[redacted]

Introduction

ODE Epidemic Models

Ordinary differential equations (ODE) models of infectious disease spread are common in mathematical epidemiology and are often used to describe and predict the evolution of infection in a population. The most common simple ODE model is the Susceptible-Infectious-Recovered (SIR) model. Consider a population of N individuals; there is no birth, death, or migration, so N remains constant. Susceptible individuals make contact with others at a rate proportional to the fraction of infectious persons in the population (proportional homogeneous mixing). Infectious individuals recover with constant rate γ. Recovered individuals have acquired immunity to infection and are not susceptible to reinfection. The ODE system describing this flow is

$$\begin{aligned} S' &= -\beta S \frac{I}{N} && (1) \\ I' &= \beta SI - \gamma I && (2) \\ R' &= \gamma I && (3) \end{aligned}$$

where $S(t)$, $I(t)$, and $R(t)$ denote the number of susceptible, infectious, and recovered individuals in the population at time t. The initial conditions usually applied are $S(0) = N - 1, I(0) = 1, R(0) = 0$. Note that

$$S(t) + I(t) + R(t) = N$$

at all times, and that we may use this constraint to effectively reduce the dimension of our problem by one (any of the three quantities S, I, or R may be calculated by subtracting the other two from N).

A major problem with these models is that they assume some form of homogeneous mixing: each person has the same probability of meeting any other person in the population. A second issue is the implicit assumption of large population size in ODE models. Effectively, the concept of "individual" is lost in the ODE framework.

Network Epidemic Models

Network or graph models are one way of explicitly modeling individuals in a population. Each person is represented by a node, and each edge of the graph represents a potentially disease-transmitting contacts between two people. Many network-type models of infectious disease spread exist in the current literature.

Of the analytic models, however, none exist which track infection on a network in continuous time.

In this project I compare ODE and network SIR models with respect to both within-epidemic dynamics, *i.e.* continuous-time disease dynamics, as well as duration, prevalence, and final size variables. For the purpose of containing the scope of this work, the SIR model of disease history was assumed; three sets of parameter values are investigated.

1 Project Outline

The outset goal of this project was to combine the theory and tools acquired in CSC 546 (University of Victoria, [redacted]) with my experience and knowledge in mathematical epidemiology to simulate the continuous-time spread of infectious disease on a network of individuals. There were four major areas of investigation undertaken to reach this goal:

1. Continuous-time stochastic simulation of ODE model (Eq. 1-3).
2. Continuous-time stochastic simulation of network model.
3. Comparison of ODE and network models.
4. Comparison of parameter sets.

As a novice programmer, the practical knowledge gained through the coding and implementing of the first two items listed above represents one major project outcome. The analysis of simulation experiments form the second major set of results.

Simulation Goals

Comparison of ODE and network models, under similar parameter configurations were performed to identify how ignoring individual differences in contact rates *i.e.* assuming homogeneous mixing, may affect epidemic outcomes or predictions. Different parameter sets were used to illustrate epidemic dynamics. The output variables collected for analysis are described in Table 1.

Simulation Parameters

The simulation parameters of interest are given in Table 2. A transmission parameter (β or T) and a recovery rate (γ) are specified for both network and ODE models. The degree of an individual (k) and the distribution of the generation interval (Y) only apply to network models.

It should be noted that the ODE transmission parameter (β) has multiple interpretations in the literature. For the purposes of drawing comparisons to the network model transmission parameter (T), the ODE analogue β is defined

Table 1: Simulation output variables. Items marked with (*) are applicable for network models only. Time series data was collected for items marked with (**).

Variable	Description
Incidence (**)	Average rate of new infections
Generation interval (*)	Time between infection of a node and transmission of infection to neighbouring node
Prevalence (**)	Proportion of population that is concurrently infected at any point during an outbreak
Final Size	Proportion of population that ever becomes infected
Duration	Elapsed time between the first infection and the last recovery in the population

Table 2: Simulation parameters. "Contact rate" refers to the rate of potentially disease transmitting contacts between susceptible and infectious individuals. Where applicable, ODE and network equivalents are given; some quantities pertain only to one type of model.

Parameter	Description	ODE	Network
Average Degree	Average per-node number of neighbours	-	$< k >$
Transmission Parameter	Per-contact probability of transmission (T) or rate of contact (β)	β	T
Mean Generation Interval	Time between infection of node and subsequent infection of neighbour	-	$\bar{Y}$
Recovery Rate	Inverse of mean infectious period (infectious period $\sim$ Exponential)	γ	γ

to be the per-susceptible rate of potentially disease-causing contacts with other persons in the population. Equations 1 - 3 model homogeneous mixing that is proportional to the infectious fraction of the population. The term βS gives the rate that the susceptible proportion experiences potentially disease-transmitting contacts. A portion of I/N of these are expected to be with infectious individuals, thus $\beta SI/N$ gives the rate of flow from the susceptible class to the infectious class.

