial infected node to Target node (IIN-TN) method, which will only focus on the infection process between the IIN and NN. The second approach using the IIN-TN method will be introduced in Section 3.2.

3. Graph Theory

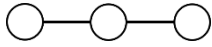
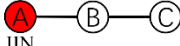
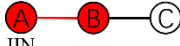

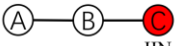
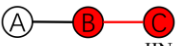

In this chapter, two methods, the holistic method and the IIN-TN method, will be introduced and explained in detail with examples about how to calculate R_0 of networks separately in Section 3.1 and Section 3.2. In general, the holistic method is a straightforward method that could demonstrate our concepts and ideas very well. Thus, in Section 3.1, four examples illustrate the calculation process. However, during this process, the shortcomings of the holistic method will be highlighted when the example networks gradually have more nodes and links. In order to overcome these shortcomings, the IIN-TN method is proposed, where IIN-TN stands for Initial Infected Node to Target Node. The IIN-TN method will significantly reduce computational complexity at the expense of accuracy. However, only the IIN-TN method allows us to handle large and complex networks. The comparison of results of the same examples from the two methods will be given as well.

3.1 First approach: Holistic method

The holistic method focuses on IIN and the local structure around it. The most important feature of this method is that it will take every neighboring node (NN) around IIN into consideration and calculate the probability of the direct infection from IIN to all NNs as a whole. This holistic consideration characteristic gives this method its name: The holistic method. A simple example of P_3 is given below.

3.1.1 Possible epidemics on P_3

With the Equation (2.7), we could derive the table of P_3 as follows:

Possible epidemics on P_3		
G : 		
G_I	$Pr(G_I)$	$R_0(G_I)$
(1.1) 	$\frac{\delta^2}{(\beta + \delta)^2}$	0
(1.2) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(1.3) 	$\frac{\beta^2}{(\beta + \delta)^2}$	1
(2.1) 	$\frac{\delta^2}{(\beta + \delta)^2}$	0
(2.2) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(2.3) 	$\frac{\beta^2}{(\beta + \delta)^2}$	1

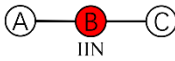
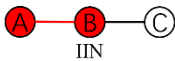
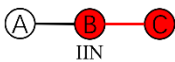

(3.1)		$\frac{\delta^2}{(\beta + \delta)^2}$	0
(3.2)		$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(3.3)		$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(3.4)		$\frac{\beta^2}{(\beta + \delta)^2}$	2

Table 1: All possible epidemics in P_3 . The hollow dot represents the susceptible node while the red dot represents the infected node. The black line indicates the link between nodes and no infection goes through it. The red line indicates infection event happens through this link.

G_I represents the graph with infection process; $Pr(G_I)$ represents the possibilities of that graph; $R_0(G_I)$ indicates the number of nodes directly infected by IIN in that graph with infection process. It should be noted that, the $R_0(G_I)$ for the graph $P_3(1.3)$ and $P_3(2.3)$ are both 1 instead of 2 because only one susceptible node has direct link with the initial infected node(IIN). According to the definition of R_0 in Section 2.2, even if the last susceptible node on the other edge is infected, it should not be considered as the R_0 of the IIN. Actually, $R_0(G_I)$ is limited by the number of NNs of IIN because only the NNs of IIN can be directly infected by IIN. A detailed discussion of $R_0(G_I)$ will be given in the part discussing about K_3 graph and K_4 graph when cycles of the infection process exist.

After listing all the possible graphs with infection process G_I , and the probability of having this graph $Pr(G_I)$ together with the basic reproductive number of this graph $R_0(G_I)$, we could calculate the R_0 of P_3 as the summation of each multiplication of $Pr(G_I)$ and $R_0(G_I)$ and then divide by the N, the number of nodes in P_3 :

$$R_0(P_3) = R_0(G) = \frac{\sum Pr(G_I) R_0(G_I)}{N} \quad (3.1)$$

Based on the above table, we can substitute all the probabilities and values to Equation (3.1) and get:

$$R_0(P_3) = \frac{4\beta\delta + 4\beta^4}{3(\beta + \delta)^2} \quad (3.2)$$

If $\beta = \delta$, substitute to Equation (3.2) and get:

$$R_0(P_3) = \frac{2}{3} \quad (3.3)$$

When $\beta = \delta$, the infection probability $\lambda = \frac{\beta}{\beta + \delta} = \frac{1}{2}$. By our intuitive understanding, the possibilities for the IIN located at the edge or the center in P_3 are $\frac{2}{3}$ and $\frac{1}{3}$ respectively. The different location of the IIN will give a different expected R_0 . The expected R_0 will be $1 \times \frac{1}{2} = \frac{1}{2}$ if the IIN is at the edge of the P_3 graph, and it will be $2 \times \frac{1}{2} = 1$ if the IIN locates at the center of the P_3 graph. Thus, the R_0 of P_3 could also be simply calculated as:

$$R_0(P_3) = \frac{1}{2} \times \frac{2}{3} + 1 \times \frac{1}{3} = \frac{2}{3} \quad (3.4)$$

The results from Equation (3.3) and Equation (3.4) are the same when substituting $\beta = \delta$, thus the result of Equation (3.2) is reasonable and satisfies our expectation.

Another important feature that we could get from the above example on P_3 is the symmetry of graph. Which is shown in below Table 2 and Table 3:

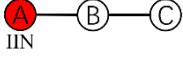
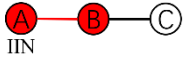
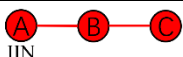
G_I	$Pr(G_I)$	$R_0(G_I)$
(1.1) 	$\frac{\delta^2}{(\beta + \delta)^2}$	0
(1.2) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(1.3) 	$\frac{\beta^2}{(\beta + \delta)^2}$	1

Table 2: Part of Table 1, presenting all the cases when Node A is the IIN. The IIN for these cases is on the left edge of P_3 graph. Except the different location of IIN, Table 2 is the same with Table 3.

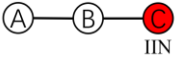
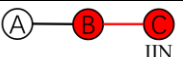
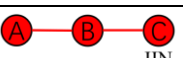
G_I	$Pr(G_I)$	$R_0(G_I)$
(2.1) 	$\frac{\delta^2}{(\beta + \delta)^2}$	0
(2.2) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(2.3) 	$\frac{\beta^2}{(\beta + \delta)^2}$	1

Table 3: Part of Table 1, presenting all the cases when Node C is the IIN. The IIN for these cases is on the right edge of P_3 graph. Except the different location of IIN, Table 3 is the same with Table 2.

Tables 2 and 3 have the same value of $Pr(G_I)$ and $R_0(G_I)$ for each corresponding G_I . Apart from that, each G_I is very similar, and the only difference is whether the IIN is on Node A or Node C. The symmetry of the graph causes this phenomenon. If we do not label each node in P_3 graph, it will look the same when IIN starts at the left edge or right edge. After we put the label, we realize they are different cases but just with the same $Pr(G_I)$ and $R_0(G_I)$. This phenomenon also happens in the K_3 and K_4 graphs.

3.1.2 Possible epidemics on $K_{3(\text{ignore cycles})}$ and K_3

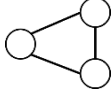
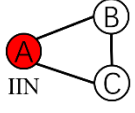
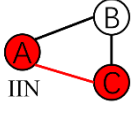
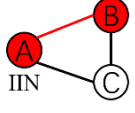
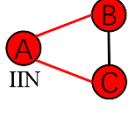
Possible epidemics on $K_{3(\text{ignore cycles})}$		
G: 		
G_I	$Pr(G_I)$	$R_0(G_I)$
(1) 	$\frac{\delta^2}{(\beta + \delta)^2}$	0
(2) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(3) 	$\frac{\beta\delta}{(\beta + \delta)^2}$	1
(4) 	$\frac{\beta^2}{(\beta + \delta)^2}$	2

Table 4: All possible epidemics on $K_{3(\text{ignore cycles})}$. Table 4 only represents the cases when Node A is the IIN. Due to symmetry of K_3 , when Node B and Node C is the IIN, all cases from them will be the same as when Node A is the IIN.

The notation “ $K_{3(\text{ignore cycles})}$ ” indicates that infection will not happen on the link between Node B and Node C as G_I in Table 4 shows. This is a special case, and the infection through the link between Node B and Node C should not be forbidden theoretically. The reason for doing this is to demonstrate the difference between $R_0(G)$ result for both considering the case with a cycle and case without cycles. Detailed comparison and discussion will be given in the next section based on the calculation result.

Based on Table 4, the basic reproductive number of $K_{3(\text{ignore cycles})}$ could be calculated as:

$$R_0(K_{3(\text{ignore cycles})}) = R_0(G) = \frac{3 \times \sum Pr(G_I) R_0(G_I)}{N} \quad (3.5)$$

As Equation (3.5) shows, the coefficient 3 is introduced by the symmetry of K_3 . Table 4 only demonstrates all possible G_I when IIN starts at Node A. Two tables identical to Table 4 can also be derived when IIN starts at Node B or Node C. Thus, there are a total of 3 times of $\sum Pr(G_I) R_0(G_I)$, where each $Pr(G_I)$ and $R_0(G_I)$ of corresponding G_I are given in Table 4.

Substitute all probabilities and values from Table 4, together with $N = 3$ to Equation (3.5):

$$R_0(K_{3(\text{ignore cycles})}) = \frac{2\beta\delta + 2\beta^2}{(\beta + \delta)^2} \quad (3.6)$$

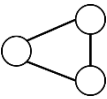
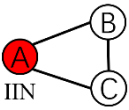
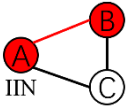
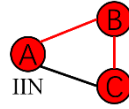
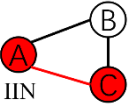
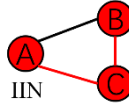
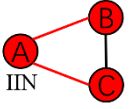
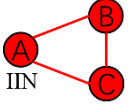
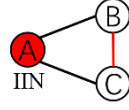
Possible epidemics on K_3		
G: 		
G_I	$Pr(G_I)$	$R_0(G_I)$
(1) 	$\frac{\delta^3}{(\beta + \delta)^3}$	0
(2) 	$\frac{\beta\delta^2}{(\beta + \delta)^3}$	1
(3) 	$\frac{\beta^2\delta}{(\beta + \delta)^3}$	1
(4) 	$\frac{\beta\delta^2}{(\beta + \delta)^3}$	1
(5) 	$\frac{\beta^2\delta}{(\beta + \delta)^3}$	1
(6) 	$\frac{\beta^2\delta}{(\beta + \delta)^3}$	2
(7) 	$\frac{\beta^3}{(\beta + \delta)^3}$	$\frac{3}{2}$
(8) 	Impossible	

Table 5: All possible epidemics on K_3 . The difference between Table 4 and Table 5 is that Table 5 will introduce the cycles, which enable the link between Node B

and Node C. Thus, Table 5 has more cases than Table 4. Again, due to the symmetry, Node B and Node C can also be the IIN and they have the same cases as when Node A is the IIN.

Compared with Table 4, Table 5 has more kinds of G_I once we allow infection to happen, through the link between Node B and Node C. The last G_I in Table 5: $K_3(8)$ is impossible because the infection will not happen between Node B and Node C when neither is infected.

Another very special G_I in K_3 is $K_3(7)$. As $K_3(7)$ shows, both Node B and Node C are infected, and the infection goes through all links in $K_3(7)$. The first instinct about $R_0(G_I)$ of $K_3(7)$ is 2, which is actually wrong. The accurate value should be $\frac{3}{2}$ which is less than 2.

This is caused by the cycle of infection because Node B may be infected by Node C instead of Node A. The same reason is true for Node C; it could be infected by Node B instead of Node A. Thus, the actual $R_0(G_I)$ of $K_3(7)$ is smaller than 2. The following Figure 1 shows all the cases that might happen:

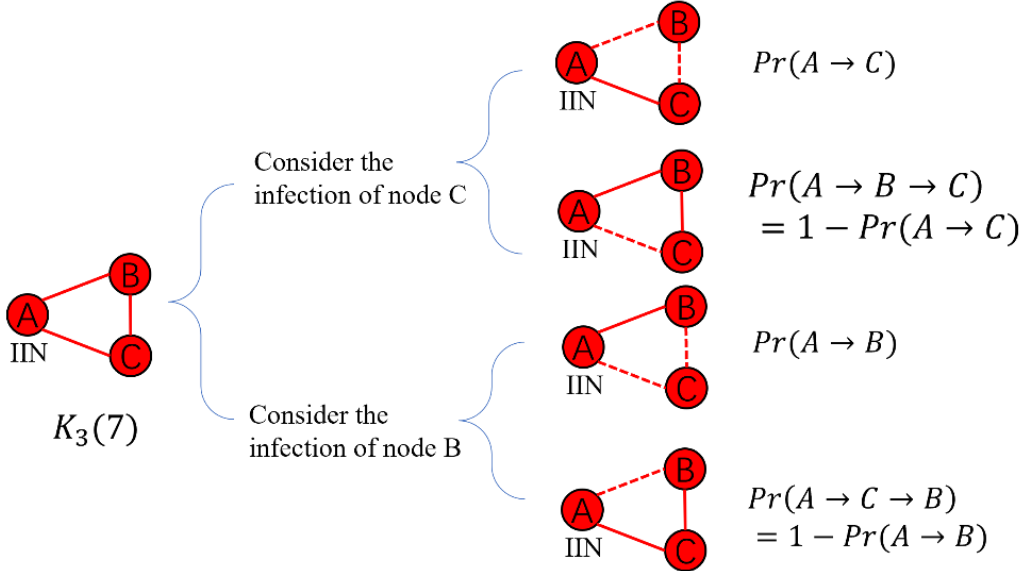


Figure 1: All possible cases of infection path in the given G_I : $K_3(7)$. The solid red lines represent the links on a faster infection path. The dashed red lines represent the links on the slower infection path. (The concept of competing paths will be given in a later section.)

Based on Figure 1, we first focus on the infection of Node C. The infection of Node C could be achieved by two paths:

Path 1: Node A \rightarrow Node C

Path 2: Node A \rightarrow Node B \rightarrow Node C

Node C is directly infected by IIN(Node A) in Path 1. Thus, only the Path 1 case could contribute to the R_0 of $K_3(7)$, while the Path 2 case cannot. We have the same for the infection of Node B. To calculate the $R_0(G_I)$ of $K_3(7)$, we need to find the four probabilities under the condition given $K_3(7)$ as the G_I :

$$Pr(A \rightarrow C | K_3(7)), Pr(A \rightarrow B \rightarrow C | K_3(7))$$

$$Pr(A \rightarrow B|K_3(7)), Pr(A \rightarrow C \rightarrow B|K_3(7))$$

The essence of the probability of occurrence of two paths is which path is faster or the entire infection process takes less time to complete. It is evident from $K_3(7)$ that Path 1 requires one infection event, and Path 2 requires two infection events. Thus, we could write the probability as:

$$Pr(A \rightarrow C|K_3(7)) = Pr(I < 2I) \quad (3.7)$$

$$Pr(A \rightarrow B \rightarrow C|K_3(7)) = Pr(I > 2I) \quad (3.8)$$

Where I represent the amount of time before one infection event happens while $2I$ represent the amount of time before two infection event happens. Thus, $Pr(I < 2I)$ is the probability that one infection event is faster than two infection events regarding the time. The derivation and result of $Pr(I < 2I)$ are given in Appendix A.2. Substitute the result to Equation (3.7) and Equation (3.8), and we have:

$$Pr(A \rightarrow C|K_3(7)) = \frac{3}{4}$$

$$Pr(A \rightarrow B \rightarrow C|K_3(7)) = \frac{1}{4}$$

Same for the infection of Node B:

$$Pr(A \rightarrow B|K_3(7)) = \frac{3}{4}$$

$$Pr(A \rightarrow C \rightarrow B|K_3(7)) = \frac{1}{4}$$

Thus, the $R_0(G_I)$ of $K_3(7)$ could be derived as:

$$R_0(G_I) = Pr(A \rightarrow C|K_3(7)) + Pr(A \rightarrow B|K_3(7)) = \frac{3}{2} \quad (3.9)$$

Based on Table 5, finally we could calculate the $R_0(G)$ of K_3 :

$$\begin{aligned} R_0(K_3) = R_0(G) &= \frac{3 \times \sum Pr(G_I) R_0(G_I)}{N} \\ &= \frac{\frac{3}{2}\beta^3 + 4\beta^2\delta + 2\beta^2\delta}{(\beta + \delta)^3} \end{aligned} \quad (3.10)$$


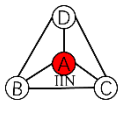
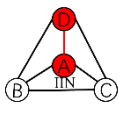
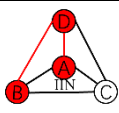
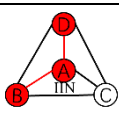
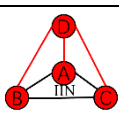
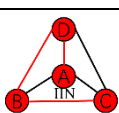
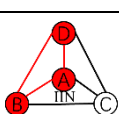
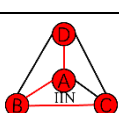
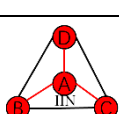
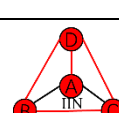
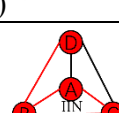
Substitute $\beta = \delta$ to Equation (3.6) and Equation(3.10) respectively:

$$R_0(K_{3(ignore\ cycles)}) = \frac{2\beta\delta + 2\beta^2}{(\beta + \delta)^2} = 1 \quad (3.11)$$

$$R_0(K_3) = \frac{\frac{3}{2}\beta^3 + 4\beta^2\delta + 2\beta^2\delta}{(\beta + \delta)^3} = \frac{15}{16} \quad (3.12)$$

The particular case $\beta = \delta$ provides us two intuitive values to help us compare with the two cases: ignore cycles and do not ignore cycles. If we consider cycles, the $R_0(K_3)$ will decrease compared to the case in which we ignore cycles. The most intuitive understanding is: the R_0 of a graph relies on the number of NNs of IIN, but cycles will make the NNs infected through other paths instead of infected directly by IIN. This is the key concept behind the definition of basic reproductive number that we proposed in this thesis. Otherwise, the R_0 of a graph G will only be the value of taking the average on the degree sequence of G .

3.1.3 Possible epidemics on K_4

Possible epidemic on K_4		
G : 		
G_I	$Pr(G_I)$	$R_0(G_I)$
(1) 	$\frac{\delta^6}{(\beta + \delta)^6}$	0
(2) 	$3 \times \frac{\beta \delta^5}{(\beta + \delta)^6}$	1
(3) 	$6 \times \frac{\beta^2 \delta^4}{(\beta + \delta)^6}$	1
(4) 	$3 \times \frac{\beta^2 \delta^4}{(\beta + \delta)^6}$	2
(5) 	$3 \times \frac{\beta^3 \delta^3}{(\beta + \delta)^6}$	1
(6) 	$6 \times \frac{\beta^3 \delta^3}{(\beta + \delta)^6}$	1
(7) 	$3 \times \frac{\beta^3 \delta^3}{(\beta + \delta)^6}$	$\frac{3}{2}$
(8) 	$6 \times \frac{\beta^3 \delta^3}{(\beta + \delta)^6}$	2
(9) 	$\frac{\beta^3 \delta^3}{(\beta + \delta)^6}$	3
(10) 	$3 \times \frac{\beta^4 \delta^2}{(\beta + \delta)^6}$	1
(11) 	$6 \times \frac{\beta^4 \delta^2}{(\beta + \delta)^6}$	$\frac{3}{2}$

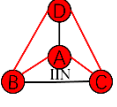
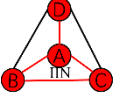
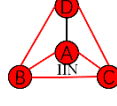
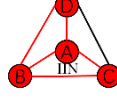
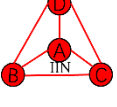
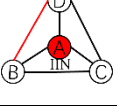
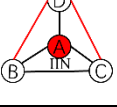
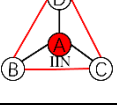
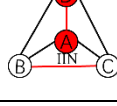
(12) 	$3 \times \frac{\beta^4 \delta^2}{(\beta + \delta)^6}$	$\frac{7}{4}$
(13) 	$3 \times \frac{\beta^4 \delta^2}{(\beta + \delta)^6}$	$\frac{5}{2}$
(14) 	$6 \times \frac{\beta^5 \delta}{(\beta + \delta)^6}$	Around 1.45 (By simulation)
(15) 	$3 \times \frac{\beta^5 \delta}{(\beta + \delta)^6}$	Around 2.07 (By simulation)
(16) 	$\frac{\beta^6}{(\beta + \delta)^6}$	Around 1.83 (By simulation)
(17) 	Impossible	
(18) 		
(19) 		
(20) 		

Table 6: All possible epidemics on K_4 . Due to the symmetry of K_4 , Table 6 only shows the cases when Node A is the IIN. Node B, C, D can also have the same cases as Node A when they are the IIN. For some case in Table 6, due to the symmetry of the case, they can be further expanded as Table 7 shows.

