Probabilistic Model Checking of Disease Spread and Prevention

Samuel Huang

Scholarly Paper for the Degree of Masters in Science

Abstract

This work investigates the benefits of model checking upon contact networks, a model of disease transmission through a population. Many such models are probabilistic in design, and thus a model checker framework that supports randomness is important. We demonstrate how to characterize a disease as it is spread in terms of the portions of the population that it affects, as well as how to evaluate and explore preventative and controlled measures to limit the disease's effects. For the purposes of this work, we focus upon several different vaccination strategies. We also explore the sensitivity of our framework that involves using a probabilistic model checker has to changes in the underlying disease transmission model; we show that many desirable properties and traits that one might want to include are easily representable using our approaches.

1 Introduction

Advances in epidemiology, the study of factors affecting health and illness of populations, can be directly beneficial to human society. A large subfield of epidemiological research focuses upon the prevention and control of disease propagation through a population. In particular, they are concerned with:

- Outbreak investigation
- Modeling
- Data collection and analysis

Developing an accurate and appropriate model for the purpose of investigating and predicting disease outbreaks is actively being pursued. If a disease were predicted to become rampant in the near future, one would like to determine the proper cause of action to avoid or otherwise reduce the impact the disease would have upon the population under

study. If one is studying records of an already passed disease's progression through a population, a natural concern may be as to how/where the disease originated.

In this work we consider the usefulness of applying existing model checking based approaches to problems posed by epidemiology, specifically those focusing on disease prevention and control. Traditional model checking algorithms work over deterministic structures, whereas so-called *probabilistic model checking* relaxes the determinism assumption to allow for models with uncertainty, or probabilistic behavior [5]. Both traditional and probabilistic model checkers, once given an acceptable model, will take queries regarding the model and rephrase them as specifications or properties of the model. The model checkers will then verify if these properties are satisfied by the model, and if so, the query receives a "yes" answer, otherwise it receives "no." This output can be rephrased to support more detailed responses, such as a system trace or other information that may be useful to the inquirer. In this work we demonstrate how to encode pertinent epidemiological problems, which often include random behavior, into an acceptable format for a probabilistic model checker, and execute queries on top of these encoded models to gain insight into the original epidemiological problem. We explore how disease prevention techniques such as vaccination can be investigated through our model checking framework, and discuss the flexibility of our approach by considering the impact that changing the underlying epidemiological model would have upon our encoding.

2 Background

We present here a study of different areas of research in both the epidemiological and model checking realms. We will demonstrate how model checking techniques can be used to concisely represent many questions of interest regarding disease prevention and control.

2.1 Compartmental Models of Epidemiology

One basic but well-characterized model of the spread of disease through a population is known as the Susceptible, Infectious, and Recovered Model, commonly referred to as the *SIR model*. The SIR model for a specific disease assumes that there are three distinct divisions of a given population of N people, specifically at some time instance t they are:

- $S(t)$ - The number of people who are susceptible (have not yet been infected) to the disease at time instance t.
- $I(t)$ - The number of people who are actively infected and can spread the disease to members of the susceptible population at time instance t.

- $R(t)$ - The number of people who were infected but have now recovered from the disease by time instance t.

Because these three sub-populations form a partition of the overall population, we have that

$$\forall t,\ S + I + R = N.$$

This model assumes a unidirectional progression of the population, namely each member of the populations follows the flow diagram:

$$Susceptible \rightarrow Infectious \rightarrow Recovered$$

The SIR model is thus more well-suited for diseases to which members of the population cannot be re-infected (perhaps due to such factors as an inherent resistance built up during a recovery period, such as is the case with most flu strains). It is easy to extend or adjust the model to allow for alternative disease patterns. Such models of this sort are termed *compartmental* models, due to their subdividing of the population into characteristic groups. Some other common compartmental models are described in Table 1.

Model Name	Flow Diagram
SIR	$Susceptible \rightarrow Infectious \rightarrow Recovered$
SEIR	$Susceptible \rightarrow Exposed \rightarrow Infectious \rightarrow Recovered$
SIS	$Susceptible \longleftrightarrow Infectious$

Table 1: Common compartmental models of disease spread in a population. The *exposed* population models a latency in becoming infected and becoming infectious.