The network transmission parameter (T) is defined as the per-contact per-time unit probability of transmission. In a discrete-time framework, the time for transmission to each neighbour of a newly infected node is then a geometrically distributed random variable with mean $1/T$. This quantity is called the "generation interval", and denoted by Y. If we suppose that $Y \sim \text{Exp}[\lambda]$, then

$$\text{Prob}(Y \leq 1) = 1 - \exp(-\lambda),$$

and from the definition of T we also see that

$$\text{Prob}(Y \leq 1) = T.$$

Substituting,

$$\begin{aligned} 1 - \exp(-\lambda) &= T \\ \exp(-\lambda) &= 1 - T \\ \lambda &= -ln[1 - T] \end{aligned}$$

Thus in the limit of continuous time, we have the generation interval Y with exponential distribution and mean $-1/ln(1-T)$:

$$Y \sim \text{Exp}[-ln(1 - T)]$$

In both the ODE and network model, the effective rates of contact between infectious persons and susceptible persons depends on the number of infectious persons. For the purposes of comparing ODE and network models, T and $< k >$ are set apriori and the value of β is given by

$$\beta = T \times < k >$$

Infectious persons are assumed to recover at a constant rate (γ). This implicitly assumes an exponential distribution for the infectious period in the ODE model (1)-(3); for comparison purposes, the network model also uses an exponentially distributed infectious period. Other distributions *e.g.* gamma or normal, are worth exploring in future work; the ODE framework could be changed to a delay-differential equation (DDE) paradigm involving integral equations to accommodate these types of modifications and make comparison with network models possible.

2 Methodology

All coding was done in Java under the Eclipse SDK v.3.6.2 environment; R v.2.11.1 was used for statistical analysis and production of figures. I would like to acknowledge [redacted] for sharing his Java source code that provided some insight into the problems I was undertaking, as well as the event list and queueing source code provided in class (CSC 546, [redacted]).

The project tasks may be broken down into four modules:

1. Build a graph of $N = 5000$ nodes with a specified degree distribution.
2. Simulate an SIR epidemic on this graph.
3. Simulate a comparable ODE model of an SIR epidemic.
4. Analyze network and ODE output; compare network and ODE model behavior under different parameter sets.

Graph Construction

Node and network classes were constructed. Nodes have an index, and infection status, and a list of neighbouring nodes. Networks are composed of a list of nodes; the links of the network are recorded in each nodes neighbour list. Due to time constraints, a binomial degree distribution was used for each graph. The pseudo-code for binomial graph construction is as follows:

```
<k> = average degree
binomial parameter n = population size
binomial parameter p = <k>/n

for each node i
     for each neighbour j of i
      r ~ Unif(0,1)
     if r < p/2 AND i != j or are already neighbours
     add j to neighbour list of i
     add i to neighbour list of j
```

Note that $p/2$ rather than p is used in the algorithm. This is because under the binomial model, p is the probability that node i is connected to j. The algorithm checks both i and j, thus each must be checked with probability $p/2$ to preserve the model assumptions. This was verified in the testing stage.

I will also acknowledge here that a significant amount of code was developed to implement the configuration model (CM) method of graph construction. The original intent of this project was to include simulation SIR epidemics on networks with exponential and other degree distributions. Unfortunately I did not successfully obtain results that could be presented in this report. The pseudo-code for the CM method of graph construction is as follows:

```
<k> = average degree
pk = probability of degree k (under chosen distribution)
n = population size
degree sequence = new List

for each node i
      draw degree d from chosen distribution
      assign d ‘‘stubs’’ to node i

if sum(degree sequence) is an odd number
     randomly choose a node and increase it’s degree by 1
     (ensures a viable number of pairings to be made)

for stub on node i
     search population for suitable node j with open stub
     (suitable nodes are those not identical
```

```
     to i, nor neighbours with i already)
     join stubs to create link between i and j

     if no suitable matches exist
     choose a suitable node k
     cut one link of node k and join to node i
```

Network SIR Simulation

The algorithm used to simulate an SIR epidemic on a constructed network of nodes is as follows:

```
T = transmission probability
1/g = mean infectious period
FEL = future event list

Initialize epidemic:
choose a random node and infect it (index case)
draw a recovery time for this node; add to FEL
for each neighbour j of i
     draw a random generation time (mean = -1/ln(1-T))
     if generation time < recovery time
          add infection of j to FEL

Work through FEL:
while length of FEL > 0
     if next event is infection of node k
          if k is susceptible
               infect node k
               schedule recovery of node k
               schedule infection of neighbours as for index case

          else discard event and continue
     if next event is recovery of node k
          recover node k
```

Metrics such as incidence and prevalence are collected and stored with a time stamp for use in both calculation of within-replication averages, and as time-series data. Note that multiple transmissions may be sent to a particular node from it's infectious neighbours, but that only the first (chronological) infection is counted. This particular infection event is used to calculate the "effective generation time" (Y_{eff}).