Due to the symmetry of K_4 , most of $Pr(G_I)$ in K_4 has coefficients except for: $K_4(1)$, $K_4(9)$, $K_4(16)$. Use $K_4(2)$ as an example to illustrate that symmetry results in a factor of 3:

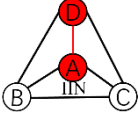
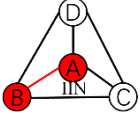
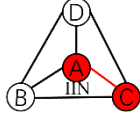
Possible G_I in $K_4(2)$		
		

Table 7: Possible G_I in $K_4(2)$. When Node A is the IIN and only one susceptible node will be infected, Node B, C, D has the same probability to be the target.

As Table 7 shows, the IIN could choose three directions, thus $K_4(2)$ actually has three

possible G_I . All three cases share the same probability $Pr(G_I) = \frac{\beta\delta^5}{(\beta+\delta)^6}$. Thus, the probability

of $K_4(2)$ is $3 \times \frac{\beta\delta^5}{(\beta+\delta)^6}$.

The $R_0(G_I)$ of most G_I in K_4 is easy to derive, but it is difficult to calculate the R_0 of $K_4(14)$, $K_4(15)$, and $K_4(16)$. This is because, for each neighboring node of IIN in these three graphs, there exist more than two paths for IIN to infect the NN if we consider cycles. Even more, some paths are dependent and share some common links. So far, since we are only able to theoretically calculate the probability of infection with only one competing path exists, we only get the numerical value of $R_0(G_I)$ for $K_4(14)$, $K_4(15)$, and $K_4(16)$ by simulation.

Based on Table 6, we can get the $R_0(K_4)$ as:

$$\begin{aligned}
 R_0(K_4) &= 3 \times \frac{\beta\delta^5}{(\beta+\delta)^6} + 6 \times \frac{\beta^2\delta^4}{(\beta+\delta)^6} + 3 \times 2 \times \frac{\beta^2\delta^4}{(\beta+\delta)^6} + 3 \times \frac{\beta^3\delta^3}{(\beta+\delta)^6} \\
 &\quad + 6 \times \frac{\beta^3\delta^3}{(\beta+\delta)^6} + 3 \times \frac{3}{2} \times \frac{\beta^3\delta^3}{(\beta+\delta)^6} + 6 \times 2 \times \frac{\beta^3\delta^3}{(\beta+\delta)^6} + 3 \times \frac{\beta^3\delta^3}{(\beta+\delta)^6} \\
 &\quad + 3 \times \frac{\beta^4\delta^2}{(\beta+\delta)^6} + 6 \times \frac{3}{2} \times \frac{\beta^4\delta^2}{(\beta+\delta)^6} + 3 \times \frac{7}{4} \times \frac{\beta^4\delta^2}{(\beta+\delta)^6} \\
 &\quad + 3 \times \frac{5}{2} \times \frac{\beta^4\delta^2}{(\beta+\delta)^6} + 6 \times 1.45 \times \frac{\beta^5\delta}{(\beta+\delta)^6} + 3 \times 2.07 \times \frac{\beta^5\delta}{(\beta+\delta)^6} \\
 &\quad + 1.83 \times \frac{\beta^6}{(\beta+\delta)^6} \\
 &= \frac{3\beta\delta^5 + 12\beta^2\delta^4 + \frac{57}{2}\beta^3\delta^3 + \frac{99}{4}\beta^4\delta^2 + 14.91\beta^5\delta + 1.83\beta^6}{(\beta+\delta)^6}
 \end{aligned}$$

3.1.4 Disadvantage of holistic method

In the above Section 3.1.1-3.1.4, we introduced the first approach using the holistic method to calculate the R_0 of the given network. However, the shortcomings of the first method are apparent:

i. Difficult to find G_I

The holistic method needs to define the status of every link in the network. As Table 5 and Table 6 show, there are 8 and 20 kinds of G_I for K_3 and K_4 , respectively, under the condition that there are only 3 and 6 links in K_3 and K_4 . If the network has more nodes and links, the complexity and the number of G_I will grow quickly. Besides, as G_I from Table 5 (8) and Table 6 (17)-(20) show, the number of impossible cases will also rapidly increase when the network becomes complex. Thus, it is difficult to find every possible G_I .

ii. Difficult to calculate $Pr(G_I)$

In the above example of K_3 and K_4 , the active and inactive links, together with the symmetry property, determine the $Pr(G_I)$ for a given G_I . However, we need to remember that the symmetry property can only be used in simple and symmetric networks such as K_3 and K_4 . Once we encounter an asymmetric network, the calculation of $Pr(G_I)$ will be much more difficult than what we have done in K_3 and K_4 .

iii. Difficult to determined $R_0(G_I)$

In the holistic method, we need first to calculate the R_0 of nodes and then derive the R_0 of the network by adding them up and taking the average. The range of R_0 of nodes is wide, ranging from 0 to the maximum number of the NNs of that chosen IIN, as Table 5 and Table 6 indicate. Apart from that, the actual value of $R_0(G_I)$ is determined by the status of links of G_I . Facing the complexity of finding G_I and $Pr(G_I)$, we cannot derive an elegant mathematical expression to calculate the R_0 of nodes.

iv. Difficult to make simplification

From the above shortcomings of i.-iii. , we can conclude that, for a large and complex network, it is very difficult to derive and calculate an analytical solution facing the complexity of both $Pr(G_I)$ and $R_0(G_I)$ if we want to consider the different infection processes in every link. Thus, we need to make some approximations and simplifications to reduce the complexity without introducing significant bias in the result. For example, we can ignore some links and not consider them when calculating $Pr(G_I)$ and $R_0(G_I)$. However, the holistic approach apparently is unsuitable if we want to make approximations because we cannot estimate and measure the significance of links. The global nature of the holistic method makes it difficult to simplify locally.

3.2 Second approach: IIN-TN method

In contrast with the holistic method, the IIN-TN (initial infected node to target node) method will not consider all the NNS of IIN once as a whole. Instead, IIN-TN will focus on the individual relation between IIN and its NNs. The main step of the whole process is as follow:

- i. Randomly select one node from the network as IIN. For this chosen IIN, all its neighboring nodes (NNs) can also be determined.
- ii. Select one of the NNs to become TN and calculate the probability that the IIN directly infects the susceptible TN. Repeat this process until all the NNs have been selected as IIN and all the probabilities have been calculated.
- iii. Sum up all the probabilities that we obtained in ii. , which is the R_0 of the chosen IIN.
- iv. Repeat step i. to step iii. until all the nodes in the network have been selected as IIN. The process of repeating step i. to step iii should be considered as independent process. The summation of R_0 of all the nodes in this network will become R_0 of the network.

The above steps only describe the general steps of the IIN-TN method. The following is a detailed explanation of the process using notations and formulas.

Between IIN and TN, apart from the direct path, there might exist other alternative paths, which are composed of links as well. Thus, in order to describe the infection process in these links, we also first need to find the graph with infection: G_I and the probability of it: $Pr(G_I)$, just like what we do in the holistic method. The difference is that instead of calculating the expected secondary infected cases of the graph with infection: $R_0(G_I)$, we need to calculate the probability that IIN directly infects TN under the condition of a given graph with infection: $Pr(IIN \rightarrow TN|G_I)$. The difference is that: the value of $Pr(IIN \rightarrow TN|G_I)$ in the IIN-TN method will be within the range of $(0,1]$, and the value of $R_0(G_I)$ in the holistic method will be within the range of $[0, number\ of\ NNs]$.

After finding $Pr(G_I)$ and $Pr(IIN \rightarrow TN|G_I)$, we can get the probability that IIN directly infect TN by the law of total probability:

$$Pr(IIN \rightarrow TN) = \sum Pr(G_I) Pr(IIN \rightarrow TN|G_I) \quad (3.13)$$

In this way, we guarantee that the IIN-TN method includes the consideration of the local structure between IIN and TN.

Then, we assign the next node of NNs to be TN and repeat the process until all nodes in NNs have been regarded as TN. For $TN \in NN$, $\mathbb{1}_{IIN \rightarrow TN}$ is a Bernoulli random variable and R_0 of IIN is the sum of dependent random variables. Thus, we have:

$$R_0(IIN) = E[(\mathbb{1}_{IIN \rightarrow 1}) + (\mathbb{1}_{IIN \rightarrow 2} | IIN \rightarrow 1) + \dots + (\mathbb{1}_{IIN \rightarrow dv} | IIN \rightarrow 1, 2, \dots, dv-1)]$$

We assume that each Bernoulli trials be independent, then we can derive R_0 of IIN as:

$$\begin{aligned} R_0(IIN) &\cong E \left[\sum_{TN \in NN} \mathbb{1}_{IIN \rightarrow TN} \right] \\ &\cong \sum_{TN \in NN} E[\mathbb{1}_{IIN \rightarrow TN}] \end{aligned}$$

$$\cong \sum_{TN \in NN} Pr(IIN \rightarrow TN) \quad (3.14)$$

Where the expectation of the independent Bernoulli random variable $E[\mathbb{1}_{IIN \rightarrow TN}]$ is the probability of success $Pr(IIN \rightarrow TN)$.

At last, sum all the R_0 of the IIN and divide the number of nodes in the network to derive the R_0 of the network:

$$R_0(Network) = \frac{\sum_{IIN \in Network} R_0(IIN)}{N} \quad (3.15)$$

Equations 3.13-3.15 demonstrate the key process of the IIN-TN method. Compared with the holistic method, in the IIN-TN method, we further divide the process of finding $R_0(IIN)$ into smaller parts. This change gives us the benefit that we can ignore the rest of the links in the graph and only focus on the links between IIN and TN. Furthermore, we can also reduce the complexity of calculating $Pr(G_I)$ and $Pr(IIN \rightarrow TN|G_I)$ by making a simplification on the alternative paths between IIN and TN. The simplification is done with the help of the structure of $K_{2,P_A} + e$, where we have an analytical solution. A more detailed introduction and explanation will be given in the following sections.

3.2.1 Introduction of $K_{2,P_A} + e$ and revisit of K_3

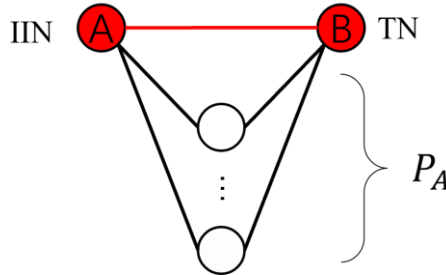


Figure 2: The graph and structure of $K_{2,P_A} + e$. The 2 nodes are Node A and Node B. The P_A nodes are the rest susceptible nodes that form P_A alternative paths. Link e is the link between Node A and Node B. The red color of Node A represents that Node A is the IIN, and the red color of Node B represents that Node B is the target node that must be infected. The red color of link e represents that this link will surely spread the epidemic, while the black color of the rest alternative paths indicates that all the paths have the chance to compete with link e .

The $K_{2,P_A} + e$ is the complete bipartite graph on two and P_A nodes plus the link e , which connects IIN(initial infected node) and TN (target node) directly. As Figure 2 shows, between IIN and TN, there are a total number of P_A alternative paths. Alternative path means the path provides another option for IIN to reach TN instead of the direct path (link e) between IIN and TN. The analytical solution of Figure 2 will change accordingly with the change in the number of alternative paths P_A . In order to derive the analytical solution, we need to derive two things based on $K_{2,P_A} + e$:

- i. $Pr(G_I)$: The probability of graph with infection process.
- ii. $Pr(IIN \rightarrow TN|G_I)$: Conditional probability of IIN infect TN directly under the condition of given G_I

An example G_I under the structure of Figure 2 is given below:

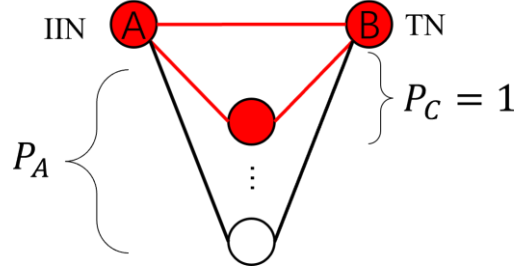


Figure 3: A possible case of $K_{2,P_A} + e$. Except from the link e , there is one competing path (which originally is an alternative path) that will compete with link e for the shortest time to infect TN. The number of P_C and P_A will determine G_I for each case.

If we define the links that spread the infection as active links, compared with Figure 2 and Figure 3, Figure 3 has two more active links. The two links form a competing path, which is denoted as $P_C = 1$. The commons and differences between an alternative path and a competing path are given below:

- i. The number of alternative paths P_A is determined and fixed once we have a graph structure like Figure 2. But the number of competing paths P_C might vary just like Figure 3 shows. The maximum number of possible competing paths is the number of alternative paths: $\max(P_C) = P_A$.
- ii. The number of links in one alternative path is denoted as L_A . If all the links in the alternative path are active, the alternative path will become a competing path. This means that if the number of active links in one alternative path is less than L_A , this alternative path should not be considered a competing path.
- iii. The existence of competing paths will lower the probability that IIN directly infects TN ($Pr(IIN \rightarrow TN|G_I)$) because the infection from IIN to TN could be reached by any competing paths or the direct path. Thus, $Pr(IIN \rightarrow TN|G_I)$ is actually determined by P_C and L_A , but irrelevant with P_A . However, $Pr(G_I)$ is determined by all of them: P_A , P_C , and L_A .

A simple example with $P_A = 1$ will demonstrate how this work as the following table shows ($K_{2,1} + e$ is same as K_3):

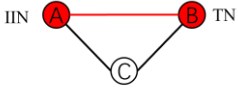
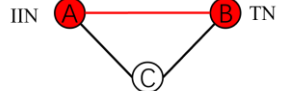
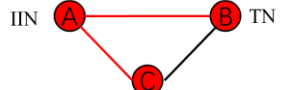
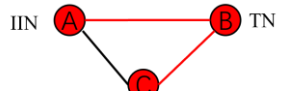
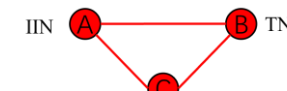
$G: K_{2,1} + e = K_3$  $P_A = 1$ $L_A = 2$		
G_I	$Pr(G_I)$	$Pr(IIN \rightarrow TN G_I)$
 $L_I = 0$ $P_C = 0$ (1)	$\frac{\beta}{\beta + \delta} \frac{\beta^0 \delta^2}{(\beta + \delta)^2}$	1
 $L_I = 1$ $P_C = 0$ (2)	$\frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^1}{(\beta + \delta)^2}$	1
 $L_I = 1$ $P_C = 0$ (3)	$\frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^1}{(\beta + \delta)^2}$	1
 $L_I = 2$ $P_C = 1$ (4)	$\frac{\beta}{\beta + \delta} \frac{\beta^2 \delta^0}{(\beta + \delta)^2}$	$\frac{3}{4}$

Table 8: All possible epidemics on $K_{2,1} + e (K_3)$. For this case, $P_A = 1$, and $P_C = [0,1]$.

Based on Table 8, the probability that IIN directly infects TN could be calculated as:

$$Pr(IIN \rightarrow TN) = Pr(A \rightarrow B) = \sum Pr(G_I) Pr(IIN \rightarrow TN|G_I) = \frac{\beta \delta^2 + 2\beta^2 \delta + \frac{3}{4}\beta^3}{(\beta + \delta)^3} \quad (3.16)$$

In fact, Table 8 and Equation 3.16 only calculate the case when node A is the IIN and node B is TN. However, when node A is IIN, node C should also be considered as TN apart from node B. By the symmetry of K_3 , the probability that node A directly infects node B should be the same as the probability that node A directly infects node C:

$$Pr(A \rightarrow B) = Pr(A \rightarrow C) \quad (3.17)$$

Thus, the R_0 of node A when A is selected as IIN could be calculate as:

$$\begin{aligned}
R_0(IIN = A) &= \sum_{TN \in NNs} Pr(IIN \rightarrow TN) \\
&= Pr(A \rightarrow B) + Pr(A \rightarrow C) \\
&= 2 \times \frac{\beta\delta^2 + 2\beta^2\delta + \frac{3}{4}\beta^3}{(\beta + \delta)^3} \\
&= \frac{\frac{3}{2}\beta^3 + 4\beta^2\delta + 2\beta^2\delta}{(\beta + \delta)^3} \tag{3.18}
\end{aligned}$$

IIN could be selected as node B and node C as well, and due to the symmetry of K_3 , the value should be the same when IIN is either A, B or C: $R_0(IIN = A) = R_0(IIN = B) = R_0(IIN = C)$. At last, the R_0 of K_3 could be calculated as:

$$\begin{aligned}
R_0(K_3) &= \frac{\sum_{IIN \in Network} R_0(IIN)}{N} \\
&= \frac{R_0(A) + R_0(B) + R_0(C)}{3} \\
&= \frac{\frac{3}{2}\beta^3 + 4\beta^2\delta + 2\beta^2\delta}{(\beta + \delta)^3} \tag{3.19}
\end{aligned}$$

Comparing the results from Equation (3.19) by the IIN-TN method and from Equation (3.10) by the holistic method, we can find that the two methods deliver the same result. The consistency of the results between the two methods supports our choice of using the less complex IIN-TN method. Furthermore, comparing Table 5 and Table 8, we can find that the complexity of the IIN-TN method is only half of the holistic method regarding the case of K_3 .

From the above example, one essential concept should be pointed out and emphasized: the $K_{2,P_A} + e$ could help us to calculate the probability of the direct infection between any IIN and any TN ($Pr(IIN \rightarrow TN)$) for any number of independent alternative paths (2 links only) between IIN and TN.

As the above example shows, when $P_A = 1$, $K_{2,P_A} + e$ becomes K_3 . The exact match of result of $K_{2,1} + e$ and K_3 is a coincidence because K_3 is one of the simplest structures and there is no alternative path in K_3 that is longer than 2 links. Actually, $K_{2,P_A} + e$ could be used as an approximation and simplification method to solve $Pr(IIN \rightarrow TN)$ by only considering the alternative path with 2 links. Another example of K_4 will be shown in the next section after the analytical solution of $K_{2,P_A} + e$ is given. In which, we will find that some links are ignored by using $K_{2,P_A} + e$ as a simplification method and how much error it will introduce.

3.2.2 Analytical solution of $K_{2,P_A} + e$ and revisit of K_4

In $K_{2,P_A} + e$, P_A (number of alternative paths) is a variable that has a range of $[1, \infty)$, and it determines the basic structure of $K_{2,P_A} + e$ once P_A is determined. Therefore, there are two variables based on P_A , such as the range of P_C and L_I : $P_C = [0, P_A]$ and $L_I = [0, 2P_A]$.