In addition to the modeling of the population partitions, compartmental models also include rules about how members of the population transition from state to state. A common approach is to take the time component t to be a continuous variable, and assume that each edge of the flow diagram has a transition rate associated with it. Such rates are represented as ordinary differentiable equations (ODEs), which allow for compact transition rules. For example, given constants β, ν, one can represent the change in sub-populations of the SIR model as follows:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\beta \cdot I(t) \cdot S(t) \\
\frac{dI(t)}{dt} &= \beta \cdot I(t) \cdot S(t) - \nu \cdot I(t) \\
\frac{dR(t)}{dt} &= \nu \cdot I(t)
\end{aligned}$$

where β is the average contact rate per unit time and ν is the recovery rate per unit time. It can be observed that these differentials sum to 0, as the SIR model assumes a fixed population, and thus these three differential equations regarding the sub-populations with the added constraint is equivalent to simply two differential equations on the transition edges of the SIR model:

$$\begin{aligned} \frac{d\left(S \to I\right)}{dt} &= \beta \cdot I\left(t\right) \cdot S\left(t\right) \\ \frac{d\left(I \to R\right)}{dt} &= \nu \cdot I\left(t\right) \end{aligned}$$

The assumption of a fixed population makes SIR-like models well-suited for relatively short-term time spans, such as a few days, weeks or months. Any longer and the model can become inappropriate due to excessive error in estimating the population. Again, variants exist that incorporate births and deaths into the population, of special importance may be deaths caused by the disease itself. With the above equations governing the population dynamics, one can visualize the dynamics of a population conforming to the SIR model. Figure 1 shows an example with $\beta = \nu = 1$, and all of the population begins healthily except for one infected person, who proceeds to spread the disease to the rest of the population as per the SIR relationships.

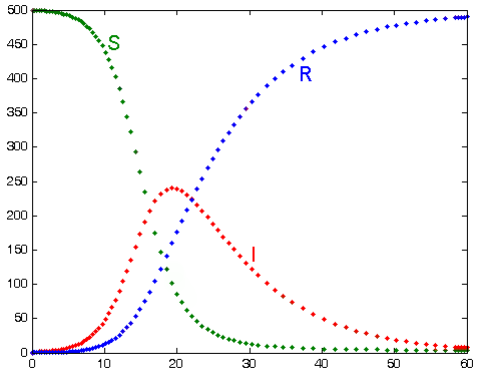


Figure 1: Dynamics of a SIR-modeled population over time. The green curve, marked S, is the percentage of the total population at a given time instance that are susceptible. The red curve, marked I, is the percentage that are infected and the blue curve, marked R, is the percentage that have recovered.

One major advantage to a compartmental model such as SIR is the relatively low number of parameters needed to represent the model; SIR in particular requires only two: β and ν, and then can model a given population where the values of S, I, and R (and consequently N) are known at some time instant. However, the oversimplification of compartmental models may warrant some criticism. The differential equations assume

homogeneous (i.i.d.) contact between all members of the population, which in many scenarios and scales, is an incorrect assumption. More complex models may be desired, which we will now consider.

2.2 Contact Networks

To address the oversimplification of the contact systems that compartmental models carry with them, one can turn to an alternative scheme known as *contact networks* [6]. Taking in one sense the other extreme, a contact network treats each person's role in a population as distinct (a heterogeneous population), and tracks properties of each individual. We can formally define a contact network.

Definition 1 *A* contact network *is a directed graph graph* $G = (V, E)$ *with edge weights where:*

- *Each node* $v \in V(G)$ *represents a person. Each* v *has a state assigned to it representing its current disease status (such as, but not limited to, susceptible, infectious, or recovered).*
- *An edge* $(v_1 \to v_2) \in E(G)$ *represents interactions between* v_1 *and* v_2 *(interactions need not be mutual).*
- *Edge weights represent relative strengths of interactions.*

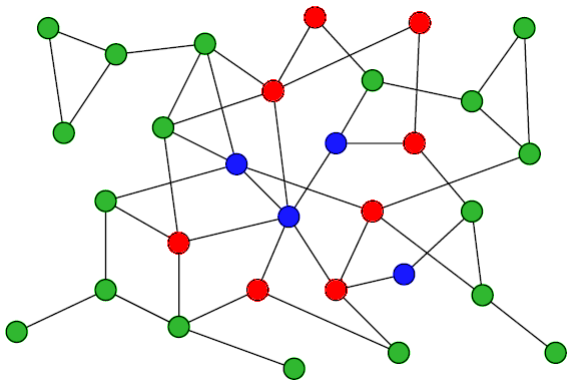


Figure 2: An example contact network. Edge weights are omitted for clarity, and we use undirected edges here to symbolic mutual interactions. The colors of the nodes represent their current disease state.