Stochastic ODE Simulation

The ODE epidemic model of Equations 1-3 was simulated using the Gillespie algorithm:

```
N = population size
B = transmission parameter
1/g = mean infectious period
S = number susceptible individuals
I = number infected individuals

S = N-1
I = 1

while I > 0
     infection rate = B*S*I/N
     recovery rate = g*I
     total rate = infection rate + recovery rate
     draw next event time t ~ Exp(total rate)
     draw r ~ Unif(0, 1)
     if r < infection rate/total rate
         S--
         I++
     else
         I--
```

As with the network SIR simulation, incidence and prevalence time series data are collected, as well as per-run calculations such as maximum prevalence and epidemic size.

Simulation Setup

Each experiment was specified by the model (ODE or network) and parameter values used. The three sets of parameter values are listed in Table 3. Every experiment consisted of five runs; the same five random seeds were used in every experiment; the total number of 5-run experiments is six. In the case of the network model, a single random seed was used to build the graph used in each experiment. As a result, a single graph configuration was used for all runs and the index case (initial infected individual) was also identical in every network run. In an ideal experimental environment, each experiment would be run multiple times over a large ensemble of graphs and index cases, so as to remove any artifacts caused by these factors.

Table 3: Parameter value sets used in simulations. Five runs each of the ODE and network models were run under each set of parameter values. Mean generation interval ($\bar{Y}$) is calculated as $[-ln(1-T)]^{-1}$.

Set	T	γ^{-1}	Y	$< k >$
A	0.25	4	3.47	5
B	0.1	4	9.49	5
C	0.1	8	9.49	5

Data Collection and Statistics

For each simulation run output variables (within-replication analysis) were calculated and recorded, and time series of incidence and prevalence were written to file. Sample mean, sample standard deviaion, and 95% confidence intervals are computed for each variable of interest (between-replication analysis). In each network experiment, there was (randomly) one run which did not result in any infections beyond the index case - the generation interval does not exist for these runs, and they are omitted from the analysis (effectively, sample size $n = 4$ for each calculation involving generation interval). ODE and network models are compared under each parameter set using the variance reduction technique of estimating the mean difference between models (model comparison).

Time series data are used here for visualization purposes and discussion points. Within a project of larger scope it would be possible to run a very large number of simulations and fit these data to an ODE model (or various ODE models), and compare the parameter estimates $\hat{\beta}$ and $\hat{\gamma}$ to the true values used in the simulation runs.

3 Results and Analysis

Individual Run Reports

Each configuration was run for five different seeds. Parameter configurations used, and real-valued output for each ODE and network simulation run are listed in Tables 4 and 5; time series output is shown in Figures 1-6.

3.1 Grand Statistics

Between replication statistics are given in Table 6; test statistics for model comparisons (between-experiment) are listed in Table 7.

For each parameter set tested, no significant differences were found between the ODE and network models observed maximum prevalence, epidemic size, and epidemic duration (Table 7). This may be related to the vary large variances observed: in most cases, the sample variance was larger than the sample mean. These were likely caused by the fact that some proportion of runs in each experiment did not result in any outbreak while the other runs produced a large outbreak of infection (see Figures 1-6 and Tables 4-5 for individual run output). The small number of sample runs produced is a third factor in producing large variance.