First, we define L_I as the total number of infected links in all alternative paths. Thus, L_I determines the power of β just as $Pr(G_I)$ in Table 8 shows. Besides, as we can see from the values of $Pr(G_I)$ in Table 8, they are composed of two fractions. The first fraction is a fixed value of $\frac{\beta}{\beta+\delta}$, which indicates the direct path e that connects IIN and TN. The second fraction will change in the numerator according to the total number of infected links L_I , but has a fixed value at the denominator: $\frac{\beta^{L_I} \delta^{2P_A - L_I}}{(\beta+\delta)^{2P_A}}$. Where $2P_A$ is the total number of links of all alternative paths, L_I is the number of active links, and $2P_A - L_I$ is the number of remaining inactive links.

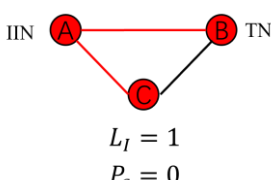
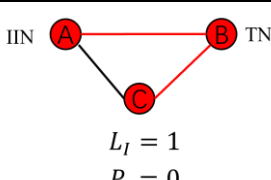
G_I	$Pr(G_I)$	$Pr(IIN \rightarrow TN G_I)$
 <p>(2)</p>	$\frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^1}{(\beta + \delta)^2}$	1
 <p>(3)</p>	$\frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^1}{(\beta + \delta)^2}$	1

Table 9: Part of Table 8. $Pr(G_I)$ and $Pr(IIN \rightarrow TN|G_I)$ are the same for both cases even the active links are different. This is because the two cases have the same L_I and P_C .

Besides, as we can see from Table 9 above, we can find that for the two cases in K_3 , they have the same $Pr(G_I)$ and $Pr(IIN \rightarrow TN|G_I)$ because the $L_I = 1$ and $P_C = 0$ are the same for both cases. The difference is the placement of the active links. In fact, for $K_{2,P_A} + e$ with a large number of P_A , the kinds of placements of active links with a fixed P_C is a large number, which could be calculated as the following function shows:

$$|\mathcal{F}(P_A, P_C, L_I)| = \begin{cases} \binom{P_A}{P_C} \frac{\prod_{k=0}^{L_I - 2P_C - 1} \binom{2P_A - 2P_C - 2k}{1}}{P(L_I - 2P_C, L_I - 2P_C)}, & \text{if } L_I - 2P_C - 1 \geq 0 \\ \binom{P_A}{P_C}, & \text{if } L_I - 2P_C - 1 < 0 \end{cases} \quad (3.20)$$

And $P_A = [1, \infty)$, $P_C = [0, P_A]$, $L_I = [0, 2P_A]$ for Equation 3.20.

Where $|\mathcal{P}(P_A, P_C, L_I)|$ is the product of two terms: $\binom{P_A}{P_C}$ and $\frac{\prod_{k=0}^{L_I-2P_C-1} \binom{2P_A-2P_C-2k}{1}}{P(L_I-2P_C, L_I-2P_C)}$.

$\binom{P_A}{P_C}$ is the binomial coefficient which could be calculated as:

$$\binom{P_A}{P_C} = \frac{P_A!}{P_C! (P_A - P_C)!} \quad (3.21)$$

The meaning of $\binom{P_A}{P_C}$ is to select P_C competing paths from the total number of P_A alternative paths.

Then, what's the meaning of $\frac{\prod_{k=0}^{L_I-2P_C-1} \binom{2P_A-2P_C-2k}{1}}{P(L_I-2P_C, L_I-2P_C)}$? The total number of links of all alternative paths in $K_{2,P_A} + e$ is $2P_A$. Because all the alternative paths in $K_{2,P_A} + e$ are composed of 2 links, one competing path will occupy two links, thus the number of remaining undefined links will be $2P_A - 2P_C$. Then, from $2P_A - 2P_C$ undefined links, we select one link as the active link, which could be represented by the binomial coefficients $\binom{2P_A-2P_C}{1}$. After this selection, the number of remaining undefined links will reduce to $2P_A - 2P_C - 2$ because we cannot select the link on the same alternative path as the $L_A = 2$ (L_A is the number of links in one alternative path). If two links in the same alternative path both become active, then this alternative path will become a competing path, which will lead to an increase of P_C and conflict with our precondition of fixed P_C . Then, based on $2P_A - 2P_C - 2$ undefined links, we can select another link to be active, which is $\binom{2P_A-2P_C-2}{1}$. Repeat the process until the number of active links reaches our precondition, the fixed L_I . The continued selection process could be represented by the product of each selection: $\prod_{k=0}^{L_I-2P_C-1} \binom{2P_A-2P_C-2k}{1}$.

However, we should not ignore the errors introduced during the selection process: one combination of selections will be overcounted many times by the order of several independent selection processes. For example, one combination of {link A, link B, link C} might be overcounted as (link A, link B, link C), (link A, link C, link B), (link B, link A, link C), (link B, link C, link A), (link C, link A, link B), (link C, link B, link A) for six permutations. Thus, we need to divide the selection result by the permutation number: $P(L_I - 2P_C, L_I - 2P_C)$, which could be calculated as:

$$P(L_I - 2P_C, L_I - 2P_C) = \frac{(L_I - 2P_C)!}{[(L_I - 2P_C) - (L_I - 2P_C)]!} = (L_I - 2P_C)! \quad (3.22)$$

because in total we need to select $L_I - 2P_C$ links one by one, thus we have $L_I - 2P_C$ independent selection process.

Equation 3.22 only gives the number of kinds of G_I for the fixed L_I and P_C . However, this is insufficient because both L_I and P_C will change if we want to derive an analytical solution of $K_{2,P_A} + e$. Based on Equation 3.22, we take the further step by considering the change of L_I , but with fixed P_C . When P_C is a fixed number, the range of L_I is determined as $L_I = [2P_C, P_A + P_C]$. The reason is simple: each competing path has two links, which means there must be at least $2P_C$ active links to form P_C competing path. The higher bound is limited by $P_A + P_C$, because if we have one more active link, it will form $P_C + 1$ competing

path, which conflicts with our predefinition of fixed P_C . Based on the range of L_I , we could write the following equation as the sum of a series of Equation 3.22, where $|\mathcal{F}(\text{number})|$ is the total number of family of placements:

$$|\mathcal{F}(\text{number})| = \sum_{L_I=2P_C}^{P_A+P_C} |\mathcal{F}(L_I, P_A, P_C)| \quad (3.23)$$

Furthermore, based on Equation 3.23, we could calculate the probability of G_I as:

$$Pr(G_I) = \frac{\beta}{\beta + \delta} \sum_{L_I=2P_C}^{P_A+P_C} \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} \right) \quad (3.24)$$

Equation 3.24 is achieved based on the precondition that P_C is a fixed value. P_C not only determines $Pr(G_I)$ as Equation 3.24 shows, but it also determines the value of $Pr(IIN \rightarrow TN|G_I)$. In $K_{2,P_A} + e$, with fixed P_C in G_I , $Pr(IIN \rightarrow TN|G_I)$ refers to the probability that the infection goes through the path faster than any competing path, including the fastest alternative path. Thus, $Pr(IIN \rightarrow TN|G_I)$ could be rewrite as:

$$Pr(IIN \rightarrow TN|G_I) = Pr(I < \min(P_C(I + I))) \quad (3.25)$$

Again, we could calculate $Pr(I < \min(P_C(I + I)))$ by the integration of the product of conditional probability:

$$\begin{aligned} Pr(I < \min(P_C(I + I))) &= \int_{t=-\infty}^{\infty} Pr(\min(P_C(I + I)) > I | I = t) f_I(i) dt \\ &= \int_{t=0}^{\infty} f_I(t) \left(1 - F_{\min(P_C(I+I))}(t) \right) dt \\ &= \int_{t=0}^{\infty} \beta e^{-\beta t} (e^{-\beta t} + \beta t e^{-\beta t})^{P_C} dt \\ &= e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) \end{aligned} \quad (3.26)$$

(The detail derivation process of Equation (3.26) is given in Appendix A.4.)

Based on Equation 3.23-3.26, we could write the result of probability that IIN direct infect TN in the graph $K_{2,P_A} + e$ with P_C range of $[0, P_A]$ as:

$$\begin{aligned} Pr(IIN \rightarrow TN) &= \sum_{P_C=0}^{P_A} Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\ &= \sum_{P_C=0}^{P_A} \left\{ \left[\frac{\beta}{\beta + \delta} \sum_{L_I=2P_C}^{P_A+P_C} \left(\mathcal{F}(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} \right) \right] \left(e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) \right) \right\} \end{aligned} \quad (3.27)$$

$$\text{Where } |\mathcal{F}(P_A, P_C, L_I)| = \begin{cases} \binom{P_A}{P_C} \frac{\prod_{k=0}^{L_I-2P_C-1} (2P_A-2P_C-2k)}{(L_I-2P_C)!}, & \text{if } L_I - 2P_C - 1 \geq 0 \\ \binom{P_A}{P_C}, & \text{if } L_I - 2P_C - 1 < 0 \end{cases}$$

The above Equation 3.27 may seem very complicated, but it is straightforward if we calculate step by step. Another example of $K_{2,P_A} + e$ with $P_A = 2$, as Figure 4 shows, will be given to illustrate the inside of Equation 3.27.

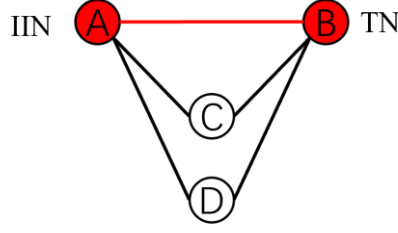


Figure 4: $K_{2,P_A} + e$ with $P_A = 2$. This structure has two alternative paths.

Starting with the outermost part of the formula, we have the summation of the multiple terms with a range of $P_C = [0, P_A]$. In this example, $P_A = 2$, thus we have the range of $P_C = [0, 2]$, which lead to 3 cases: $P_C = 0$; $P_C = 1$; $P_C = 2$.

When $P_C = 0$, the range of L_I could be derived : $L_I = [2P_C, P_A + P_C] = [0, 2]$. Thus, L_I also has three cases: $L_I = 0$; $L_I = 1$; $L_I = 2$. Then, we can find the kinds of G_I and calculate the number for each kinds using $|\mathcal{F}(P_A, P_C, L_I)|$:

$$\begin{aligned}
 |\mathcal{F}(P_A, P_C, L_I)| &= \binom{P_A}{P_C} \frac{\prod_{k=0}^{L_I-2P_C-1} \binom{2P_A-2P_C-2k}{1}}{P(L_I-2P_C, L_I-2P_C)} \\
 &= \begin{cases} \binom{2}{0} & \text{when } L_I = 0 \\ \binom{2}{0} \frac{\binom{4}{1}}{P(1,1)} & \text{when } L_I = 1 \\ \binom{2}{0} \frac{\binom{4}{1}\binom{2}{1}}{P(2,2)} & \text{when } L_I = 2 \end{cases} \\
 &= \begin{cases} 1 & \text{when } L_I = 0 \\ 4 & \text{when } L_I = 1 \\ 4 & \text{when } L_I = 2 \end{cases} \quad (3.28)
 \end{aligned}$$

Besides, the probability $Pr(G_I)$ for each case (not considering the number of each kind):

$$Pr(G_I) = \frac{\beta}{\beta + \delta} \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} = \begin{cases} \frac{\beta}{\beta + \delta} \frac{\beta^0 \delta^4}{(\beta + \delta)^4} & \text{when } L_I = 0 \\ \frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^3}{(\beta + \delta)^4} & \text{when } L_I = 1 \\ \frac{\beta}{\beta + \delta} \frac{\beta^2 \delta^2}{(\beta + \delta)^4} & \text{when } L_I = 2 \end{cases} \quad (3.29)$$

Equation 3.28 shows 3 kinds of G_I when $P_A = 2$ and $P_C = 0$, and each kind has 1, 4, and 4 possible epidemics, respectively. The following Table 10 summarizes the content of Equation 3.28 and Equation 3.29 and gives the schematic graph of possible epidemics:

L_I	G_I	$Pr(G_I)$
$L_I = 0$		$\frac{\beta}{\beta + \delta} \frac{\beta^0 \delta^4}{(\beta + \delta)^4}$
$L_I = 1$		$4 \times \frac{\beta}{\beta + \delta} \frac{\beta^1 \delta^3}{(\beta + \delta)^4}$
$L_I = 2$		$4 \times \frac{\beta}{\beta + \delta} \frac{\beta^2 \delta^2}{(\beta + \delta)^4}$

Table 10: Possible epidemics for $K_{2,2} + e$ when $P_A = 2$ and $P_C = 0$. L_I has range $[0,2]$.

Furthermore, we need to calculate $Pr(IIN \rightarrow TN|G_I)$ when $P_C = 0$:

$$Pr(IIN \rightarrow TN|G_I) = e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) = 1 \quad (3.30)$$

At last, based on Equation 3.28-3.30 we could calculate $Pr(IIN \rightarrow TN)$ when $P_C = 0$:

$$\begin{aligned}
Pr(IIN \rightarrow TN) &= Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\
&= \frac{\beta}{\beta + \delta} \left[\sum_{L_I=2P_C}^{P_A+P_C} \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} \right) \right] (e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1)) \\
&= \frac{\beta}{\beta + \delta} \left[\sum_{L_I=0}^2 \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^4} \right) \right] \times 1 \\
&= \frac{\beta}{\beta + \delta} \left(\frac{\beta^0 \delta^4}{(\beta + \delta)^4} + \frac{4\beta^1 \delta^3}{(\beta + \delta)^4} + \frac{4\beta^2 \delta^2}{(\beta + \delta)^4} \right) \quad (3.31)
\end{aligned}$$

After dealing with the case when $P_C = 0$, when $P_C = 1$ and $P_C = 2$, we repeat the process of Equation 3.28 and 3.29, and we can write another 2 tables similar to Table 10:

L_I	G_I	$Pr(G_I)$
$L_I = 2$		$2 \times \frac{\beta}{\beta + \delta} \frac{\beta^2 \delta^2}{(\beta + \delta)^4}$
$L_I = 3$		$4 \times \frac{\beta}{\beta + \delta} \frac{\beta^3 \delta^1}{(\beta + \delta)^4}$

Table 11: Possible epidemics for $K_{2,2} + e$ when $P_A = 2$ and $P_C = 1$. L_I has range: $[2, 3]$.

L_I	G_I	$Pr(G_I)$
$L_I = 4$		$\frac{\beta}{\beta + \delta} \frac{\beta^4 \delta^0}{(\beta + \delta)^4}$

Table 12: Possible epidemics for $K_{2,2} + e$ when $P_A = 2$ and $P_C = 2$. L_I could only be 4 for this case.

Again, we can calculate $Pr(IIN \rightarrow TN|G_I)$ for the two cases of Table 11 and Table 12:

$$\begin{aligned}
 Pr(IIN \rightarrow TN|G_I) &= e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) \\
 &= \begin{cases} \frac{3}{4}, & \text{when } P_C = 1 \\ \frac{17}{27}, & \text{when } P_C = 2 \end{cases} \quad (3.32)
 \end{aligned}$$

And the $Pr(IIN \rightarrow TN)$:

$$\begin{aligned}
 Pr(IIN \rightarrow TN) &= Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\
 &= \frac{\beta}{\beta + \delta} \left[\sum_{L_I=2P_C}^{P_A+P_C} \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} \right) \right] (e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1)) \\
 &= \begin{cases} \frac{\beta}{\beta + \delta} \left[\sum_{L_I=2}^3 \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^4} \right) \right] \times \frac{3}{4}, & \text{when } P_C = 1 \\ \frac{\beta}{\beta + \delta} \left[\sum_{L_I=4}^4 \left(f(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^4} \right) \right] \times \frac{17}{27}, & \text{when } P_C = 2 \end{cases}
 \end{aligned}$$

$$= \begin{cases} \frac{\beta}{\beta + \delta} \left(\frac{2\beta^2\delta^2}{(\beta + \delta)^4} + \frac{4\beta^3\delta^1}{(\beta + \delta)^4} \right) \times \frac{3}{4}, & \text{when } P_C = 1 \\ \frac{\beta}{\beta + \delta} \left(\frac{\beta^4\delta^0}{(\beta + \delta)^4} \right) \times \frac{17}{27}, & \text{when } P_C = 2 \end{cases} \quad (3.33)$$

In Equation 3.31 and Equation 3.33, we have derived the terms for different values of P_C . According to Equation 3.27, the last step is to sum up all the terms:

$$\begin{aligned} Pr(IIN \rightarrow TN) &= \sum_{P_C=0}^{P_A} Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\ &= \sum_{P_C=0}^2 Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\ &= \frac{\beta}{\beta + \delta} \left(\frac{\beta^0\delta^4}{(\beta + \delta)^4} + \frac{4\beta^1\delta^3}{(\beta + \delta)^4} + \frac{4\beta^2\delta^2}{(\beta + \delta)^4} \right. \\ &\quad \left. + \left(\frac{2\beta^2\delta^2}{(\beta + \delta)^4} + \frac{4\beta^3\delta^1}{(\beta + \delta)^4} \right) \times \frac{3}{4} + \left(\frac{\beta^4\delta^0}{(\beta + \delta)^4} \right) \times \frac{17}{27} \right) \\ &= \frac{\beta^1\delta^4 + 4\beta^2\delta^3 + \frac{11}{2}\beta^3\delta^2 + 3\beta^4\delta^1 + \frac{17}{27}\beta^5\delta^0}{(\beta + \delta)^5} \end{aligned} \quad (3.34)$$

The result in Equation 3.34 is the analytical result of $K_{2,2} + e$. For $K_{2,P_A} + e$ with a high P_A value, we can still follow the same process and get the analytical solution. With the analytical result of $K_{2,2} + e$, we can regard it as the simplification structure of K_4 as Figure 5 shows.

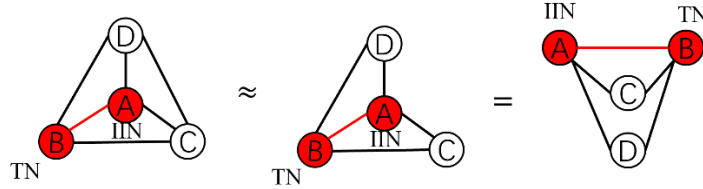


Figure 5: K_4 could be simplified as $K_{2,2} + e$ by ignoring one links. For this case, the link between Node D and Node C has been ignored. Based on this simplification, the analytical solution of $K_{2,2} + e$ could be used as the approximation result of K_4 .