Thus a contact network encapsulates each individual's distinct role in the larger population. An example visualization is shown in Figure 2. The interactions considered are usually those that are conducive to the spread of the particular disease in question

(e.g. sharing of a cup may be used for colds, but not sexually transmitted diseases). On top of such a network, we can again impose a per-person state transition such as was seen for the SIR model, namely $S \to I \to R$. Given such a contact network, various types of transmission models have been proposed. A common theme assumes discrete time steps [7], and defines the probability of one infectious node infecting its susceptible neighbor, given that it is infected for τ time units, is

$$p_{i,j} = 1 - (1 - r_{i \to j})^{\tau} ,$$

where $r_{i \to j}$ is the edge weight along edge $(i \to j)$. If a continuous time scale is used, this can be extended to

$$p_{i,j} = 1 - \exp\{-r_{i \to j} \tau\} .$$

The edge weights provide another level of heterogeneity past the graph structure; in particular the susceptibility of an individual varies from individual to individual (here based upon the neighboring nodes, and the relative edge weights) [6]. While certainly a more detailed and flexible model, such specific interactions may not be available. Furthermore, such a detailed model requires a large number of parameters. The structure of the network must be specified, as does the weights on each edge. This may lead to computational challenges when scaling up the population under question.

To address these issues, one can take an intermediate approach, where a node instead represents a set of people that are expected to behave *homogeneously*, and imposes a contact network between such meta-nodes. Disease behavior inside such a meta-node can degenerate to less complex models, such as SIR. Epigrass [3], a tool intended for studying the spread of disease over networks, uses a similar model in considering bus routes over a set of regions in Brazil. Figure 3 shows an example transformation from geographic region to contact network.

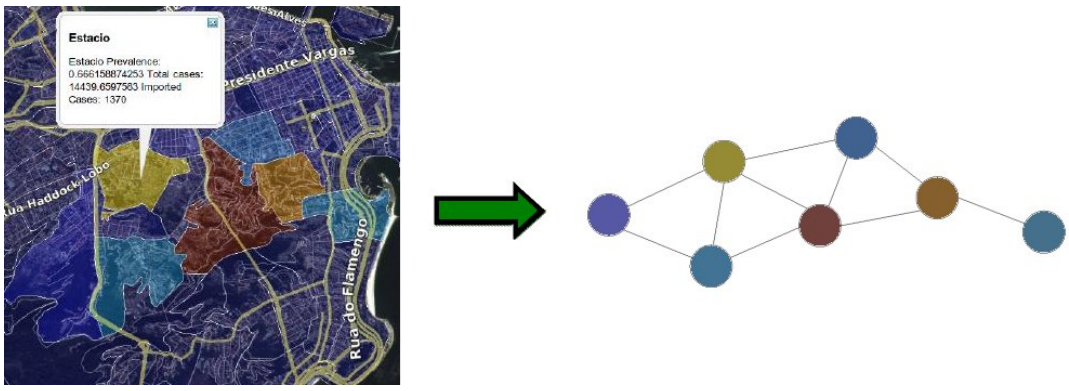


Figure 3: Example dataset used by Epigrass, to construct a contact network modeling contacts made on bus routes.

2.3 Probabilistic Model Checking

Once we have a particular chosen model for disease spread over a population, be it as simple as an SIR model, or utilizing a more complex approach such as a contact network, a natural question to ask is "What sorts of characteristics can we model?" Perhaps more to the point, we'd like our model to be able to answer questions that are relevant to the inspection and understanding of disease spread. In particular, as mentioned in Section 1, we might be interested in questions related to prevent and/or control disease. Such questions may be:

1. "At time instance t, what fraction of the total population is infected?"
2. "Allowing the disease to run its full course, what fraction of the total population becomes infected?"
3. "Will person p (or group of people g) become infected within t time units of the onset of the disease?"