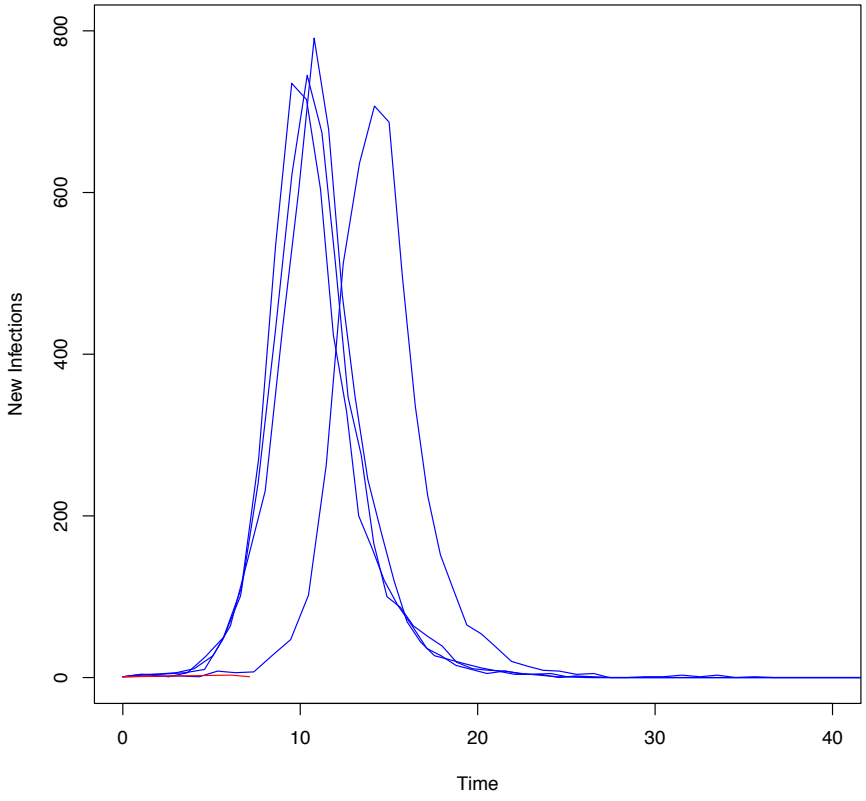

Figure 1: Comparison of simulated ODE (red) and network (blue) incidence curves. Parameter set A: transmission probability $T = 0.25$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 4$. No ODE run produced an outbreak. Three network runs had similar large outbreaks, one network outbreak was delayed, and one did not grow.

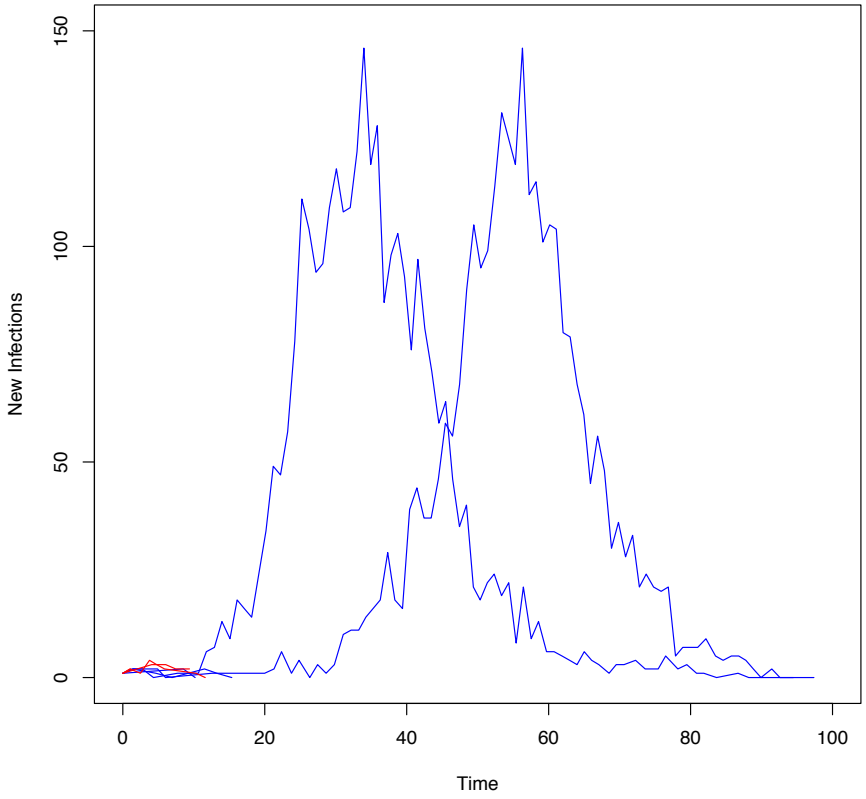

Figure 2: Comparison of simulated ODE (red) and network (blue) incidence curves. Parameter set B: transmission probability $T = 0.1$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 4$. Two network runs resulted in outbreaks of disease; one of these had a delayed start time. No other simulations produced visible outbreaks.