Thus, the $Pr(IIN \rightarrow TN)$ of K_4 could be regarded as the analytical solution we have derived for $K_{2,2} + e$, which is shown in Equation 3.34. Thus, based on Equation 3.34, we could calculate the R_0 of node A by regarding node B, node C, and node D as TN in turn. By symmetry of K_4 , the $Pr(IIN \rightarrow TN)$ between each pair of nodes is the same and we could write R_0 of node A:

$$\begin{aligned} R_0(IIN = A) &= \sum_{TN \in NNS} Pr(IIN \rightarrow TN) \\ &= Pr(A \rightarrow B) + Pr(A \rightarrow C) + Pr(A \rightarrow D) \\ &= 3Pr(A \rightarrow B) \end{aligned}$$

$$= 3 \times \frac{\beta^1 \delta^4 + 4\beta^2 \delta^3 + \frac{11}{2}\beta^3 \delta^2 + 3\beta^4 \delta^1 + \frac{17}{27}\beta^5 \delta^0}{(\beta + \delta)^5} \quad (3.35)$$

Apart from the case when node A is IIN, the rest of the nodes in K_4 could also be regarded as IIN, and we can get the same result of R_0 as the one given in Equation 3.35 due to the symmetric property of K_4 . Thus, we could write R_0 of K_4 as:

$$\begin{aligned} R_0(K_4) &= \frac{\sum_{IIN \in Network} R_0(IIN)}{N} \\ &= \frac{R_0(A) + R_0(B) + R_0(C) + R_0(D)}{4} \\ &= \frac{4R_0(A)}{4} \\ &= 3 \times \frac{\beta^1 \delta^4 + 4\beta^2 \delta^3 + \frac{11}{2}\beta^3 \delta^2 + 3\beta^4 \delta^1 + \frac{17}{27}\beta^5 \delta^0}{(\beta + \delta)^5} \end{aligned} \quad (3.36)$$

So far, we have two results of $R_0(K_4)$, one from the holistic method and one from the IIN-TN method. We can compare them by substituting $\beta = \delta = 1$ to both analytical results:

$$\begin{aligned} R_0(K_4) &= \begin{cases} \frac{3\beta\delta^5 + 12\beta^2\delta^4 + \frac{57}{2}\beta^3\delta^3 + \frac{99}{4}\beta^4\delta^2 + 14.91\beta^5\delta + 1.83\beta^6}{(\beta + \delta)^6}, & \text{by Holistic method} \\ 3 \times \frac{\beta^1\delta^4 + 4\beta^2\delta^3 + \frac{11}{2}\beta^3\delta^2 + 3\beta^4\delta^1 + \frac{17}{27}\beta^5\delta^0}{(\beta + \delta)^5}, & \text{by TNN - IN method} \end{cases} \\ &= \begin{cases} 1.3840, & \text{by Holistic method} \\ 1.3246, & \text{by IIN - IN method} \end{cases} \end{aligned} \quad (3.37)$$

As the results stated in Equation 3.37 show, the two methods almost give the same result. The extremely small error proves that the IIN-TN method is correct and makes a good approximation for the K_4 . However, by theory, the result from the IIN-IN method should be bigger than the result of the Holistic method. The Holistic method used in K_4 considers every link for each case as Table 6 shows. For example, for the case (14), (15), and (16) in Table 6, since our method cannot derive analytical solutions for them, the value of $R_0(G_I)$ can only be given by simulation. As for the IIN-TN method, as Figure 5 shows, some link has been ignored when we use the $K_{2,P_A} + e$ structure. That is the reason why the IIN-TN method will give a higher result of $R_0(\text{Network})$ than the Holistic method by theory. However, the simulation result will vary for each trial, which means the simulation will not provide a fixed exact value all the time. The simulation method will be introduced in Chapter 4. Since the error is so small, in the K_4 example, we tend to believe this abnormality is introduced by the simulation error that we use in the Holistic method.

When facing a more complex network, the shortcoming of using the approximation might be more obvious because the number of alternative paths that are

longer than two links might be significant in some scenarios. In those cases, the results after the approximation will be higher than the theoretical value. This is the error introduced by the approximation and we believe this is a valuable direction worth further research in the future.

4. Verification of Graph Theory by Simulation

In Chapter 3, the calculation and derivation in graph theory have been given mathematically, but there is a lack of results from simulation to support our calculation. Thus, in this chapter, we will introduce how we use MATLAB to simulate the epidemics in the networks that appearing in Chapter 3. The code of MATLAB will be given in Appendix C.

4.1 Verification of $K_3(7)$

In Chapter 3, Section 3.1.2, the first non-integer value of $R_0(G_I)$ is introduced when we try to find the R_0 of K_3 as the following Table 13 shows:

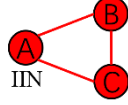
Possible epidemics on K_3	
G_I	$R_0(G_I)$
	$\frac{3}{2}$

Table 13: Part of Table 5: case (7) of Table 5. All links are active, thus the infection of Node B and Node C are the result of competing between different links. Detailed calculation process of this case was given in Figure 1 and Equation 3.7-3.9.

In Equation 3.9, we give the numerical result of above G_I as the sum of two conditional probabilities:

$$R_0(G_I) = Pr(A \rightarrow C | K_3(7)) + Pr(A \rightarrow B | K_3(7)) = \frac{3}{2} \quad (3.9)$$

For the G_I in Table 13, it has a total of 3 links. Thus, in MATLAB, we could use the following code to generate the random time that the link needs for the link to become active and spread the epidemic to the node on the other side(Full code is given in Appendix C):

```

1 infection_rate = 1;
2 curing_rate = 1;
3 effect_spr_rate = infection_rate / curing_rate;
4 link1 = exprnd(1 / infection_prob);
5 link2 = exprnd(1 / infection_prob);
6 link3 = exprnd(1 / infection_prob);

```

Table 14: Part of source code of K3_R0_simulation in Appendix C.

Where *infection_rate* in the above code is β , *curing_rate* is δ , and the *effect_spr_rate* is the effective spreading rate τ . The MATLAB command *exprnd* will generate the random

number from the exponential distribution with a mean of $\frac{1}{\tau}$. The reason for $\text{mean} = \frac{1}{\tau}$ is simple. Please recall that, in Section 2.1, we have explained that the infection process and curing process are two independent Poisson processes, and the ratio $\tau = \frac{\beta}{\delta}$ is the effective infection (spreading) rate [5] [6] based on the two Poisson processes. Thus, we could use the *exprnd* command to generate the random time.

Based on the time for each link, we could further compare the time for different paths and find out which path needs the shortest time. Suppose the path with the shortest time is the alternative path. In that case, it means in this round of the simulation, the IIN does not directly infect the neighboring node, and it cannot be counted as the R_0 of IIN; if the direct path (the link that directly connects IIN and neighboring node) has the shortest time, which means in this round of simulation, IIN directly infects the neighboring node. It should be counted as the R_0 of IIN.

It should be noted that we cannot get the results directly from one round of simulation. According to the law of large numbers (LLN), the larger the number of times that the simulation is repeated, the closer will the average result be to the expected value [7]. Thus, only after a huge round of simulation, we could get close to the real probabilities of the direct infection and the real $R_0(G_I)$ for a given G_I that we want to explore. The number of rounds of simulation for different structures in our program is set to be 100,000, which we believe is a balanced choice that guarantees acceptable errors on the one hand and would not cost too much calculation resources on the other hand. Besides, in order to make the easy comparison, we also assign the special value of $\beta = 1$, and $\delta = 1$ in simulation. The aim of doing this is to stay consistence with the theory part and in this way, we could easily find whether the simulation and theoretical derivation are matched or not. The following Table 15 shows the results of several attempts of simulation:

1	>> K3_R0_simulation
2	R0: 1.5012
3	>> K3_R0_simulation
4	R0: 1.5019
5	>> K3_R0_simulation
6	R0: 1.5009
7	>> K3_R0_simulation
8	R0: 1.5011

Table 15: 4 results in Command window (MATLAB) after running K3_R0_simulation 4 times. The results could be regarded as the same. Some variation is caused by the number generated by the *exprnd()* function.

As the results in Table 15 above show, the result of each time of simulation is different; this is because the result depends on the randomly generated number, which is reasonable to have four different results. Although the precise numbers are not perfectly $\frac{3}{2}$, but the accuracy of the simulation is very high because they are very close to $\frac{3}{2}$. Thus, the results also prove the correctness of our derivation in Equation 3.9.

4.2 Verification of three cases in K_4

In Table 6 of Section 3.1.3, we give three simulation results for the three cases of G_I : (14), (15), and (16) as the following Table 16 shows:

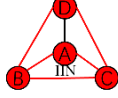
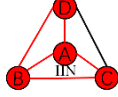
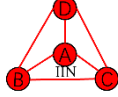
Possible epidemic on K_4	
G_I	$R_0(G_I)$
	Around 1.45 (By simulation)
	Around 2.07 (By simulation)
	Around 1.83 (By simulation)

Table 16: Part of Table 6: cases (14), (15), (16) of Table 6. The R_0 for these three cases is not derived based on calculation but on simulation by the method mentioned in Table 14. The reason is given in section 3.1.3.

Using the same method as given in Section 4.1.1, we conduct the simulation for the G_I in Table 16, and the results are presented in the following Table 17:

```

1 >> K4_14_R0_simulation
2 R0: 1.4456
3 >> K4_15_R0_simulation
4 R0: 2.0749
5 >> K4_16_R0_simulation
6 R0: 1.8338

```

Table 17: Results in Command window (MATLAB) after running simulation for (14), (15), and (16).

Due to the similarity of the code for the three cases, we only provide the full version of code of (14) in Appendix C. The code of the other two cases can be easily obtained by modifying the part about comparing different paths. The reason why we can only provide the $R_0(G_I)$ results by simulation for the cases (14), (15), and (16) in Table 6 is that we do not have an analytical solution that perfectly solves the G_I of the three cases. Currently, only the analytical solution that deals with the independent alternative paths with 2 links ($K_{2,p_A} + e$) is derived and stated in Section 3.2.2, Equation 3.27. However, an approximation result for K_4 , after using the IIN-TN method, can also be given in Section 3.2.2. The verification of the IIN-TN method will be given in the next section.

4.3 Verification of $Pr(IIN \rightarrow TN|G_I)$ in $K_{2,P_A} + e$

In section 3.2.1 , the graph and structure of $K_{2,P_A} + e$ are given in Figure 2 as below shows:

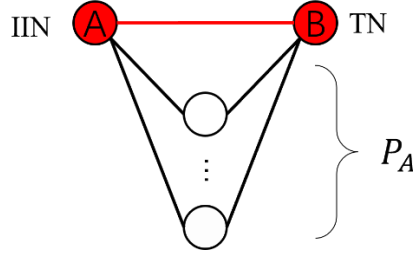


Figure 2: The graph and structure of $K_{2,P_A} + e$.

Based on this structure, we further give the mathematical expression of $Pr(IIN \rightarrow TN|G_I)$, the conditional probability under the given G_I with variable P_C (number of the competing path) in Section 3.2.2 as Equation 3.25 and Equation 3.26 states:

$$\begin{aligned} Pr(IIN \rightarrow TN|G_I) &= Pr(I < \min(P_C(I + I))) \\ &= e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) \end{aligned} \quad (3.25 \text{ \& } 3.26)$$

We also provide the full analytical solution of $K_{2,P_A} + e$ by using the IIN-TN method as Equation 3.27 states:

$$\begin{aligned} Pr(IIN \rightarrow TN) &= \sum_{P_C=0}^{P_A} Pr(G_I) Pr(IIN \rightarrow TN|G_I) \\ &= \sum_{P_C=0}^{P_A} \left\{ \left[\frac{\beta}{\beta + \delta} \sum_{L_I=2P_C}^{P_A+P_C} \left(\mathcal{F}(P_A, P_C, L_I) \times \frac{\beta^{L_I} \delta^{2P_A-L_I}}{(\beta + \delta)^{2P_A}} \right) \right] \left(e^{P_C+1} (P_C + 1)^{-P_C-1} \Gamma(P_C + 1, P_C + 1) \right) \right\} \end{aligned} \quad (3.27)$$

$$\text{Where } |\mathcal{F}(P_A, P_C, L_I)| = \begin{cases} \binom{P_A}{P_C} \frac{\prod_{k=0}^{L_I-2P_C-1} (2P_A-2P_C-2k)}{(L_I-2P_C)!}, & \text{if } L_I - 2P_C - 1 \geq 0 \\ \binom{P_A}{P_C}, & \text{if } L_I - 2P_C - 1 < 0 \end{cases}$$

It should be noted that the simulation program, which concepts are introduced in Section 4.1.1, can only be used to simulate a given G_I once after every edit. Thus, it will be too complex to implement a simulation program to verify the whole of Equation 3.27 since there are too many kinds of G_I within it if the P_A is a large number. However, we can achieve the simulation for some certain G_I , which only have competing paths. It will be easier for us to obtain the structure of $K_{2,P_A} + e$ in the program when only the competing paths exist. We can do this by adding the number of competing paths consecutively by a for loop in MATLAB. This kind of G_I is shown in the following Figure 6:

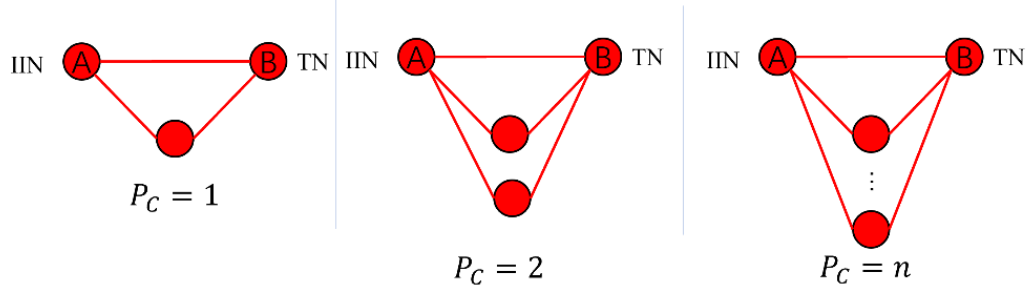


Figure 6: $K_{2,p_A} + e$ with only show competing paths. The alternative paths are not shown for the simplicity of the figure.

Thus, on the one hand, we can simulate the direct infection probability between IIN and TN under the existence of competing paths, which is $Pr(IIN \rightarrow TN|G_I)$. On the other hand, the mathematical expression of the conditional probability is also given as Equation 3.26 shows, which is easy to achieve in MATLAB. The following Figure 7 shows the comparison of the mathematical results and the simulation results for the number of competing paths from 1 to 50:

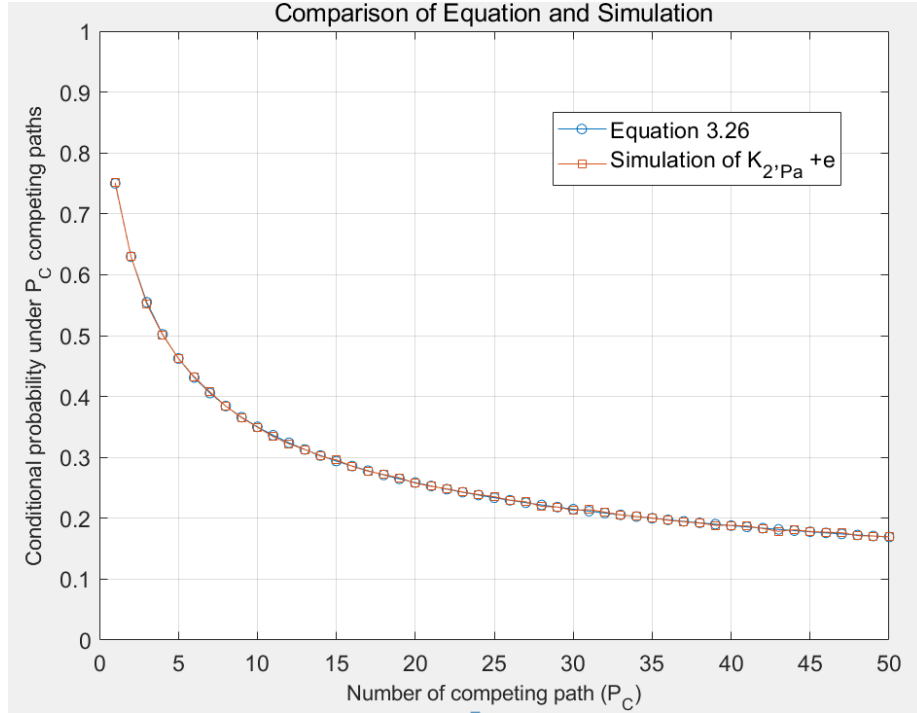


Figure 7: Comparison of results from the mathematical calculation and simulation for the conditional probability of direct infection under $K_{2,p_A} + e$ structure with only competing paths exist.(Full code is given in Appendix C.)

As Figure 7 indicates, the two curves representing the results from the calculation and simulation almost overlap, which means that our simulation and calculation are perfectly matched, and our derivation is correct.

5. Application and Results

In this chapter, we first implement the IIN-TN method through MATLAB and apply it to the human contact network in the real world. Based on the real-world network, we could get the distribution of R_0 . Based on the distribution, we will further explore the differences in the results of R_0 between the IIN-TN method and the traditional Star method with different infectious diseases that have occurred in history, for example, Influenza, Measles, SARS-CoV-1, and SARS-CoV-2.

5.1 Code implementation of IIN-TN method

In Section 3.2.2, we provide the analytical solution of structure $K_{2,p_A} + e$, and further utilize it as the fundamental structure of the IIN-TN method. The whole process of the IIN-TN method has been implemented in MATLAB (Full code is provided in Appendix C). The method could be applied to any network. Some examples of simple networks, such as K_3 and K_4 , are given below in Figure 8 (All examples provided in this section use the special value $\beta = 1$ and $\delta = 1$ for consistency with the examples in the above sections) :

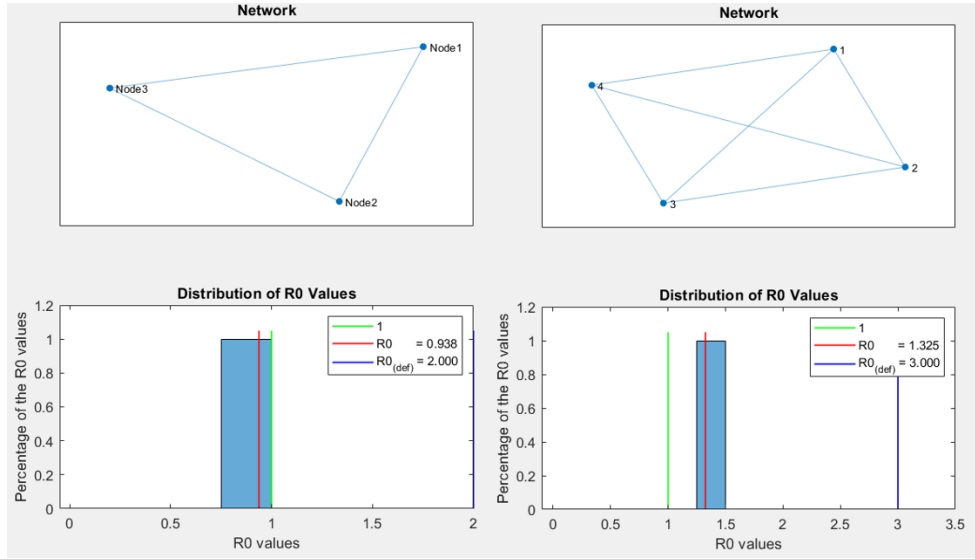


Figure 8: The R_0 distribution of K_3 and K_4 using the IIN-TN method in MATLAB. (Full code is provided in Appendix C). The green line represents 1, which is the threshold value that will determine whether the epidemic will spread out or die out. The red line represents the R_0 value that is derived from the IIN-TN method. The blue line represents the $R_{0(def)}$, which will be introduced in Equation 5.1. The blue histogram shows the percentage of $R_0(node)$ values, where each $R_0(node)$ value is derived for each node when the node is the IIN. For the two cases in Figure 8, since both K_3 and K_4 are symmetric structures, every node is the same, $R_0(node)$ has the same value, and the percentage is 1.

In each graph of Figure 8, there are three vertical lines in the color green, red, and blue. Which separately represents 1, R_0 of the network by IIN-TN method and $R_{0(def)}$. The R_0 of K_3

and K_4 are 0.938 and 1.325 separately as Figure 8 shows. The two results match the results that we derive in Section 3.2.1, Equation 3.19, and Section 3.2.2, Equation 3.37. The consistency of results shows that our implementation of code is successful.