If we consider the second question above, it might seem strange at first, because in both the common SIR model and contact networks, it is quite possible that the entire population will become infected (and then recover) exactly once. However, this brings us to the issue of disease prevention. We have previously assumed that the population is initially entirely in the "susceptible" state (except for one infected individual). However, the introduction of such things as vaccinations to our model would result in individuals being initialized to be in the "recovered" state. Of course, some models (including formulations of contact networks as mentioned above) are probabilistic and will not always given the same answer for deterministic questions of this sort. We can relax our queries somewhat by rephrasing them with a probabilistic flavor:

1. "At time instance t, what *is the likelihood* that the fraction of the total population is infected *is at least* c?"
2. "Allowing the disease to run its full course, what *is the likelihood* that the fraction of the total population becomes infected *is at least* c??"
3. "What *is the likelihood* that person p (or group of people g) becomes infected within t time units of the onset of the disease?"

The realm of model checking allows us to pose questions of the sort above, both probabilistic and non-probabilistic. Research in *temporal logics* allow us to make assertions over some existing state-machine-like construct as well as handle a notion of time. There are many flavors of temporal logic, one pertinent one is known as Continuous Stochastic Logic, or CSL [1]. CSL deals with models that both work over continuous time and involve uncertainty in the queries and the model, as above. tCTL, another form of temporal logic, handles continuous time but no uncertainty. We will focus on

CSL here; the syntax and semantics for tCTL are similar. We first define the formal structure that a CSL statement works over, a Continuous Time Markov Chain (CTMC), which can be thought of a continuous time version of a labeled state machine:

Definition 2 *A continuous time Markov Chain M is a 4-tuple $(S, \mathcal{A}, \Lambda, l)$ with components defined as follows:*

- $S = \{s_1, s_2, \ldots, s_n\}$ *is a finite set of states.*
- $\mathcal{A}$ *is a finite set of* atomic propositions*.*
- $\Lambda : S \times S \to \mathbb{R}$ *is the* transition rate matrix*.*
- $l : S \mapsto 2^{\mathcal{A}}$ *is the* labeling function*.*

Λ *has the following properties:*

- *Off-diagonal entries are non-negative.*
- *The diagonal elements $\lambda_{j,j} = -(\sum_{i \neq j} \lambda_{j,i})$.*
- *When at state s_j, in time dt we can transition to state s_k (for $k \neq j$) with probability $\lambda_{j,k} dt$.*

The atomic propositions correspond to a set of labels that each state may have. Thus, we can conceive a continuous transition from state to state such that at any fixed time point t, we have a set of properties (atomic propositions) satisfied, corresponding to what state we are currently in. Similar structures exist to correspond with tCTL. Now, given our notion of a CTMC, we can formally define the CSL temporal logic.

Definition 3 *A CSL formula Φ defined over a CTMC M is one of the form*

$$\Phi ::= \mathit{true} \mid a \mid \Phi \wedge \Phi \mid \neg\Phi \mid \mathcal{S}_{\bowtie p}(\Phi) \mid \mathcal{P}_{\bowtie p}(\varphi)$$

$$\varphi ::= X\Phi \mid \Phi \mathcal{U} \Phi \mid \Phi \mathcal{U}^{\leq t} \Phi$$

where $a \in \mathcal{A}$, $p \in [0, 1]$, $t \in \mathbb{R}_{>0}$, and $\bowtie \in \{\leq, =, \geq\}$ Each Φ is also called a state-formula, and each φ is called a path-formula.

The interested reader is invited to consult Baier et al.[2] for a full description of the CSL semantics. A CSL formula can represent the queries listed above:

- "At time instance t, what *is the likelihood* that the fraction of the total population is infected *is at least c*?"
  $\rightarrow \mathcal{P}_{\geq ?}\left(\texttt{true}\, \mathcal{U}^{\leq t}\ (\texttt{I/N > c})\right)$

- "Allowing the disease to run its full course, what *is the likelihood* that the fraction of the total population becomes infected *is at least c*?"

  $\rightarrow \mathcal{S}_{\geq ?}\,(\texttt{I/N > c})$

- "What *is the likelihood* that person p (or group of people g) becomes infected within t time units of the onset of the disease?"