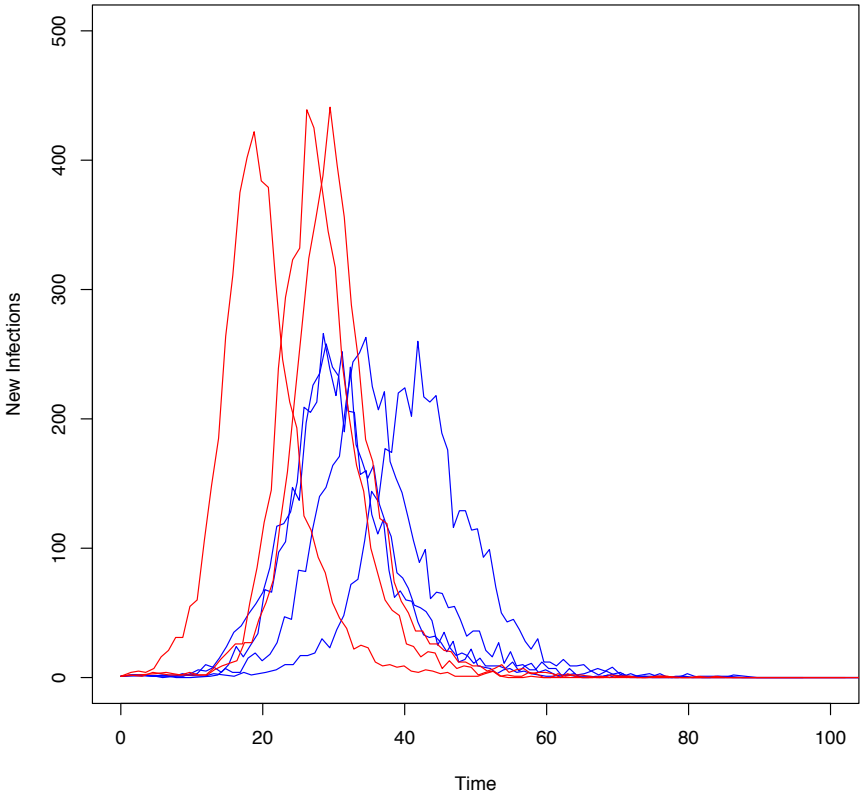

Figure 3: Comparison of simulated ODE (red) and network (blue) incidence curves. Parameter set C: transmission probability $T = 0.1$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 8$. Three of five ODE runs and four of five network runs produced visible outbreaks. The ODE outbreaks were happened earlier and faster than the network outbreaks.

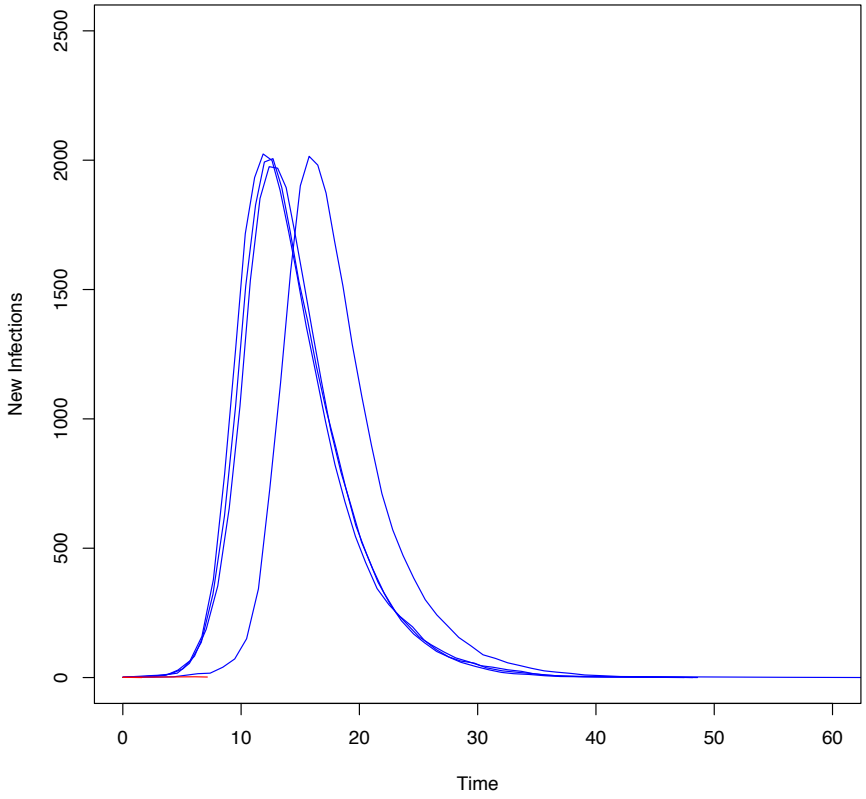

Figure 4: Comparison of simulated ODE (red) and network (blue) prevalence curves. Parameter set A: transmission probability $T = 0.25$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 4$. No ODE run produced an outbreak. Three network runs had similar large outbreaks, one network outbreak was delayed, and one did not grow.