Apart from the R_0 of a network derived by IIN-TN method, we use another definition $R_{0(def)}$ as a comparison, which is the ratio between effective spreading rate τ and epidemic threshold τ_c :

$$R_{0(def)} \triangleq \frac{\tau}{\tau_c} [8] \quad (5.1)$$

Because we know that for both case:

- i. $R_0 < 1$ [4, 9]
- ii. $\tau < \tau_c$ [4, 5]

the epidemic dies out. Where τ is the effective infection rate for a SIS model [5, 6]:

$$\tau = \frac{\beta}{\delta} \quad (5.2)$$

and τ_c is the epidemic threshold which could be derived from the largest eigenvalue of adjacency matrix A of the network [5]:

$$\tau_c = \frac{1}{\lambda_{max}(A)} \quad (5.3)$$

Thus, based on Equation 5.1-5.3, we can derive the equation for $R_{0(def)}$ as:

$$R_{0(def)} = \frac{\beta \lambda_{max}(A)}{\delta} \quad (5.4)$$

The blue histogram in Figure 9 represents the percentage of $R_0(node)$ values. For each node in the network, the IIN-TN method will find the R_0 of a node when the node is regarded as the IIN. Thus, the percentage of $R_0(node)$ values is the ratio, which is calculated as the ratio of the number of $R_0(node)$ that is within the specific range to the total number of nodes.

Because of the symmetry of K_3 and K_4 , each node in K_3 or K_4 is identical, thus the R_0 of each node is also identical and both the distribution histograms only have 1 bin in Figure 8. This phenomenon only appears in the complex graph. The distribution histogram is different from a random graph which is constructed based on the Erdős–Rényi model [10]:

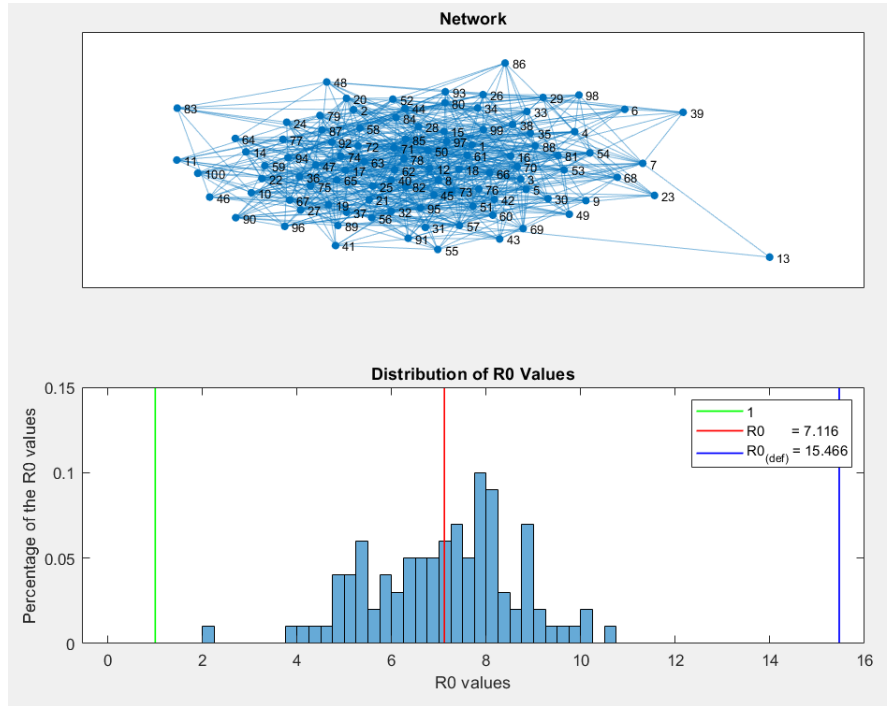


Figure 9: The R_0 distribution of an ER-random graph using the IIN-TN method. The shape of the distribution is a bell shape, which satisfies the properties of an ER-random graph.

The network $G_{ER}(n,p)$ in Figure 9 is created based on Erdős-Rényi model with node number $n = 100$, and the probability for each link $p = 0.15$. The distribution histogram of $G_{ER}(100,0.15)$ is the shape of normal distribution, which makes sense and matches the characteristic of the Erdős-Rényi model.

Apart from the IIN-TN method, another method, the Star method also be implemented. The following Figure 10 shows the degree distribution of $G_{ER}(100,0.15)$:

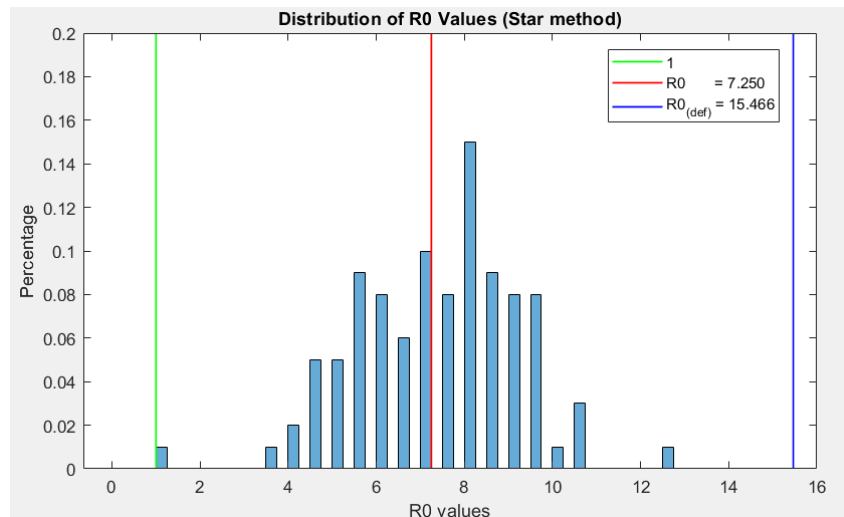


Figure 10: The R_0 distribution of an ER-random graph using Star method. The Star method will count the degree of each node with the consideration of infection probability. The Star method could also be regarded as degree distribution times the infection probability.

The Star method will not consider the cycles in the network. It only counts the degree of each node times the infection probability. Thus, the R_0 of the network by the Star method will be smaller than the R_0 of the network by the IIN-TN method as Figure 9 and Figure 10 show. Please note that all the figures in this section are carried out under the premise that $\beta = 1$ and $\delta = 1$. Thus, the detailed discussion in this distribution will be meaningless. However, in the next section, the two methods will be implemented in a real human-contact network with the data from the infectious disease that truly appears in human history.

5.2 Application in human-contact network

5.2.1 Dataset

Dataset 1 [11]

In 2018, an undirected network of around 100 students from the Massachusetts Institute of Technology (MIT) was collected by using location data from GPS. During the 9 months, the research records every face-to-face contact between the students. In this network, a node represents a person, and a link between two nodes indicates the corresponding person has physical contact. Figure 11 shows the real human-contact network after implementing the data processing on the original data (Code about data processing is given in Appendix C):

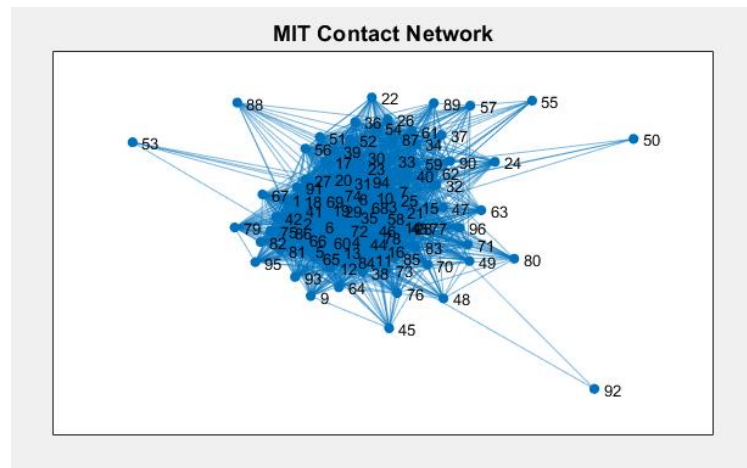


Figure 11: The human-contact network of 96 students from MIT. The total number of nodes is 96 and total number of links is 2539.

Dataset 2 [12]

This high-density network is constructed based on the high-resolution data of CPIs. Where CPIs stand for the close proximity interactions. Researchers used wireless sensor network technology to collect the CPI data from an American high school within a typical day. They collected 762,868 CPIs among 788 individuals. This means the whole network has 788 nodes in total. The entire network is too extensive for our program to run. The time needed is unacceptable when implementing our algorithm on this network based on my current device. Thus, to find

the balance between computation cost and scale of the network, we resample the dataset and only focus on the first 100 nodes.

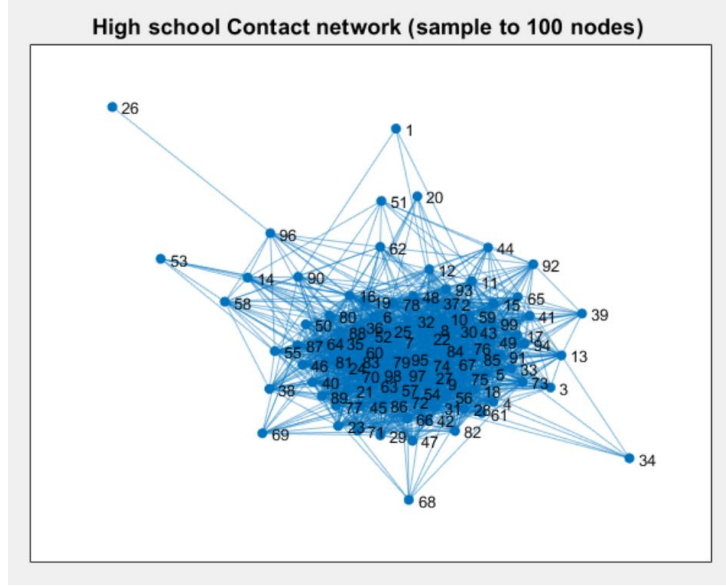


Figure 12: The high-resolution human contact network of around 100 students from an American high school. The number of nodes is 99 and number of links is 1707.

5.2.2 Infectious Disease

The data of four diseases are shown in the following Table 17.

Disease	Influenza	Measles	SARS-CoV-1	SARS-CoV-2 (COVID19)
Infectious Period (days)	1 [13]	5 [14]	4.5 [15]	10.91[16]
Curing rate (δ)	1	$\frac{1}{5}$	$\frac{2}{9}$	$\frac{1}{10.91}$
Empirical R_0	1.28 [17]	15 [18]	2.87 [19]	2.45 [16]

Table 17: Empirical data of four infectious disease.

These four infectious diseases are all notorious in history. According to Table 17, they have very different infectious characteristics. Take the two SARS-CoV diseases as an example.

Severe Acute Respiratory Syndrome Coronavirus 1 (SARS-CoV-1) and Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) are two distinct viruses. Still, they are closely related and belong to the same family of viruses known as Coronavirus. Among the many differences, the most significant difference is the scope and outcome of the two final spreads. The outbreak of SARS-CoV-1 occurred between 2002 and 2003. The virus primarily spread from Guangdong Province in China to other parts of China and then to various countries, leading to a global outbreak. The most affected areas were in Asia, including China, Hong Kong, Taiwan, Singapore, and Vietnam. SARS-CoV-2, known as COVID-19, has exhibited a much broader and sustained global spread than SARS-CoV-1. Since its emergence in late 2019 in Wuhan, China, COVID-19 has rapidly spread to nearly every corner of the world. COVID-19

has spread to multiple continents and countries, and the virus continues to be a primary concern even after significant efforts to control its transmission through vaccination campaigns, public health measures, and travel restrictions.

Although these four infectious diseases perform very differently, we use four indicators to measure them:

- i. Infectious period
- ii. Curing rate (δ)
- iii. Empirical basic reproductive number (R_0)
- iv. Infection rate (β)

Where the infectious period refers to the number of days that an infected patient remains contagious. Usually, we cannot directly find the curing rate of a disease from the analytical data in the academic paper. However, the curing rate is the reciprocal of the infectious period. Once we get the infectious period, it is very easy to derive the curing rate of a disease by finding the reciprocal of the infectious period. In addition to the cure rate, the infection rate is also a value that is difficult to quantify clinically. However, we have the empirical R_0 of the diseases. Based on the Star method, we could find the infection rate β that makes the R_0 of the human contact network become identical to the empirical R_0 . Tables 18 and 19 give the infection rate β of the four diseases based on the network in Figure 11 and Figure 12 with empirical R_0 in Table 17:

Disease	Influenza	Measles	SARS-CoV-1	SARS-CoV-2 (COVID19)
Infection rate (β) for Dataset 1	0.0248	0.0792	0.0128	0.0045

Table 18: Calculated infection rate based on human contact network from dataset 1 and empirical R_0 in Table 17.

Disease	Influenza	Measles	SARS-CoV-1	SARS-CoV-2 (COVID19)
Infection rate (β) for Dataset 2	0.0385	0.1539	0.0202	0.0070

Table 19: Calculated infection rate based on human contact network from dataset 2 and empirical R_0 in Table 17.

5.3 Results

With the curing rate in Table 17 and the infection rate in Table 18. We could get the distribution histogram of the diseases in the human contact network by the method of IIN-TN. Meanwhile, the distribution histogram by the Star method will also be given along with the histogram by the IIN-TN method for easier comparison. Figure 12-15 are the results for four diseases in human contact network from MIT.

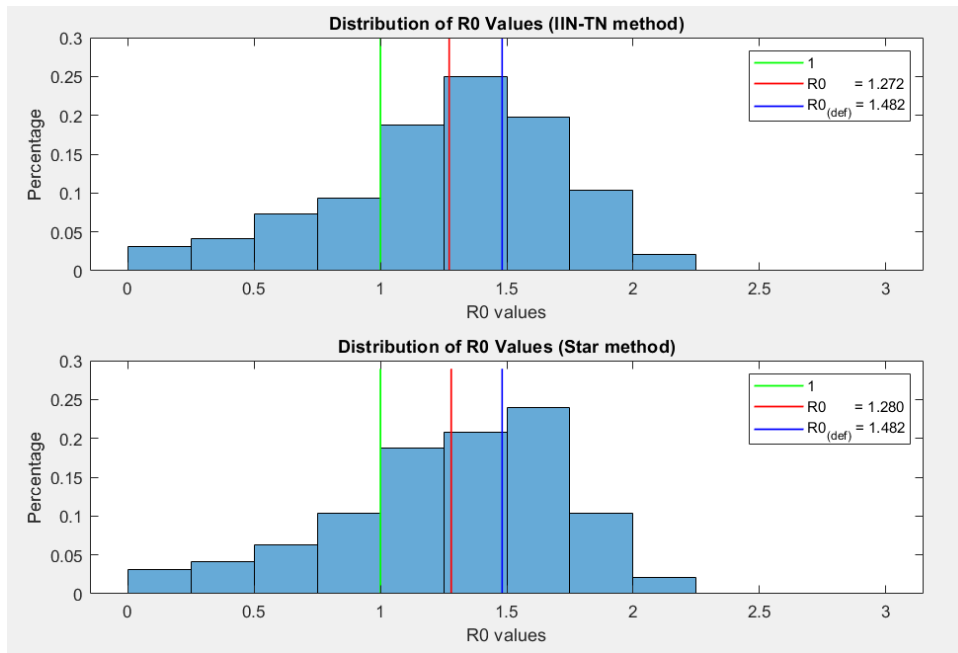


Figure 12: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for Influenza based on the human contact network of MIT.

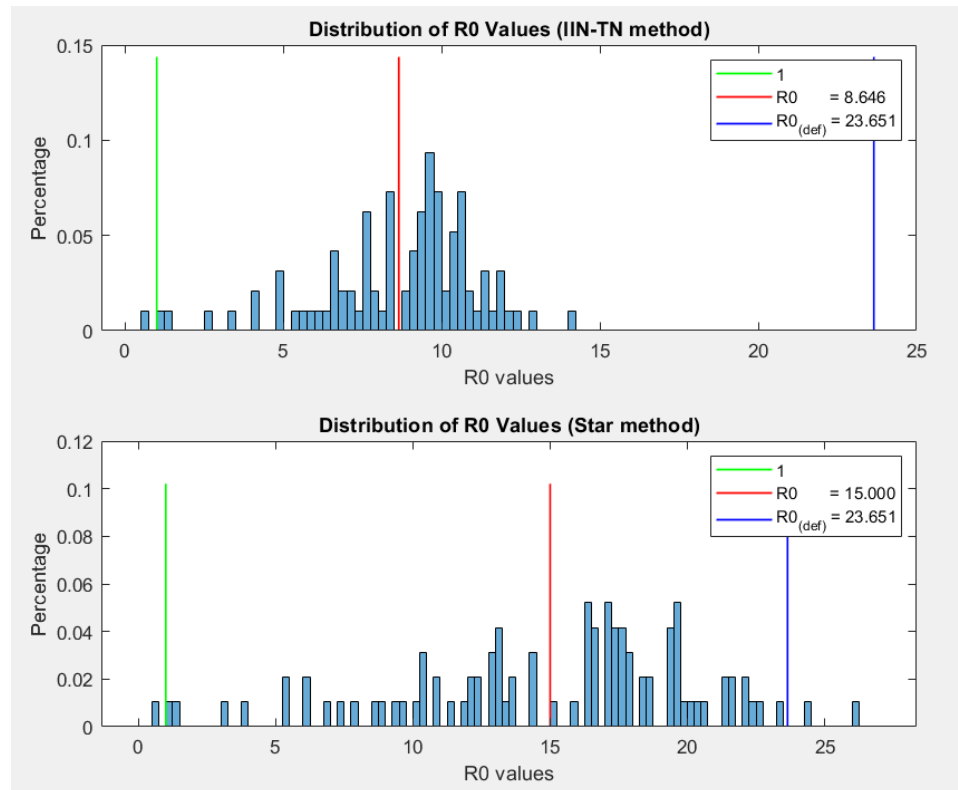


Figure 13: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for Measles based on the human contact network of MIT.

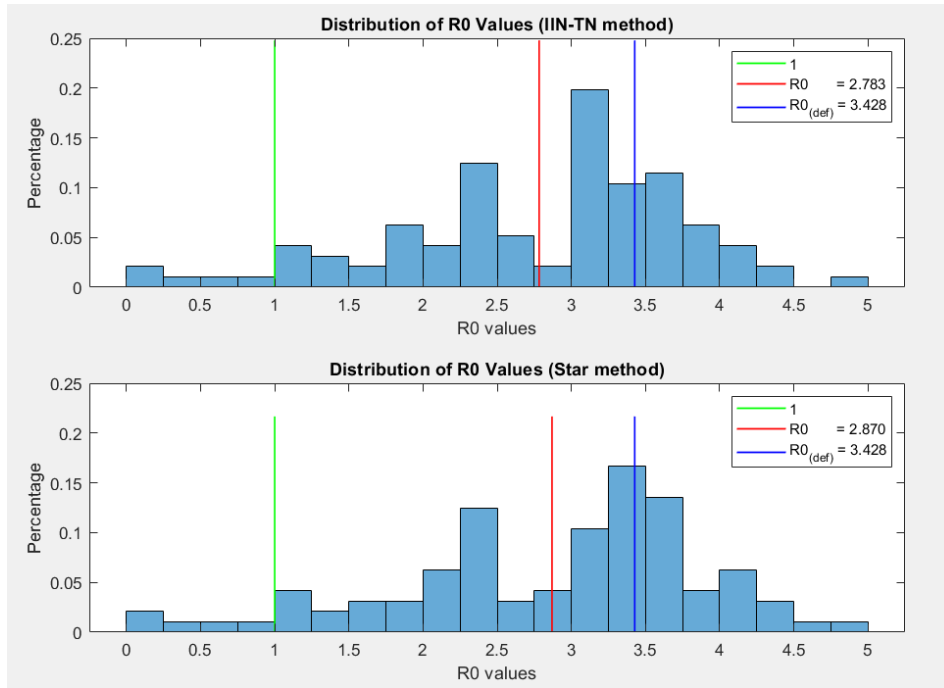


Figure 14: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for SARS-CoV-1 based on the human contact network of MIT.