  $\rightarrow \mathcal{P}_{\geq ?}\left(\texttt{true}\ \mathcal{U}^{\leq t}\ \ (p \text{ is infected})\right)$

Here a question mark (?) appears in the formulae in the location of whatever value we wish to find; in the case of the third example, we wish to find the corresponding probability. That is to say, rather than just verifying whether or not the CSL formula is satisfied or not, we also wish to find a satisfying value for the unspecified variable. In most cases, the underlying algorithms used for the model checking need not change at all. In the next section, we discuss a tool called PRISM that incorporates this functionality natively. The atomic propositions can be encoded to allow numerical values, not just boolean expressions, as is done here (which allows us to refer to the number of infected people `I` or total number `N`, for example).

3 Methodology

We demonstrate how to perform a set of queries over a chosen disease model. The chosen model is first encoded as a CTMC, and then the queries are written as CSL formulae, which are run on the CTMC models. We use PRISM [4], an available tool for performing this probabilistic model checking.

3.1 Our Disease Model

Similar to Epigrass, we adopt an intermediate model that lies between a true per-individual contact network and a homogeneous SIR model. Specifically, we have a contact network with meta-nodes where each node represents a portion of the population that can be considered homogeneous. Each individual meta-node is modeled to function as its own population, so an SIR model exists for each meta-node. Furthermore, edges between meta-nodes indicate points of contact between populations. These edges facilitate the transfer of disease from one meta-node to another. We take here a simplifying assumption that for two populations A and B, the likelihood of any given member $a \in A$ coming into contact with member $b \in B$ is the same as any other member $b' \in B$, and vice-versa. Formally, we have for a given meta-node u (same as the SIR transmission rules),

$$\frac{d\left(S_u \to I_u\right)}{dt} = \beta \cdot I_u\left(t\right) \cdot S_u\left(t\right) \quad (1)$$

$$\frac{d\left(I_u \to R_u\right)}{dt} = \nu \cdot I_u\left(t\right) \quad (2)$$

and for an edge $(u \to v)$ between meta-nodes with edge weight $r_{u \to v}$:

$$\frac{d\left(S_v \to I_v\right)}{dt} = \beta \cdot r_{u \to v} \cdot \frac{I_u\left(t\right)}{pop_u} \cdot \frac{S_v\left(t\right)}{pop_v} \quad (3)$$

This final rule is similar to the SIR rule for infecting susceptible members of the population; except that the infectious members come from population u and the susceptible members are in population v. It should be noted that much like the SIR model, we assume that for any given population (meta-node), the members are fixed. This means what while we can simulate individuals between populations interacting, they remain in their respective groups. Much as before with the SIR model, one could design a model involving changing populations, if needed.

3.2 PRISM - A Tool for Probabilistic Model Checking

As briefly mentioned earlier, we use PRISM, a tool for performing probabilistic model checking, including the checking of CSL formulae over CTMCs. PRISM utilizes its own language for encoding state-machine-like structures such as CTMCs. The language is fairly simple; code samples are given in Figure 4 and Figure 5. These modules are templates for each node and edge, and we ground these templates out for each node and edge that we have in our network. The variables `rate_I` `rate_R` correspond to the β and ν parameters from Section 2.1, which are constant over all nodes and edges of the graph. Guard conditions are placed on each action that can take place, to ensure the proper circumstances (for example, the infection rule should only trigger if there is at least one person who is infected and who can become infected). With this in mind, it should be easy to see the direct encoding of Equations (1) - (3), and how to modify the rules if we choose to change the disease transmission model.

Following the encoding of the nodes and edges, we also utilize the *cost/reward* structures available in PRISM. These simplify the eventual CSL formulas that we formulate. PRISM supports a system of rewards that are awarded when certain states are encountered or conditions are met. For our purposes, we utilize the reward feature to access information about the model's state that is interesting to us, in particular each node's breakdown of susceptible, infectious, and recovered members. This is accomplished with the simple reward structure presented in Figure 6. In this figure, we create a reward which is here named "`I_i`," which takes the value $\frac{I_i}{pop_i}$, the fraction of node i's population that is infectious. We can similarly create rewards `S_i` and `R_i` for each node i in

```
const int pop_i = 120;
global S_i : [0..pop_i] init 100;
global I_i : [0..pop_i] init 10;
global R_i : [0..pop_i] init 10;

module node_i
  // infection within node i
  [] (S_i > 0) & (I_i > 0) ->
    (rate_I*(I_i)*(S_i/pop_i)) : (S_i'=S_i-1)&(I_i'=I_i+1);

  // recovery within node i
  [] (I_i > 0) ->
    (rate_R*I_i) : (I_i'=I_i-1)&(R_i'=R_i+1);
endmodule
```