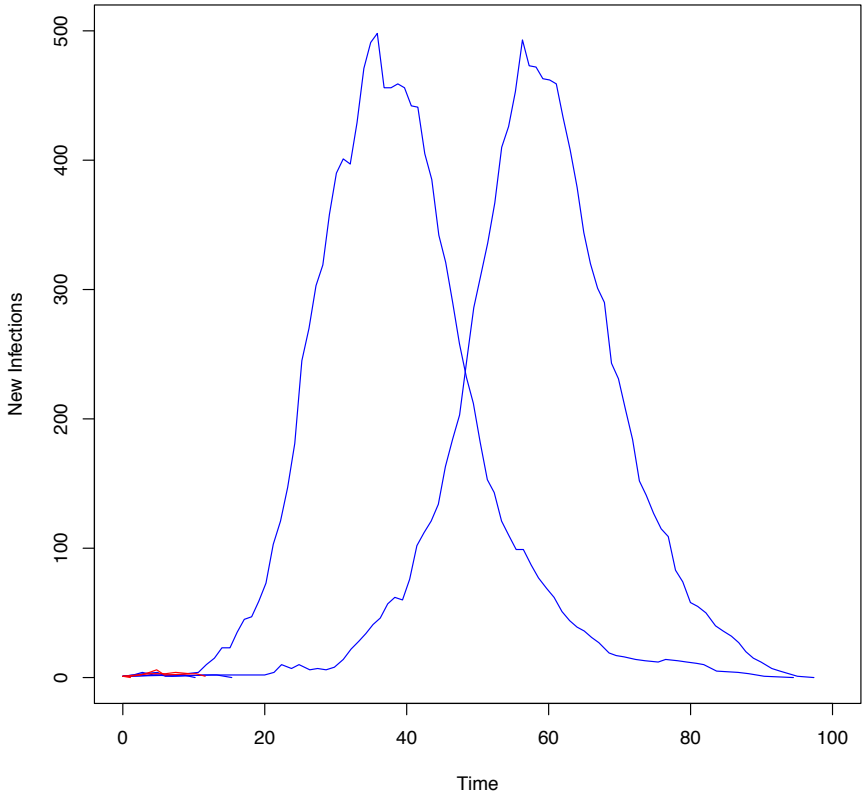

Figure 5: Comparison of simulated ODE (red) and network (blue) prevalence curves. Parameter set B: transmission probability $T = 0.1$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 4$. Two network runs resulted in outbreaks of disease; one of these had a delayed start time. No other simulations produced visible outbreaks.

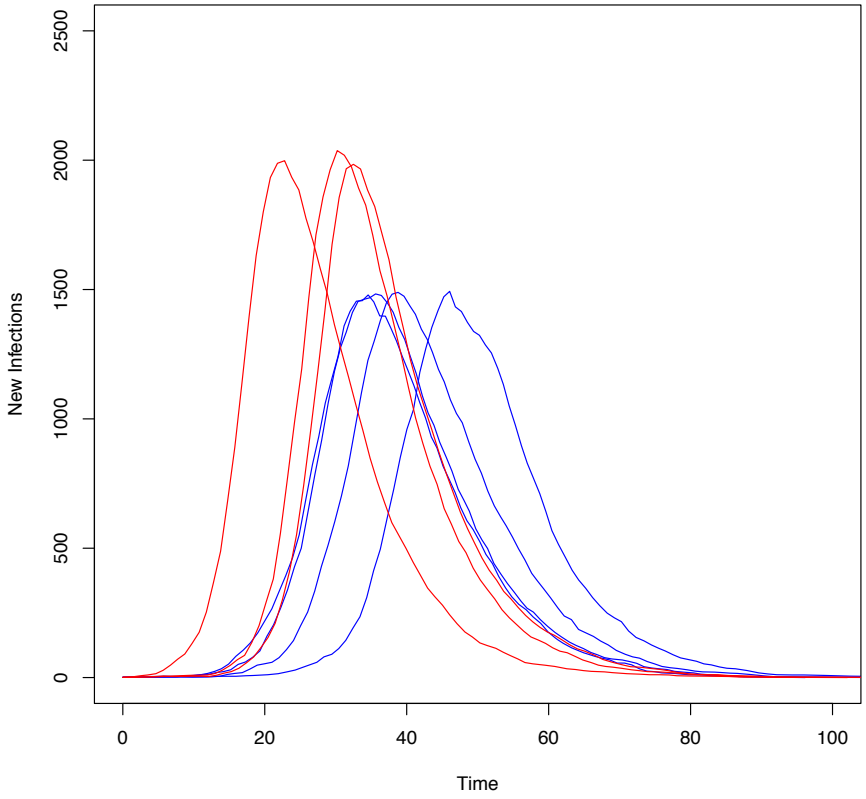

Figure 6: Comparison of simulated ODE (red) and network (blue) prevalence curves. Parameter set C: transmission probability $T = 0.1$; average degree $< k >= 5$; mean infectious period $\gamma^{-1} = 8$. Three of five ODE runs and four of five network runs produced visible outbreaks. The ODE outbreaks were larger and happened earlier than the network outbreaks.