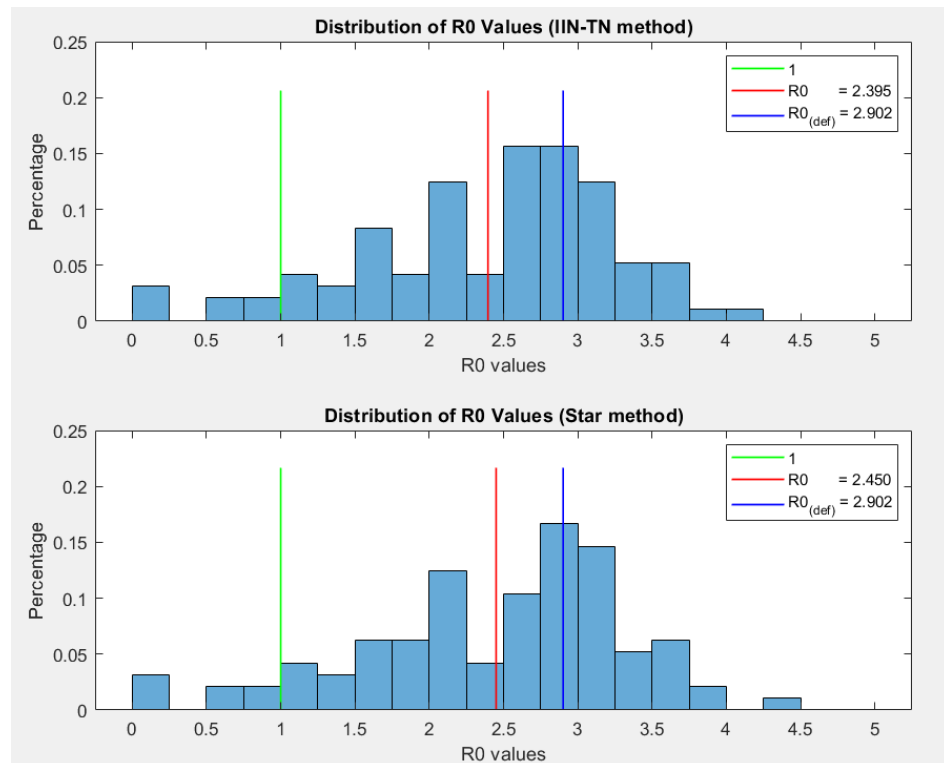


Figure 15: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for SARS-CoV-2 based on the human contact network of MIT.

Following the same method and procedure, we can also get the distribution histogram of four diseases by two methods with the curing rate in Table 17 and infection rate in Table 19. The results for four diseases in the human contact network from American high school are shown in Figure 16-19.

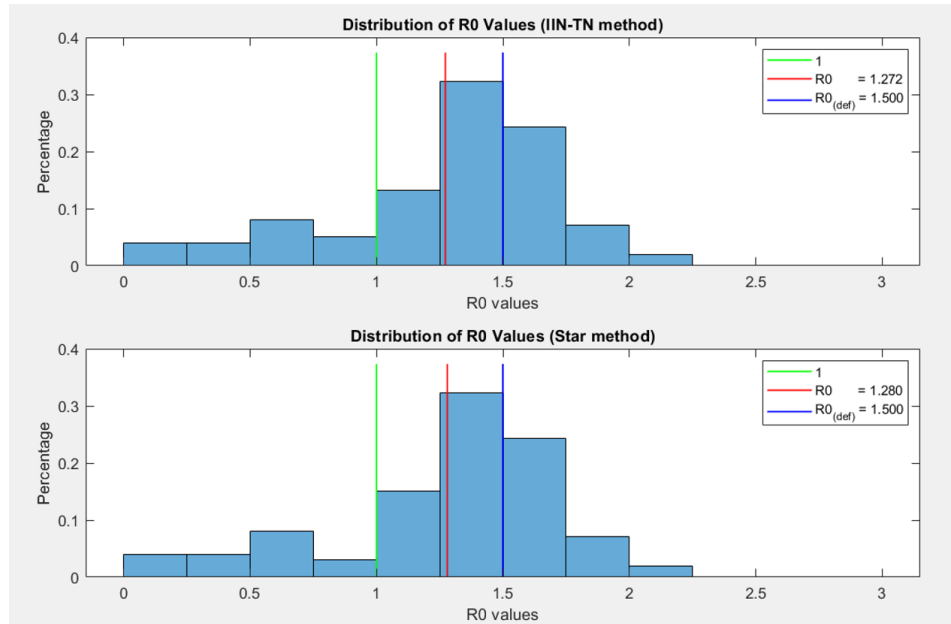


Figure 16: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for Influenza based on the human contact network of high school.

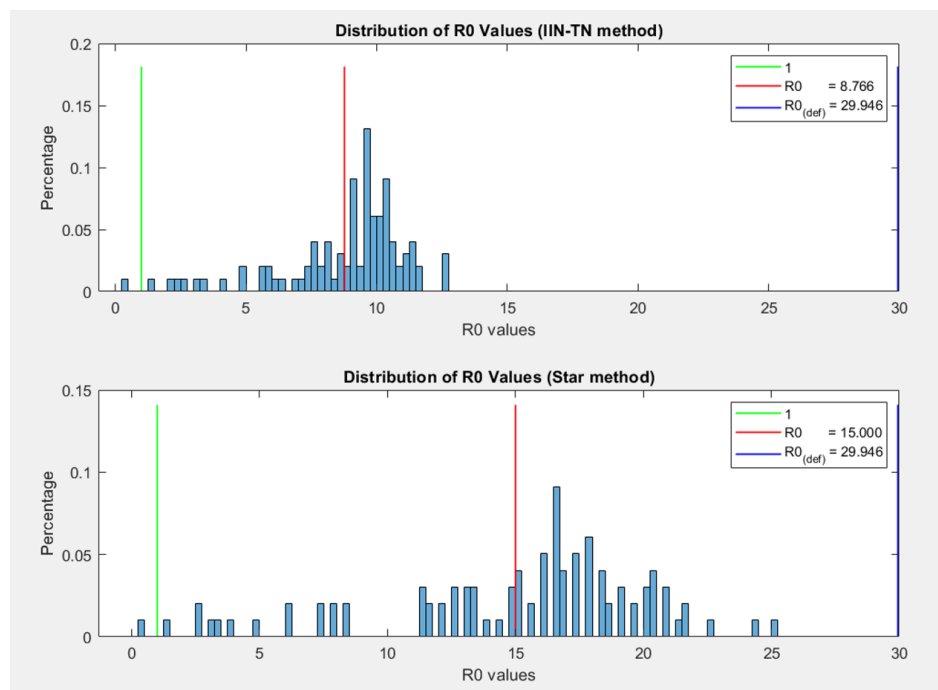


Figure 17: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for Measles based on the human contact network of high school.

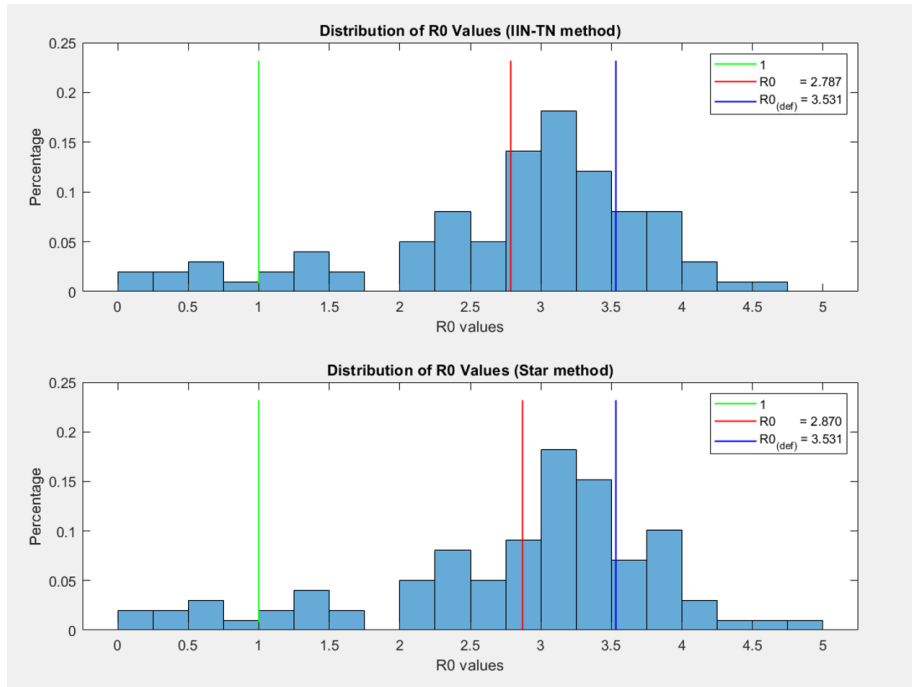


Figure 18: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for SARS-CoV-1 based on the human contact network of high school.

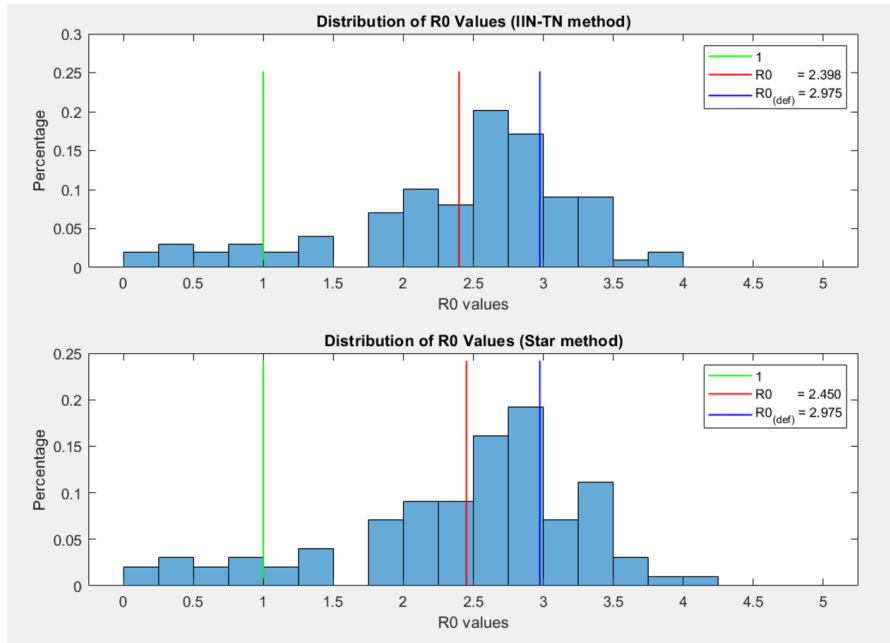


Figure 19: The comparison between two distributions of R_0 values derived by the IIN-TN method and Star method separately for SARS-CoV-2 based on the human contact network of high school.

From the eight figures for two human contact networks, we can find that the R_0 of the network by the IIN-TN method is smaller than the Star method for four cases. This phenomenon matches our expectation because the nature of the IIN-TN method is to consider the alternative path between the neighboring two nodes, which will decrease the direct infection probability. However, according to R_0 results in Figure 12-19, the reduction of R_0 of the network is not significant for the three diseases: Influenza, SARS-CoV-1, and SARS-CoV-2. The only obvious reduction happens in Measles. This is also reflected in the distribution histogram: only the Measles experienced a significant distribution shift from the higher part to the lower part. This could be explained by the infection probability and effective infection rate of diseases as stated in the following Table 20:

Disease	Influenza	Measles	SARS-CoV-1	SARS-CoV-2 (COVID19)
Infection probability $(\frac{\beta}{\beta+\delta})$	0.0242	0.2836	0.0545	0.0468
Effective infection rate $(\frac{\beta}{\delta})$	0.0248	0.3960	0.0121	0.043

Table 20: Infection probability and Effective infection rate of four disease.

According to Table 20, Measles has a significantly larger infection rate and effective infection rate compared with other three diseases. On the one hand, the more significant effective infection rate, together with a larger infection probability, could indicate that Measles will spread much faster in the network than other diseases, which means there is more possibility for the infection to go through the alternative paths instead of the direct link. On the other hand, the more significant infection probability will make the results derived from Equation 3.27 decrease more compared with a more negligible infection probability. Both could explain why Measles has a more noticeable change. This could also lead to a conclusion that the IIN-TN method will be more effective for a disease with a more significant infection probability and a more effective infection rate.

6. Conclusion and Future Research

This chapter is the last chapter that will summarize what has been achieved in this thesis and provide some possible direction for future research.

6.1 Conclusion

In this thesis project, based on the original concepts of the basic reproductive number (R_0) of infectious disease, we propose a new approach to calculate the R_0 based on the structure of the network and implement this method in MATLAB and, last, apply it to the human contact network.

In Chapter 1 and Chapter 2, we first give the introduction of the basic concept of epidemiology, especially the parameter in the SIS model, and then derive the infection probability of a link equal to $\lambda = \frac{\beta}{\beta + \delta}$. Based on this infection probability, in Chapter 3, we

first use the Holistic method to find the R_0 of some simple network: P_3 , K_3 , and K_4 . Then, we propose a new method: the IIN-TN method, which could take the alternative paths between two nodes into consideration when finding the direct infection probability between IIN and TN: $Pr(IIN \rightarrow TN)$. The consistent results of the IIN-TN method with the holistic method in K_3 , and K_4 prove the correctness of IIN-TN method. Then, we further expand the IIN-TN method to a large and complex network with the help of $K_{2,p_A} + e$ structure. The analytical solution is given in Section 3.2.2. Meanwhile, we also point out the shortcomings of only adopting $K_{2,p_A} + e$ structure to a complex network at the end of Chapter 3.

In Chapter 4, based on the essence of the SIS model, we propose a simulation method that could randomly generate the infection time of a link. Then, this simulation method is used to verify the theory discussed in Chapter 3. After the successful verification, the analytical solution in Chapter 3 is implemented as MATLAB algorithms in Chapter 5 and applied to a real human contact network with data from different infectious diseases.

The application results indicate that the IIN-TN method indeed gives a lower R_0 value of the network compared with the R_0 given in the method, which follows the traditional definition. This finding meets our expectations and shows that the IIN-TN method have the ability to further analyze the local structure of a network. Apart from that, according to the results from different infectious diseases, the IIN-TN method will be more effective if the disease has a relatively larger infection probability and effective infection rate.

The calculation and derivation of R_0 have been discussed for a long time. Some researchers try to estimate it by empirical data from a statistical perspective [2, 3], and some modelers try to derive it using epidemic models from a mathematical perspective [9, 20, 21]. However, none emphasized the contact network and linked the R_0 with the network itself. This is why some scholars think R_0 cannot reflect the information from the underlying contact network[4]. However, the IIN-TN method proposed in this paper could directly calculate R_0

from any contact network. Furthermore, the IIN-TN method could give the R_0 distribution based on the contact network, which could further reveal the properties of the contact network. Thus, we believe this paper is a solid addition to the field of finding R_0 because this paper connects the contact network with R_0 directly, and we wish our work could bring more inspiration to this area.

6.2 Future research

However, there are still two directions that are worth further research:

- i. In Chapter 3, we only give the analytical solution of structure $K_{2,p_A} + e$. The limitation of this structure is obvious: it can only deal with the alternative path of 2 links. For the alternative path has 3 links or more, our method will ignore it by default. This simplification will make the results larger than the theoretical result. There may be many competing paths that has 3 links or more, which will lower the theoretical results than the results derived from our IIN-TN method.
- ii. The human contact network from MIT used in Chapter 5 is a relatively small network with only 96 persons. Besides, all the people in the network are students from MIT, which has the limitation that the network does not have a powerful universality for an epidemic outbreak in society. Thus, a more extensive network with all kinds of people could provide more accurate and meaningful data after applying the IIN-TN method.

Bibliograph

- [1] K. Dietz, "The estimation of the basic reproduction number for infectious diseases," *Statistical methods in medical research*, vol. 2, no. 1, pp. 23-41, 1993.
- [2] S. W. Park, B. M. Bolker, D. Champredon, D. J. D. Earn, M. Li, J. S. Weitz, B. T. Grenfell, and J. Dushoff, "Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak," *Journal of the Royal Society, Interface*, vol. 17, no. 168, pp. 20200144, 2020.
- [3] G. Chowell, H. Nishiura, and L. s. M. A. Bettencourt, "Comparative estimation of the reproduction number for pandemic influenza from daily case notification data," *Journal of The Royal Society Interface*, vol. 4, no. 12, pp. 155-166, 2007.
- [4] A. Socievole, F. De Rango, C. Scoglio, and P. Van Mieghem, "Assessing network robustness under SIS epidemics: The relationship between epidemic threshold and viral conductance," *Computer Networks*, vol. 103, pp. 196-206, 2016.
- [5] P. Van Mieghem, J. Omic, and R. Kooij, "Virus Spread in Networks," *IEEE/ACM Transactions on Networking*, vol. 17, no. 1, pp. 1-14, 2009.
- [6] B. Qu, and H. Wang, "SIS Epidemic Spreading with Heterogeneous Infection Rates," *IEEE Transactions on Network Science and Engineering*, vol. 4, no. 3, pp. 177-186, 2017.
- [7] M. Dekking, "A modern introduction to probability and statistics : understanding why and how," Springer texts in statistics, Springer, 2005.
- [8] R. Persoons, M. Sensi, B. Prasse, and P. Van Mieghem, "Transition from time-variant to static networks: timescale separation in NIMFA SIS," arXiv preprint arXiv:2305.12446, 2023.
- [9] P. van den Driessche, "Reproduction numbers of infectious disease models," *Infectious Disease Modelling*, vol. 2, no. 3, pp. 288-303, 2017/08/01/, 2017.
- [10] P. Erdős, and A. Rényi, "On the evolution of random graphs," *Publ. math. inst. hung. acad. sci*, vol. 5, no. 1, pp. 17-60, 1960.
- [11] K. Abbas, "Real Human Contact processed data," 2018.
- [12] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, "A high-resolution human contact network for infectious disease transmission," *Proc Natl Acad Sci U S A*, vol. 107, no. 51, pp. 22020-5, Dec 21, 2010.
- [13] A. Cori, A. J. Valleron, F. Carrat, G. Scalia Tomba, G. Thomas, and P. Y. Boëlle, "Estimating influenza latency and infectious period durations using viral excretion data," *Epidemics*, vol. 4, no. 3, pp. 132-138, 2012/08/01/, 2012.
- [14] M. J. Keeling, and B. T. Grenfell, "Understanding the persistence of measles: reconciling theory, simulation and observation," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 269, no. 1489, pp. 335-343, 2002.
- [15] Z. Zhang, "The outbreak pattern of SARS cases in China as revealed by a mathematical model," *Ecological Modelling*, vol. 204, no. 3, pp. 420-426, 2007/06/16/, 2007.
- [16] C. B. Biggerstaff M, Cucunubá ZM, Dinh L, Ferguson NM, Gao H, et al, *Early insights from statistical and mathematical modeling of key epidemiologic parameters of COVID-19*, Emerg Infect Dis, Emerg Infect Dis, 2020 Nov
- [17] M. Biggerstaff, S. Cauchemez, C. Reed, M. Gambhir, and L. Finelli, "Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review

- of the literature," *BMC Infectious Diseases*, vol. 14, no. 1, pp. 480, 2014/09/04, 2014.
- [18] F. M. Guerra, S. Bolotin, G. Lim, J. Heffernan, S. L. Deeks, Y. Li, and N. S. Crowcroft, "The basic reproduction number (R_0) of measles: a systematic review," *The Lancet Infectious Diseases*, vol. 17, no. 12, pp. e420-e428, 2017/12/01/, 2017.
 - [19] J. Liu, W. Xie, Y. Wang, Y. Xiong, S. Chen, J. Han, and Q. Wu, "A comparative overview of COVID-19, MERS and SARS: Review article," *Int J Surg*, vol. 81, pp. 1-8, Sep, 2020.
 - [20] P. van den Driessche, and A.-A. Yakubu, "Demographic population cycles and R_0 in discrete-time epidemic models," *Journal of Biological Dynamics*, vol. 13, no. sup1, pp. 179-200, 2019/03/15, 2019.
 - [21] N. Hernández-Cerón, Z. Feng, and P. van den Driessche, "Reproduction numbers for discrete-time epidemic models with arbitrary stage distributions," *Journal of Difference Equations and Applications*, vol. 19, no. 10, pp. 1671-1693, 2013/10/01, 2013.
 - [22] S. National Institute of, Technology, and F. W. J. Olver, *NIST digital library of mathematical functions : companion to the NIST handbook of mathematical functions*, Cambridge: Cambridge University Press for the NIST, 2010.