Figure 4: PRISM code for a node of the contact network.

```
// edge weight
const double trans_u_v = 2;

module edge_u_v
  // models pop_u infecting pop_v
  [] (S_v > 0) & (I_u > 0) ->
     (rate_I*trans_u_v*(I_u/pop_u)*(S_v/pop_v)) :
     (S_v'=S_v-1)&(I_v'=I_v+1);
endmodule
```

Figure 5: PRISM code for an edge of the contact network.

```
rewards "I_i"  true : I_i / pop_i; endrewards
```

Figure 6: PRISM code for creating a reward for the existing model.

the graph. Reward queries, a PRISM extension of CSL, can then be used to compute the *expected* value of this reward at a specified time instance. For example, the query `R{"I_i"}=? [I = T]` asks "what is the expected fraction of node i that is infected at the instantaneous time T?" In this sample query, the standalone symbol `I` is part of the PRISM reward query structure, standing for "instantaneous," and should not be taken to be have anything to do with the infectious portion of the population. Following the creation of such reward structures, we can effectively profile the expected outcome of

any population in the network model, or by creating rewards summing all the nodes' statistics, produce an aggregate summary akin to Figure 1.

3.3 PRISM's Verification Approach

PRISM is termed a probabilistic model checker. As discussed in Section 1, this implies that verification is performed on top of a model that incorporates some random behavior. The method of verification is left open. Many traditional approaches seek to answer queries exactly, meaning that with complete certainty the answer that they return is correct. However, this oftentimes leads to issues in computational complexity. Efficient data structures such as binary decision diagrams (BDDs) allow drastic speed-ups in many cases, but do not solve the problem in general. An alternative approach instead tries to output an *approximate* solution, which gives a high probability guarantee of being correct rather than full confidence. PRISM uses such an approach, through a combination of analysis of the underlying model and a sampling of random execution traces.

4 Experimentation

Given the formulations of the models we have from Section 3, we have constructed a model with 7 populations (meta-nodes) and use SIR to model disease spread inside each node, as discussed in Section 3.1. A visualization of the contact network used is shown in Figure 7. We then discuss using vaccination as a disease prevention/control mechanism.

4.1 Disease Progression, No Prevention

The first and simplest initialization that we consider is when all but one person in the overall population is susceptible, and one individual is chosen to be infectious. Here, we select node "a" to begin with one infectious person. We present results in the form of a time series set of data, showing snapshots of the population as the disease propagates. To obtain the values needed to produce this data, we use the reward-based queries discussed in Section 3.2 to obtain approximations to the expected values of the population breakdowns. Figure 8 shows several key points during the disease progression. A series of graph visualizations of our model at different time point accompany the plot, which allows us a more distributed view of the population's health at a particular point in time.

4.2 Disease Prevention - Vaccination

Following the work done above in modelling the progression of disease, we can also consider the impacts of taking preventative or responsive measures at the disease onset.

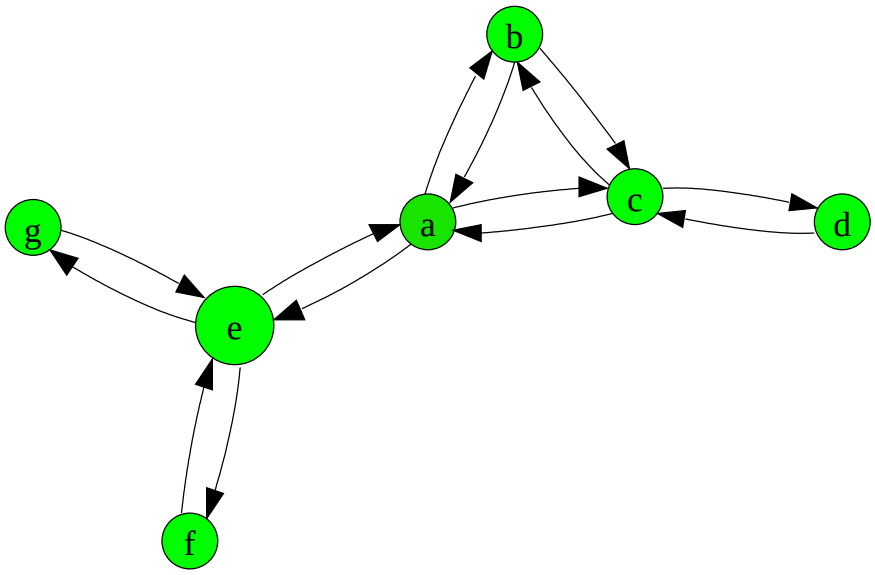


Figure 7: The contact network being modeled. Each node represents a different population which interact with other populations as determined by the edge set. Node size is relative to the total population contained in that node.