Table 4: Simulation results of stochastic ODE model. Parameter sets are described in Table 2. Final size (s), maximum prevalence (I_{max}), and epidemic duration (T_{tot}) are given for each of five independent runs per parameter configuration.

Parameter Set	Run	s	I_{max}	T_{tot}
A	1	4958	0.4782	43.65
	2	4967	0.4864	51.2933
	3	1	2E-04	0.2088
	4	4963	0.474	45.52
	5	1	2E-04	0.209
B	1	13	0.0014	11.64
	2	10	0.0010	10.31
	3	1	2E-04	0.417
	4	2	4E-04	1.073
	5	1	2E-04	0.418
C	1	4889	0.4002	114.927
	2	4897	0.4108	113.26
	3	1	2E-04	0.501
	4	4886	0.4016	102.09
	5	1	2E-04	0.502

4 Discussion and Conclusion

As expected, epidemic size increased with increased transmission parameter values (Set A versus Set B; Tables 4,5; Figs. 1, 2, 4, 5). Increasing the mean infectious period also increased epidemic size and speed (Set C versus Set B; Tables 4,5; Figs 2, 3, 5, 6). Clearly there is a large divide between those epidemics that "take off" and infect a large proportion of the population, and those that die off after infecting only a few individuals. The ODE simulations seemed more prone to dying out than the network simulations.

The large sample variance in each experiment is likely the root of the statistical insignificance for all mean difference tests. If we omit those epidemics that died out and consider only the visible peaks in each figure presented, seem to be several potential differences. Consider parameter set C, which featured the longer mean infectious period ($\gamma^{-1} = 8$), but the smaller transmission value ($T = 0.1$). Figure 3 shows that the corresponding ODE outbreaks occur sooner and progress faster (larger incidence) than the network outbreaks. Each model type appears to have one incidence peak which is simply shifted in time between simulation runs. The prevalence curves for this parameter set (Figure 6) also reflect this shifted peak feature, and the tendency of the ODE outbreak to occur sooner and with a steeper slope than the network epidemic.

Table 5: Simulation results of network with binomial degree distribution. Parameter sets are described in Table 2. Final size (s), maximum prevalence (I_{max}), effective mean generation interval ($\widehat{Y_{eff}}$) and epidemic duration (T_{tot}) are given for each of five independent runs per parameter configuration.

Parameter Set	Run	s	I_{max}	$\widehat{Y_{eff}}$	T_{tot}
A	1	4553	0.4062	1.27	63.81
	2	4554	0.4094	1.27	47.47
	3	4585	0.4066	1.29	48.186
	4	1	2E-04	NA	0.3879
	5	4570	0.4052	1.27	48.59
B	1	2	4E-04	8.52	10.19
	2	2935	0.1022	2.59	94.48
	3	2851	0.0994	2.71	97.37
	4	1	2E-04	NA	0.38
	5	6	4E-04	3.53	15.53
C	1	4277	0.2994	3.413	135.31
	2	4305	0.2992	3.34	104.44
	3	4263	0.2984	3.37	109.22
	4	1	2E-04	NA	0.7758
	5	4268	0.2964	3.222	96.145

Figure 2 suggests that under reduced transmission ($T = 0.1$) and mean infectious period ($\gamma^{-1} = 4$) the network model is more likely to show an outbreak. However, due to the small sample size ($n = 5$), it is impossible to tell if this is by chance. A large number of simulations need to be run to compare the probability of an epidemic dying out under the ODE model assumptions as compared to the network model framework. One "run" might consist of several individual epidemic simulations, from which a mean probability of extinction is calculated. The between-run grand mean of a large number of ODE and network experiments would then be compared.

4.1 Learning Experience

In the course of this project I worked with or attempted to work with several new tools. Overall, I learned a great deal about functionally performing simulations, although there are some pieces of work that remain unfinished for me.

OMNet++ proved the biggest challenge to me. I was able to work through the tutorials and learn the steps in building a simulation under this environment. The graphical tools were a benefit to my lack of knowledge in program structure, although admittedly I might have understood more if forced to hard code each component of the .ned file myself! Ultimately my problem with OM-

Table 6: Average simulation output (mean ± standard deviation) for ODE and network (NW) models. Parameter sets are described in Table 2. Average final size (s), maximum prevalence (I_{max}), effective mean generation interval ($\widehat{Y_{eff}}$) and epidemic duration (T_{tot}) are computed. Each experiment consisted of five runs; sample statistics are computed using $n = 5$ except in the case of $\widehat{Y_{eff}}$, where $n = 4$ valid observations were available in each network experiment.