Appendix

Appendix A: Derivation

A.1 The derivation of $Pr(I < R)$

Both the infection and curing processes are Poisson process with rate β and δ separately. In Section 2.1, we derive the two equations for the time before the event happening from the perspective of exponential distribution. Here we provide another approach by using the Erlang distribution. The Erlang distribution describe the k independent exponential variables, when $k = 1$, the Erlang distribution could be simplified to exponential distribution. (When $k \neq 1$, Erlang distribution is an effective tool when we face multiple independent Poisson processes. The application of Erlang distribution when $k \neq 1$ is given in Section A.2.)

Thus, we could get the two equations from Erlang distribution by substituting $k = 1$ and $\lambda = \beta$ or $\lambda = \delta$ to PDF of Erlang distribution: $f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}$ to get the functions:

$$f_I(t) = \beta e^{-\beta t}, \quad (1)$$

$$f_R(r) = \delta e^{-\delta r}. \quad (2)$$

Then, we could derive $Pr(I < R)$ with (1), (2) by implementing law of total probability:

$$\begin{aligned} Pr(I < R) &= \int_{-\infty}^{\infty} Pr(I < R | R = r) f_R(r) dr \\ &= \int_{-\infty}^{\infty} Pr(I < r) f_R(r) dr \\ &= \int_{r=0}^{\infty} \int_{t=0}^r f_I(t) f_R(r) dt dr \\ &= \int_{r=0}^{\infty} \int_{t=0}^r (\beta e^{-\beta t}) (\delta e^{-\delta r}) dt dr \\ &= \int_{r=0}^{\infty} \left[\int_{t=0}^r (\beta e^{-\beta t}) dt \right] (\delta e^{-\delta r}) dr \\ &= \int_{r=0}^{\infty} (1 - e^{-\beta r}) (\delta e^{-\delta r}) dr \\ &= \int_{r=0}^{\infty} \delta e^{-\delta r} dr - \int_{r=0}^{\infty} \delta e^{-(\beta+\delta)r} dr \\ &= 1 - \frac{\delta}{\beta + \delta} = \frac{\beta}{\beta + \delta}. \end{aligned} \quad (3.1)$$

A.2 The derivation of $Pr(I_1 < I_2 + I_3)$

I_1 , I_2 , and I_3 are the infection time from three identical and independent infection process described by Poisson process. I_1 represents the time that direct path needs to spread the infection from infected node to the susceptible node. $I_2 + I_3$ represents the time that the competing path needs to spread the infection from infected node to the susceptible node. $Pr(I_1 < I_2 + I_3)$ is the probability that the susceptible node is infected by the infected node

through direct path (has only one link) instead of through the competing path (has two links).

To solve this probability, we can use the law of total probability as Equation (3) shows, derive the function as:

$$\begin{aligned}
Pr(I_1 < I_2 + I_3) &= Pr(I_2 + I_3 > I_1) \\
&= \int_{t=-\infty}^{\infty} Pr(I_2 + I_3 > I_1 | I_1 = t) f_I(t) dt \\
&= \int_{t=-\infty}^{\infty} Pr(I_2 + I_3 > t) f_{I_1}(t) dt \\
&= \int_{t=0}^{\infty} \int_{s=t}^{\infty} f_{I_2+I_3}(s) ds f_{I_1}(t) dt \\
&= \int_{t=0}^{\infty} f_I(t) \left(\int_{s=t}^{\infty} f_{I_2+I_3}(s) ds \right) dt \\
&= \int_{t=0}^{\infty} f_I(t) \left(1 - \int_{s=0}^t f_{I_2+I_3}(s) ds \right) dt \\
&= \int_{t=0}^{\infty} f_I(t) (1 - F_{I_2+I_3}(t)) dt.
\end{aligned} \tag{4}$$

Since infection process 2 and 3 are identical Poisson process, we could get the CDF function $F_{I_2+I_3}(t)$ directly from Erlang distribution by substituting $k = 2$ and $\lambda = \beta$ to (5)

$$F(x) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n \tag{5}$$

and get:

$$F_{I_2+I_3}(t) = 1 - e^{-\beta t} - \beta t e^{-\beta t}. \tag{6}$$

Then, substitute (1), (6) to (4):

$$\begin{aligned}
Pr(I_1 < I_2 + I_3) &= \int_{t=0}^{\infty} \beta e^{-\beta t} (e^{-\beta t} + \beta t e^{-\beta t}) dt \\
&= \int_{t=0}^{\infty} (\beta e^{-2\beta t} + \beta^2 t e^{-2\beta t}) dt \\
&= \int_{t=0}^{\infty} \beta e^{-2\beta t} dt + \int_{t=0}^{\infty} \beta^2 t e^{-2\beta t} dt.
\end{aligned} \tag{7}$$

Then, we could calculate (7) separately.

To calculate $\int_{t=0}^{\infty} \beta e^{-2\beta t} dt$, substitute $u = -2\beta t, du = -2\beta dt$, and this also gives a new lower bound $u = -\infty$, and a new upper bound $u = 0$:

$$\int_{t=0}^{\infty} \beta e^{-2\beta t} dt = \frac{1}{2} \int_{-\infty}^0 e^u du = \frac{1}{2} \tag{7.1}$$

Similarly, to calculate $\int_{t=0}^{\infty} \beta^2 t e^{-2\beta t} dt$, using integrate by part and substitute

$$g = -\frac{e^{-2\beta t}}{2\beta}, \quad \frac{d}{dt} g = e^{-2\beta t} \text{ to it:}$$

$$\begin{aligned}
\int_{t=0}^{\infty} \beta^2 t e^{-2\beta t} dt &= \int_{t=0}^{\infty} \beta^2 t g' dt = \beta^2 \left([tg]_0^{\infty} + \int_{t=0}^{\infty} t' g dt \right) \\
&= \left[-\frac{t\beta e^{-2t\beta}}{2} \right]_0^{\infty} + \int_{t=0}^{\infty} \frac{\beta e^{-2\beta t}}{2} dt \\
&= \frac{-\infty\beta e^{-2\infty\beta}}{2} + \frac{0\beta e^{-0}}{2} + \frac{1}{4} \\
&= 0 + 0 + \frac{1}{4} = \frac{1}{4}
\end{aligned} \tag{7.2}$$

Then, sum up (7.1) and (7.2) to (7) and we get:

$$Pr(I_1 < I_2 + I_3) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \tag{8}$$

A.3 The derivation process of $Pr(I_1 < L_A I)$

L_A is the number of links in one alternative path. $L_A I$ means the time that needed for a competing path with L_A links in this path to spread infection from the infected node to target node. $Pr(I_1 < L_A I)$ is the probability that the infection in direct path is faster than competing path.

If we assign $L_A = 2$, repeat the derivation process A.2 , we could calculate

$$Pr(I_1 < I_2 + I_3 + I_4) = \frac{7}{8}, \tag{9}$$

because we can use the same method in (7.2) to get

$$\int_{t=0}^{\infty} \frac{1}{2 \times 1} \beta^3 t^2 e^{-2\beta t} dt = \frac{1}{8}. \tag{10}$$

Furthermore, we could find a general equation for the probability that infection process takes the direct path instead of competing path with L_A links as (only single competing path exists) :

$$Pr(I_1 < L_A I) = 1 - \frac{1}{2^{L_A}} \tag{11}$$

The proof is given below:

Follow the same process in (4) , we can prove (11) as following shows:

$$\begin{aligned}
Pr(I_1 < L_A I) &= Pr(L_A I > I_1) \\
&= \int_{t=-\infty}^{\infty} Pr(L_A I > I_1 | I_1 = t) f_I(t) dt \\
&= \int_{t=-\infty}^{\infty} Pr(L_A I > t) f_{I_1}(t) dt \\
&= \int_{t=0}^{\infty} \int_{s=t}^{\infty} f_{L_A I}(s) ds f_{I_1}(t) dt \\
&= \int_{t=0}^{\infty} f_I(t) \left(\int_{s=t}^{\infty} f_{L_A I}(s) ds \right) dt \\
&= \int_{t=0}^{\infty} f_I(t) \left(1 - \int_{s=0}^t f_{L_A I}(s) ds \right) dt \\
&= \int_{t=0}^{\infty} f_I(t) (1 - F_{L_A I}(t)) dt.
\end{aligned} \tag{12}$$

Where $F_{L_A I}(t)$ is the CDF function which indicate that infection process takes the competing path with L_A links in total. Since the infection along each link in this path are identical and independent, we could get the CDF function of $F_{L_A I}(t)$ directly from Erlang distribution by substituting $k = n$ and $\lambda = \beta$ to (5), and get:

$$F_{L_A I}(t) = 1 - e^{-\beta t} - \beta t e^{-\beta t} - \frac{1}{2 \times 1} (\beta t)^2 e^{-\beta t} - \dots - \frac{1}{(L_A - 1)!} (\beta t)^{L_A} e^{-\beta t}. \quad (13)$$

Substitute (1), (13) to (12):

$$\begin{aligned} & Pr(I_1 < L_A I) \\ &= \int_{t=0}^{\infty} \left(e^{-2\beta t} + \beta^2 t e^{-2\beta t} + \frac{1}{2 \times 1} \beta^3 t^2 e^{-2\beta t} + \dots + \frac{1}{(L_A - 1)!} \beta^{L_A} t^{L_A-1} e^{-2\beta t} \right) dt. \end{aligned} \quad (14)$$

The total number of terms in the integration of (14) is L_A , which is the same number of links in the competing path. The integration results of former three terms already be given in (7.1), (7.2), (10). Actually, all terms in (14) have the same structure with the last term:

$$\int_{t=0}^{\infty} \left(\frac{1}{(L_A - 1)!} \beta^{L_A} t^{L_A-1} e^{-2\beta t} \right) dt, \quad (15)$$

and a general result of (15) will help us to solve (14).

Here, we could take use of one equation from the list of integrals of exponential functions:

$$\int_0^{\infty} (x^N e^{-ax}) dx = \frac{N!}{a^{N+1}} (N = 0, 1, 2, \dots, Re(a) > 0), \quad (16)$$

to solve (15) because they have same structure.

Compare with (15) and (16), we could substitute $x = t$, $a = 2\beta$, $N = L_A - 1$ to (16):

$$\int_0^{\infty} (t^{L_A-1} e^{-2\beta t}) dt = \frac{(L_A - 1)!}{(2\beta)^{L_A}}, \quad (17)$$

and further substitute (17) to (15):

$$\begin{aligned} & \int_{t=0}^{\infty} \left(\frac{1}{(L_A - 1)!} \beta^{L_A} t^{L_A-1} e^{-2\beta t} \right) dt \\ &= \frac{\beta^{L_A}}{(L_A - 1)!} \times \int_0^{\infty} (t^{L_A-1} e^{-2\beta t}) dt \\ &= \frac{\beta^{L_A}}{(L_A - 1)!} \times \frac{(L_A - 1)!}{(2\beta)^{L_A}} \\ &= \frac{1}{2^{L_A}}. \end{aligned} \quad (18)$$

Based on (7.1), (7.2), (10) and (18), we could write (14) as:

$$Pr(I_1 < L_A I) = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots + \frac{1}{2^{L_A}}. \quad (19)$$

Equation (19) is the sum of a geometric series, which could be solved by using geometric series formular and the result is:

$$Pr(I_1 < L_A I) = 1 - \frac{1}{2^{L_A}} \quad (20)$$

A.4 The derivation of $Pr(I < \min(P_C(I + I)))$

I is the time needed for the direct path. P_C is the number of competing paths, and $P_C(I + I)$ indicates there are total number of P_C independent time for all competing paths (each path is composed of 2 links). Thus, $\min(P_C(I + I))$ refers to the shortest time among them. And $Pr(I < \min(P_C(I + I)))$ is the probability that the direct path is faster than any competing path because I is even smaller than the shortest time.

Repeat what we have done in A.2 and A.3:

$$\begin{aligned}
 Pr(I < \min(P_C(I + I))) &= Pr(\min(P_C(I + I)) > I) \\
 &= \int_{t=-\infty}^{\infty} Pr(\min(P_C(I + I)) > I | I = t) f_I(t) dt \\
 &= \int_{t=-\infty}^{\infty} Pr(\min(P_C(I + I)) > t) f_I(t) dt \\
 &= \int_{t=0}^{\infty} \int_{s=t}^{\infty} f_{\min(P_C(I+I))}(s) ds f_I(t) dt \\
 &= \int_{t=0}^{\infty} f_I(t) \left(\int_{s=t}^{\infty} f_{\min(P_C(I+I))}(s) ds \right) dt \\
 &= \int_{t=0}^{\infty} f_I(t) \left(1 - \int_{s=0}^t f_{\min(P_C(I+I))}(s) ds \right) dt \\
 &= \int_{t=0}^{\infty} f_I(t) (1 - F_{\min(P_C(I+I))}(t)) dt
 \end{aligned} \tag{20}$$

Every single competing path in $P_C(I + I)$ is composed of two links, similarly as (5) and (6) in A.2, we can get the CDF for one competing path $F_{I+I}(t)$ directly from Erlang distribution by substituting $k = 2$ and $\lambda = \beta$ to

$$F(x) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n$$

and get:

$$F_{I+I}(t) = 1 - e^{-\beta t} - \beta t e^{-\beta t} \tag{21}$$

Because the total number of competing paths is P_C , we need to find the relationship between $F_{I+I}(t)$ and $F_{\min(P_C(I+I))}(t)$. $F_{I+I}(t)$ is the CDF of $Pr(I + I < t)$, thus $1 - F_{I+I}(t)$ will represents $Pr(I + I > t)$. Then, we further take $1 - (1 - F_{I+I}(t))^{P_C}$ to make sure that the shortest infection time is smaller than t , which is exactly the CDF that we want to find:

$F_{\min(P_C(I+I))}(t)$. This gives us following equation:

$$\begin{aligned}
 F_{\min(P_C(I+I))} &= 1 - (1 - F_{I+I}(t))^{P_C} \\
 &= 1 - (1 - (1 - e^{-\beta t} - \beta t e^{-\beta t}))^{P_C} \\
 &= 1 - (e^{-\beta t} + \beta t e^{-\beta t})^{P_C}
 \end{aligned} \tag{22}$$

Substitute (22) back to (20):

$$\begin{aligned} Pr(I < \min(P_C(I + I))) &= \int_{t=0}^{\infty} f_I(t) \left(1 - 1 + (e^{-\beta t} + \beta t e^{-\beta t})^{P_C}\right) dt \\ &= \int_{t=0}^{\infty} \beta e^{-\beta t} (e^{-\beta t} + \beta t e^{-\beta t})^{P_C} dt \end{aligned} \quad (23)$$

Compare (23) with (7):

$$Pr(I < \min(P_C(I + I))) = \int_{t=0}^{\infty} \beta e^{-\beta t} (e^{-\beta t} + \beta t e^{-\beta t})^{P_C} dt \quad (23)$$

$$Pr(I_1 < I_2 + I_3) = \int_{t=0}^{\infty} \beta e^{-\beta t} (e^{-\beta t} + \beta t e^{-\beta t}) dt \quad (7)$$

We could easily find that the two equations have very similar form, and actually we could get (7) from (23) by substituting $P_C = 1$ to (23). The match of two equations proves the corrections of our derivation.

The result of integration in (23) will give us the analytical solution for any P_C that we want to choose. However, the direct integration of the power of a summation of exponential functions is difficult to derive, especially when P_C is large and the binomial expansion will be very complex. In this paper, a tricky method is used to solve this. Some examples in the above section, such as (7.1),(7.2) show that, the integration result of terms with infection rate β does not consist of β anymore. This finding is reasonable since we only want to compare the infection time of different path and get the corresponding probability, which should be irrelevant with infection rate. We can change our time scale, and we could further simplify (23) by substituting $\beta = 1$:

$$\begin{aligned} Pr(I < \min(P_C(I + I))) &= \int_{t=0}^{\infty} e^{-t} (e^{-t} + t e^{-t})^{P_C} dt \\ &= \int_{t=0}^{\infty} e^{-t} (e^{-t} (t + 1))^{P_C} dt \\ &= \int_{t=0}^{\infty} e^{-t} (e^{-t})^{P_C} (t + 1)^{P_C} dt \\ &= \int_{t=0}^{\infty} \frac{e^{-t(P_C+1)}}{(t + 1)^{-P_C}} dt \end{aligned} \quad (24)$$

Let $u = t + 1$, thus we have $t = u - 1$ and $dt = d(u - 1) = du$, substitute to (24):

$$Pr(I < \min(P_C(I + I))) = e^{P_C+1} \int_{u=1}^{\infty} \frac{e^{-u(P_C+1)}}{u^{-P_C}} du \quad (25)$$

In (25), $\int_{u=1}^{\infty} \frac{e^{-u(P_C+1)}}{u^{-P_C}} du$ matches the form of generalized exponential integral. Thus, we could further simplify (25) by following Equation 8.19.3 in NIST Digital Library of Mathematical Functions:

$$E_p(z) = \int_1^{\infty} \frac{e^{-zt}}{t^p} dt$$

Then, (25) could be represented as:

$$Pr(I < \min(P_C(I + I))) = e^{P_C+1} E_{-P_C}(P_C + 1) \quad (26)$$

Where $E_{-P_C}(P_C + 1)$ is the generalized exponential integral, which could further extend as the function of gamma function according to Equation 8.19.1 in NIST Digital Library of Mathematical Functions [22]:

$$E_p(z) = z^{p-1} \Gamma(1-p, z)$$

Thus, the final result could be represented as

$$Pr(I < \min(P_C(I+I))) = e^{P_C+1} (P_C+1)^{-P_C-1} \Gamma(P_C+1, P_C+1) \quad (27)$$

Where $\Gamma(P_C+1, P_C+1)$ is the incomplete gamma function.

The results of $Pr(I < \min(P_C(I+I)))$ for some value of P_C are given below:

$$\begin{aligned} Pr(I < \min(P_C(I+I))) &= \frac{3}{4}, & \text{if } P_C = 1 \\ &= \frac{17}{27}, & \text{if } P_C = 2 \\ &= \frac{71}{128}, & \text{if } P_C = 3 \\ &= \frac{1569}{3125}, & \text{if } P_C = 4 \end{aligned} \quad (28)$$

A.5 The derivation of $Pr(I < \min(\text{multiple path}))$

In A.4, we already derived the solution of the probability that the direct path is faster than any competing paths with 2 links. In this thesis, only the solution of A.1, A.2, and A.4 have been used in the mathematical derivation and program development. However, it is still worth further exploration on the probability for a more general case.