This would typically be done to limit the number of total infected population or perhaps the total number of infection people of a target type (such as pregnant women, or young children, etc). There are a number of ways that one could consider doing this, we focus on a common tactic - vaccination. Vaccination impacts our model by shifting a portion of the initial populations to the safe "recovered" state whereby they will not be able to infect or become infected. Total vaccination will, in theory, prevent any outbreak from occurring. However, vaccination on such a large scale is costly and usually impractical fiscally. One naive approach is to uniformly vaccinate a fraction of the entire population indiscriminately. However, one might expect that a better solution is possible, and indeed, vaccinating specific, critical points of a population can be a viable tactic, such as more fully targeting the neighbors of a diseased node in the contact network [8]. While leaving other nodes "open to disease" so to speak by administering little to no vaccination, this nevertheless can effectively prevent disease spread due to the spreading nature of disease. In particular, the definition of a "contact" between populations here is crucial and must be well-suited for the disease under study. The issue of finding such critical points in more complex models is oftentimes non-trivial given more complex network.

It is very easy for our model to express an alteration such as vaccination; the initial population distribution changes from being all susceptible to containing some recovered members, and the queries and simulations are performed the exact same way. The question of "key" points in a population is more clear in our contact network formulation as well; we might wish to vaccinate all nodes that serve as bridges (formally that have