Parameter Set	Model	s	$\widehat{Y_{eff}}$	I_{max}	T_{tot}
A	ODE	2978 (±2717.62)	-	0.288 (± 0.263)	28.19 (± 25.69)
A	NW	3652.6 (± 2041.35)	1.275 (± 0.01)	0.326 (± 0.182)	41.69 (± 24.07)
B	ODE	5.4 (±5.68)	-	6.4E-04 (±5.3E-04)	4.77 (±5.69)
B	NW	1159 (±1583.2)	4.338 (±2.82)	0.041 (±0.055)	43.59 (±48.09)
C	ODE	2934.8 (±2678.2)	-	0.2426 (±0.221)	66.26 (±60.23)
C	NW	3422.8 (±1912.9)	3.34 (±0.082)	0.239 (±0.133)	89.18 (±51.54)

Net++ was not understanding it's capabilities. I worked very hard to figure out how to setup a network of a relatively large number of nodes (e.g. N=1000+) without adding each one individually. I could not do this, and I certainly could not figure out how to set the degree distribution for such a network. I still do not know if it is possible. Before this project I had not worked with any specialized simulation environments. They do exist for mathematical epidemiology, and I may investigate them in the future. It will be important to understand the intended purpose and capabilities of the application first, however.

My programming skills and computer saavy improved over the course of this project (and during the course of CSC 546). I was not able to utilize shared code ([redacted] Java graph building package) because I was running out of time due to overcommitting to OMNet++, and because I did not have the immediate skills or knowledge to integrate new items into my working environment. Given more time I believe I would have worked through this.

The greatest benefit to me was the formalization of the discrete event-time simulation, in particular, the future event list. The continuous-time network SIR that I was able to implement using the sample code from class and my understanding of disease models taught me a lot about project design. It is also one of the better pieces I have written in terms of adaptability and modularity. Model verification/validation is an area I need to work on - rather than being

happy with code that simply compiles (an accomplishment in itself for a mathematician) I need to carefully check what it is doing. This will allow me more quality time to design and run experiments.

4.2 Insights and Improvements

In theoretical network epidemic modeling, dynamical models (essentially, those that output time-series data) are practically non-existent. The majority of work is based on percolation theory, and accounts for only "end" results, such as final size, epidemic duration, and probability of an epidemic. I know of one recent dynamical model, but it is a so-called generation model. In this model of disease, the infectious period is exactly one time step, and individuals transmit infection only at the end of this time, and then immediately recover. Many people are working in network dynamics and dynamic infectious disease modeling, but simulations provide a valuable insight to what is occurring while the epidemic is actually occurring. In this sense, the main benefit of my project design was the capacity to collect time-series data.

The most obvious improvement to my approach would be to fulfill the original goal of producing networks with varying degree distributions. The average degree parameter was not explored in this report, as the comparison of models and correct formulation of both model and analyses was my primary goal in the end. The framework and ideas explored in this project are a solid basis for simulation of ODE and network models. Comparison of models, parameter estimation, exploration of epidemic dynamics, and calculation of health-care costs or benefits are all within the scope of applications.

Table 7: Comparison of ODE and network (NW) model output. Mean difference ($\bar{D}$), standard deviation (s_D), 95% confidence interval, test statistic (T_4) and corresponding p-value (p) were computed for each variable under each model and parameter set. Final size (s), maximum prevalence (I_{max}),and epidemic duration (T_{tot}). Parameter sets are described in Table 2.

Parameter Set (& Model)	Variable	$\bar{D}$	s_D	95% C.I	T_4	p-value
A (ODE-NW)	s	-674.6	39.33	(-5663.3, 4313.1)	-0.375	0.726
	I_{max}	-0.0377	0.373	(-0.501, 0.426)	-0.226	0.833
	T_{tot}	-13.51	39.33	(-62.35, 35.32)	-0.768	0.485
B (ODE-NW)	s	-1153.6	1583.06	(-3119.2, 812.0)	-1.63	0.179
	I_{max}	-0.040	0.055	(-0.108, 0.028)	-1.62	0.181
	T_{tot}	-38.82	47.91	(-98.30, 20.67)	-1.812	0.144
C (ODE-NW)	s	-488.0	3865.54	(-5287.70, 4311.70)	-0.282	0.792
	I_{max}	0.0039	0.300	(-0.369, 0.377)	0.0289	0.978
	T_{tot}	-22.92	85.29	(-128.82, 82.98)	-0.6009	0.580