We are happy to point out that, based on the previous derivation, we can find a general solution for the probability that the direct path is faster than multiple independent paths. In Equation (20), we have:

$$Pr(I < \min(P_C(I+I))) = \int_{t=0}^{\infty} f_I(t) (1 - F_{\min(P_C(I+I))}(t)) dt$$

And $F_{\min(P_C(I+I))}(t)$ could be further calculated by using Erlang distribution in Equation (22):

$$F_{\min(P_C(I+I))} = 1 - (e^{-\beta t} + \beta t e^{-\beta t})^{P_C}$$

Similarly, we can follow the same procedure:

$$Pr(I < \min(\text{multiple path})) = \int_{t=0}^{\infty} f_I(t) (1 - F_{\min(\text{multiple path})}(t)) dt \quad (29)$$

In which $F_{\min(\text{multiple path})}(t)$ can be represented as:

$$F_{\min(\text{multiple path})}(t) = 1 - [1 - F_1(t)] [(1 - F_2(t)) \dots [1 - F_n(t)]] \quad (30)$$

Where $F_1(t)$, $F_2(t)$, and $F_n(t)$ are the CDF of different competing paths. Each CDF could be found by using Erlang distribution by substitution $k = \text{number of links}$ in that path and $\lambda = \beta$ to

$$F(x) = 1 - \sum_{n=0}^{k-1} \frac{1}{n!} e^{-\lambda x} (\lambda x)^n$$

Thus, the probability could be presented as the integration of the function:

$$\begin{aligned} Pr(I < \min(\text{multiple path})) &= \\ &\int_{t=0}^{\infty} f_I(t) (1 - [1 - F_1(t)] [(1 - F_2(t)) \dots [1 - F_n(t)]] dt \end{aligned} \quad (31)$$

Appendix B: Nomenclature

B.1 List of Abbreviations

Abbreviation	Definition
IIN	Initial infected node
NN	Neighboring node
TN	Target node
P_3	Graph with path on 3 nodes
$K_{3(ignore\ cycles)}$	Complete tripartite graph without considering cycles of infection
K_3	Complete tripartite graph
K_4	Complete 4-partite graph
$K_{2,P_A} + e$	The complete bipartite graph on two and P_A nodes plus the link e , which connect IIN and TN directly

B.2 List of Notations

Notation	Explanation
G	Graph
G_I	Graph with infection process
A	Adjacency matrix of a graph
N	Number of nodes in graph G
L	Number of links in graph G
P_A	Number of alternative paths
P_C	Number of competing path (alternative path becomes competing path once every link in alternative path propagates infection)
L_I	Number of active link (active link: links which propagate infection) of all alternative paths (same number as infection events)
L_A	Number of links in one alternative path
$R_0(G)$	Basic reproductive number of graph G
$R_0(G_I)$	Basic reproductive number of graph with infection process
$R_0(A)$	Basic reproductive number of node A in graph G
$Pr(A \rightarrow B)$	Probability of infected Node A directly infect susceptible Node B
$Pr(G_I)$	Probability of graph with infection process
$Pr(I)$	Probability of infection event happens between an infected node and a susceptible node
I	The amount of time before the infection event happens
R	The amount of time before the curing event happens
nI	The amount of time required for n consecutive infection events (infection happen through n links)

Appendix C: Codes

C.1 Code in Section 4.1

```
1 % set parameters
2 num_simulations = 100000; % simulation times
3 infection_rate = 1;
4 curing_rate = 1;
5 effect_spr_rate = infection_rate/curing_rate; % effective infection rate
6
7 % set counter
8 A_direct_infections = 0;
9 B_direct_infections = 0;
10
11 for i = 1:num_simulations
12     % simulate time for 3 paths
13     link1 = exprnd(1 / effect_spr_rate);
14     link2 = exprnd(1 / effect_spr_rate);
15     link3 = exprnd(1 / effect_spr_rate);
16     % find which path is quickest
17     if link1 < link2 + link3
18         A_direct_infections = A_direct_infections + 1;
19     end
20     if link2 < link1 + link3
21         B_direct_infections = B_direct_infections + 1;
22     end
23 end
24
25 % find probability
26 A_prob = A_direct_infections / num_simulations;
27 B_prob = B_direct_infections / num_simulations;
28 R0 = A_prob + B_prob;
29 % display result
30 fprintf('R0: %.4f\n', R0);
```

C.2 Code in Section 4.2

```
1 % set parameters
2 num_simulations = 100000; % simulation times
3 infection_rate = 1;
4 curing_rate = 1;
5 effect_spr_rate = infection_rate/curing_rate;% effective infection rate
6
7 % set counter
8 A_direct_infections = 0;
9 B_direct_infections = 0;
10
11 for i = 1:num_simulations
12
13     % simulate time for 5 links
14     link1 = exprnd(1 / effect_spr_rate);
15     link2 = exprnd(1 / effect_spr_rate);
16     link3 = exprnd(1 / effect_spr_rate);
17     link4 = exprnd(1 / effect_spr_rate);
18     link5 = exprnd(1 / effect_spr_rate);
```



```

19
20 % find which path is quickest
21 if (link1<link2+link3) && (link1<link2+link4+link5)
22     A_direct_infections=A_direct_infections+1;
23 end
24
25 if (link2<link1+link3) && (link2<link1+link4+link5)
26     B_direct_infections=B_direct_infections+1;
27 end
28
29 end
30
31 % find probability
32 A_prob = A_direct_infections / num_simulations;
33 B_prob = B_direct_infections / num_simulations;
34 R0 = A_prob + B_prob;
35 % display result
36 fprintf('R0: %.4f\n', R0);

```

C.3 Code in Section 4.3

```

1 clc
2 clear
3 close all
4
5 %% Code implementation of Equation 3.26
6 %Initialize result container
7 condi_prob_results_1 = zeros(1, 50);
8 for i = 1:50
9     P_A = i;
10
11     for j = 0:P_A %The range of P_C is [0,P_A]
12         P_C = j;
13
14         %Calculate the probability of direct infection at different P_C
15         condi_prob = exp(P_C+1) * (P_C+1)^(-P_C-1) * igamma(P_C+1,P_C+1);
16
17     end
18     condi_prob_results_1(i) = condi_prob;
19 end
20
21 %% Code implmentation of Structure of Figure 2, Section 3.2.1
22 % Setting parameters
23 num_simulations = 100000; % Simulation times
24 infection_rate = 1;
25 curing_rate = 1;
26 effect_spr_rate = infection_rate/curing_rate;% effective infection rate
27
28 n=50; %Set the number of additional paths besides direct_path
29 condi_prob_results_2 = zeros(1, n); %Initialize result container
30
31
32 %Initialize the loop of path, the number of paths increases from small
33 for j =1:n
34

```

```

35 % Initialize counter
36 direct_infections = 0;
37 chain_infections = 0;
38
39 %Initialize the simulation loop, each simulation should be independent
40 for i = 1:num_simulations
41     direct_time = exprnd(1 / effect_spr_rate);
42
43     path_time = zeros(1, j);
44     path_time(1)=direct_time;
45     for k = 1:j
46         link_number = 2; %Specify how many links there are in each competing path
47         chain_path_time = exprnd(1 / effect_spr_rate,1,link_number);
48         path_time(k+1) = sum(chain_path_time);
49     end
50
51     if min(path_time)==direct_time
52         direct_infections = direct_infections + 1;
53     else
54         chain_infections = chain_infections + 1;
55     end
56
57     % calculation ratio
58     direct_infection_ratio = direct_infections / num_simulations;
59     chain_infection_ratio = chain_infections / num_simulations;
60     % store results
61     condi_prob_results_2(j)=direct_infection_ratio;
62 end
63 end
64
65 %% Plot the figure
66 figure;
67 plot(conda_prob_results_1, '-o', 'DisplayName', 'Equation 3.26');
68 hold on;
69 plot(conda_prob_results_2, '-s', 'DisplayName', 'Simulation of K_2,P_a +e');
70
71 legend;
72 xlabel('Number of competing path (P_C)');
73 ylabel('Conditional probability under P_C competing paths');
74 ylim([0, 1]);
75
76 title('Comparison of Equation and Simulation');
77 grid on;
78 hold off;

```

D.4 Code in Section 5.1

Network generation: K_3 , K_4 , and $G_{ER}(100, 0.15)$.

```
1 %% Construct the K3 graph
2 G_k3 = graph();
3 % add 3 nodes to graph
4 G_k3 = addnode(G_k3, {'Node1', 'Node2', 'Node3'});
5 % add links for the 3 nodes
6 G_k3 = addedge(G_k3, 'Node1', 'Node2');
7 G_k3 = addedge(G_k3, 'Node2', 'Node3');
8 G_k3 = addedge(G_k3, 'Node3', 'Node1');
9 % plot the graph
10 plot(G_k3)
11
12 %% Construct the K4 graph
13 G_k4 = graph();
14 % add node
15 numNodes = 4;
16 G_k4 = addnode(G_k4, numNodes);
17 % add link
18 probability = 1; % set the prob of each link to be 1
19 for i = 1:numNodes
20     for j = i+1:numNodes
21         if rand < probability
22             G_k4 = addedge(G_k4, i, j);
23         end
24     end
25 end
26 % plot(G_k4)
27
28 %% Construct ER-random graph
29 % This ER random graph has 100 node, and p=0.15 for link between each pair
30 G_ER = graph();
31 % add node
32 numNodes = 100;
33 G_ER = addnode(G_ER, numNodes);
34 % add random links
35 probability = 0.15; % initial the probability of each link
36 for i = 1:numNodes
37     for j = i+1:numNodes
38         if rand < probability
39             G_ER = addedge(G_ER, i, j);
40         end
41     end
42 end
43 plot(G_ER)
```

Implementation of IIN-TN method

```

1 %% Implement IIN-TN emthod
2
3 %% Load the network to be tested here
4 input_network = G_ER;
5 adjacency_matrix = adjacency(input_network);
6 %initial infection rate and curing rate
7 infection_rate = 1;
8 curing_rate = 1;
9
10 %% calculate R0 of each node based on network
11 % Initialize the number of nodes in the network
12 num_nodes = numnodes(input_network);
13 % Initialize a backup of the adjacency matrix
14 adjacency_matrix_copy = adjacency_matrix;
15 % Initialize result storage structure
16 results = struct('IIN', [], 'TN', [], 'Num2linksPath', [], 'Probability', []);
17 % Initialize the index of result
18 index = 1;
19
20 for i = 1:num_nodes
21     % Initialize IIN
22     IIN = i;
23     % Traverse the adjacent nodes of node IIN (find TN)
24     TN = find(adjacency_matrix(IIN, :));
25     for j = 1:length(TN)
26         %Make a backup first and then delete the
27         %link that directly connection two nodes
28         adjacency_matrix_copy_A = adjacency_matrix_copy;
29         adjacency_matrix_copy_A(IIN,TN(j)) =0;
30         adjacency_matrix_copy_A(TN(j),IIN) =0;
31
32         %look for the shortest path to connect the two and record the number of 2linksP
33         ath
34         counter = 0;
35         while true
36             shortest_path = shortestpath(graph(adjacency_matrix_copy_A), IIN, TN(j));
37             if length(shortest_path) == 3
38                 adjacency_matrix_copy_A = Edge_remove(adjacency_matrix_copy_A,shortest_pa
39 th);
40                 counter = counter + 1;
41             else
42                 break
43             end
44         end
45         results(index).IIN = IIN;
46         results(index).TN = TN(j);
47         results(index).Num2linksPath = counter;
48         index = index + 1;
49     end
50 end
51
52 % Calculate the probability of each IIN-TN
53 for i = 1:length(results)
54     num_2linkpaths = results(i).Num2linksPath;
55     results(i).Probability = Prob_IIN_TN(infection_rate,curing_rate,num_2linkpaths);
56 end

```

```

57
58 % Calculate R0 for each node
59 results_node = struct('Node', [], 'R0', []); %initial
60 for i = 1:num_nodes
61     results_node(i).Node = i;
62     results_node(i).R0 = 0;
63 end
64
65 for i = 1:length(results)
66     for j = 1:length(results_node)
67         if results(i).IIN == results_node(j).Node
68             results_node(j).R0 = results_node(j).R0 + results(i).Probability;
69         end
70     end
71 end
72
73 %% Calculate the R0 of the network based on results_node and draw the results
74
75 % find R0 for the whole graph by using R0 for each node
76 total_sum = 0;
77 R0_list = zeros(1, num_nodes);
78 for i = 1:length(results_node)
79     total_sum = total_sum + results_node(i).R0;
80     R0_list(i)=results_node(i).R0;
81 end
82 R0_network = total_sum/num_nodes;
83
84 %print out the result
85 fprintf('The R0 of this network by IIN-TN method is:')
86 disp(R0_network)
87
88 % find R0 by definition
89 eigenvalues = eig(adjacency_matrix);
90 max_eigenvalue = max(eigenvalues);
91 R0_def = (infection_rate / curing_rate)* max_eigenvalue;
92 fprintf('The R0 of this network by definition:')
93 disp(R0_def)
94
95
96 figure;
97 subplot(2, 1, 1);
98 plot(input_network)
99 title('Network');
100
101 subplot(2, 1, 2);
102 plot_limit = ceil( max(R0_list) );
103 edges = linspace(0, plot_limit, plot_limit*4+1 );
104 histogram(R0_list, edges, 'Normalization', 'probability');
105 title('Distribution of R0 Values');
106 xlabel('R0 values');
107 ylabel('Percentage of the R0 values');
108
109 % Add vertical bars to highlight the values of R0 and R0_def
110 hold on
111 [counts, ~] = histcounts(R0_list, edges); %Find the number of each hist
112 max_count = max(counts);
113 line_position = max_count / num_nodes + 0.05; %Find the highest reached position of hist

```

```

114 % Draw a vertical line of 1 in color green
115 line([1 1], [0 line_position], 'Color', 'g', 'LineWidth', 1);
116 % Draw a vertical line of R0 in color red
117 line([R0_network R0_network], [0 line_position], 'Color', 'r', 'LineWidth', 1);
118 % Draw a vertical line of R0_def in color blue
119 line([R0_def R0_def], [0 line_position], 'Color', 'b', 'LineWidth', 1);
120
121 % Edit legend
122 legend('', '1', sprintf('R0      = %.3f', R0_network), sprintf('R0_{(def)} = %.3f', R0_def))
123 hold off

```

Implementation of Star method(the part should run after the above part in C.4)

```

1 %% Repeat the process with star method
2 % Calculate R0 of each node (star method))
3 results_node_star_method = struct('Node', [], 'R0', []); % Initialize result container
4 for i = 1:num_nodes
5     results_node_star_method(i).Node = i;
6     results_node_star_method(i).R0 = 0;
7 end
8
9 for i = 1:length(results_IIN_TN)
10     for j = 1:length(results_node_star_method)
11         if results_IIN_TN(i).IIN == results_node_star_method(j).Node
12             results_node_star_method(j).R0 = results_node_star_method(j).R0 + ...
13                 infection_rate/(infection_rate + curing_rate);
14         end
15     end
16 end
17
18 total_sum_star_method = 0;
19 R0_list_star_method = zeros(1, num_nodes);
20 for i = 1:length(results_node_star_method)
21     total_sum_star_method = total_sum_star_method + results_node_star_method(i).R0;
22     R0_list_star_method(i)=results_node_star_method(i).R0;
23 end
24 R0_network_star_method = total_sum_star_method/num_nodes;
25
26 %print out the result
27 fprintf('The R0 of this network by Star method is:')
28 disp(R0_network_star_method)
29
30 figure;
31 % plot the result
32 plot_limit = ceil( max(max(R0_list_star_method) , max(R0_list)) );
33 edges = linspace(0, plot_limit, plot_limit*4+1 );
34 histogram(R0_list_star_method, edges, 'Normalization', 'probability');
35 title('Distribution of R0 Values (Star method)');
36 xlabel('R0 values');
37 ylabel('Percentage');
38 hold on
39 [counts, ~] = histcounts(R0_list_star_method, edges);
40 max_count = max(counts);
41 line_position = max_count / num_nodes + 0.05;
42 line([1 1], [0 line_position], 'Color', 'g', 'LineWidth', 1);
43 line([R0_network_star_method R0_network_star_method], [0 line_position], 'Color', 'r', 'LineWid

```

```

44 th', 1);
45 line([R0_def R0_def], [0 line_position], 'Color', 'b', 'LineWidth', 1);
46 legend('1',sprintf('R0          = %.3f', R0_network_star_method),sprintf('R0_{(def)} = %.3f', R0
47 _def))
48 hold off

```

Functions used in above

1. Prob_IIN_TN.m

```

1 function [result] = Prob_IIN_TN(infection_rate, curing_rate, num_paths)
2 % This function is the Equation 3.27 in Thesis paper
3 % infection_rate, curing_rate, num_paths are the parameters that should
4 % be given
5
6 P_A = num_paths;
7 if P_A == 0
8     result = 1;
9 else
10    result = 0;
11    for j = 0:P_A % the range of P_C is [0,P_A]
12        P_C = j;
13        prob_sum = 0;
14        for k = 2*P_C:(P_A+P_C) %the range of L_I is [2*P_C,P_A+P_C]
15            L_I = k;
16            % check the output
17            fprintf('%s = %g,\n', 'P_A', P_A);
18            fprintf('%s = %g,\n', 'P_C', P_C);
19            fprintf('%s = %g,\n', 'L_I', L_I);
20
21            % Count the number of each infection graph
22            % f: This variable corresponds to f (P_A,P_C,L_I) in the formula
23            if L_I-2*P_C-1<0
24                f = nchoosek(P_A, P_C);
25                % fprintf('%s = %g,\n', 'f', f);
26            else
27                molecular = 1;
28                for k = 0:(L_I-2*P_C-1)
29                    comb = nchoosek( 2*P_A-2*P_C-2*k , 1);
30                    molecular = molecular*comb;
31                end
32                denominator = factorial(L_I-2*P_C);
33                f = nchoosek(P_A, P_C)*molecular/denominator;
34                % fprintf('%s = %g,\n', 'f', f);
35            end
36
37            % Calculate the probability corresponding to each f
38            prob_molecular = (infection_rate^L_I) * ( curing_rate^(2*P_A-L_I) );
39            prob_denominator = (infection_rate+curing_rate)^(2*P_A);
40            prob = prob_molecular/prob_denominator;
41            % fprintf('%s = %g,\n', 'probability', prob);
42
43            %Calculate the probability sum of all situations where L_I is in range:[2*P
44            _C,P_A+P_C]
45            prob_sum = prob_sum + prob*f;
46        end
47        fprintf('%s = %g,\n', 'prob_sum', prob_sum);

```

```

48     direct_path = infection_rate /(infection_rate + curing_rate);
49     total_prob = direct_path*prob_sum;
50
51     %Calculate the probability of direct infection at different P_C
52     condi_prob = exp(P_C+1) * (P_C+1)^(-P_C-1) * igamma(P_C+1,P_C+1);
53
54     %Calculate the final result
55     result = condi_prob*total_prob + result;
56 end
57 end
58 end

```

2. Edge_remove.m

```

1 function [new_adjacency_matrix] = Edge_remove(adjacency_matrix,shortest_path)
2 % delete the corresponding links in the
3 % adjacency_matrix according to the shortest path
4
5 num_edges_in_path = length(shortest_path) - 1;
6 for i = 1:num_edges_in_path
7     edge_to_remove = [shortest_path(i), shortest_path(i+1)];
8     adjacency_matrix(edge_to_remove(1), edge_to_remove(2))=0;
9     adjacency_matrix(edge_to_remove(2), edge_to_remove(1))=0;
10 end
11 new_adjacency_matrix = adjacency_matrix;
12 end

```

D.5 Code in Section 5.2.1

Data preprocessing

```

1 %% Preprocessing (sorting) of realHumanContactMIT
2 sortedData = sortrows(realHumanContactMIT, 2);
3 sortedData = sortrows(sortedData, 1);
4 G_MIT = graph();
5 % Set the length
6 transmit_index = length(sortedData.id1)-1;
7 % Establish a point-to-point connection
8 for i = 1:transmit_index
9     disp(i)
10    node1 = sortedData.id1(i);
11    node2 = sortedData.id2(i);
12    if node2 ~= sortedData.id2(i+1)
13        G_MIT = addedge(G_MIT, node1, node2);
14        plot(G_MIT)
15    end
16
17    if i == transmit_index
18        G_MIT = addedge(G_MIT, node1, node2);
19    end
20 end
21 plot(G_MIT)

```