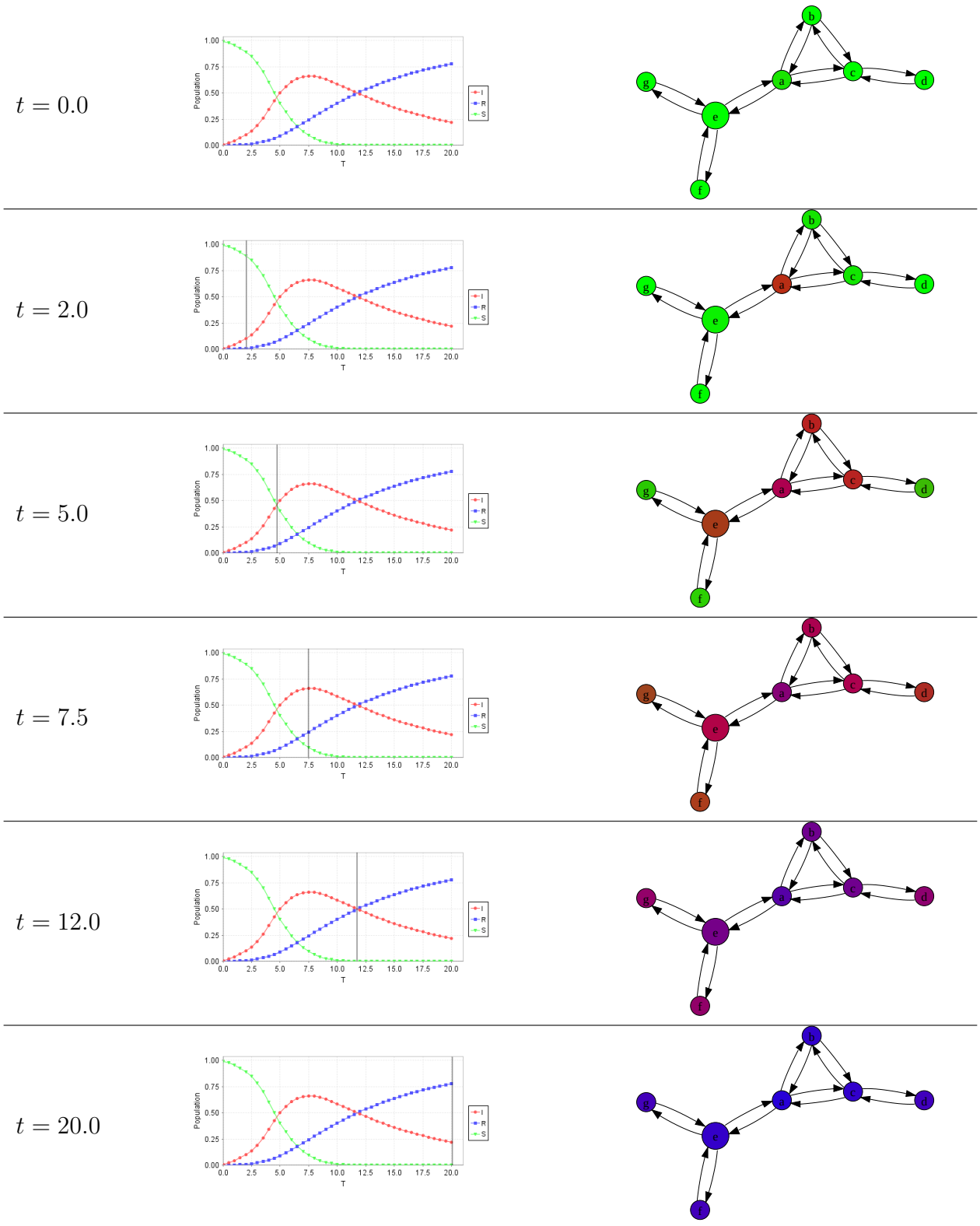


Figure 8: Progression of disease in non-vaccinated case. Overall fractions of the population that are susceptible, infectious, and recovered are shown in the plots over a time interval of $t \in [0, 20]$. Contact network visualization is also presented, where the color of each meta-node indicates the constituents of its particular population. The color scheme is kept the same as in Figure 1.

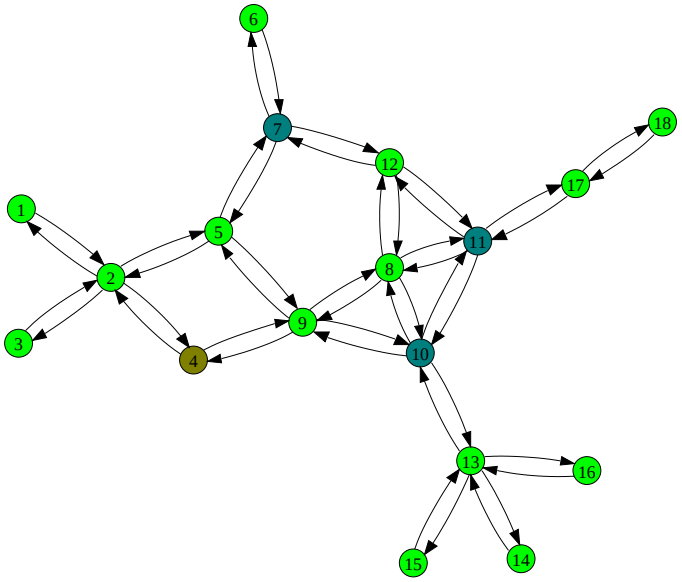


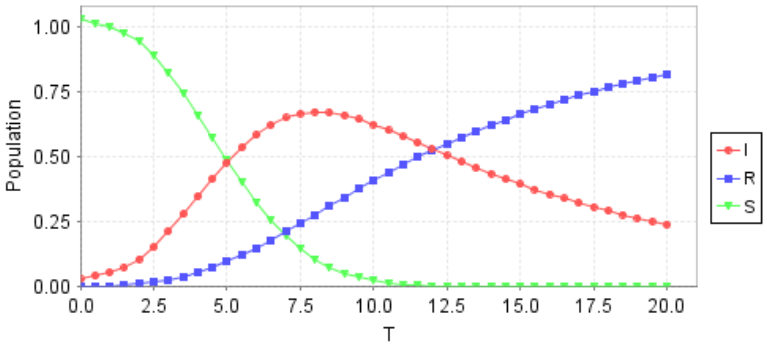
Figure 9: The contact network being modeled. Each node represents a different population which interact with other populations as determined by the edge set. Node size is relative to the total population contained in that node.

a high betweenness centrality, a measurement often used in social network research) to the rest of the population. In the example contact network show in Figure 8, if node f or g were initially the only infected populations, then we would want to vaccinate node e which is the only way that the disease could spread to other populations in the network.

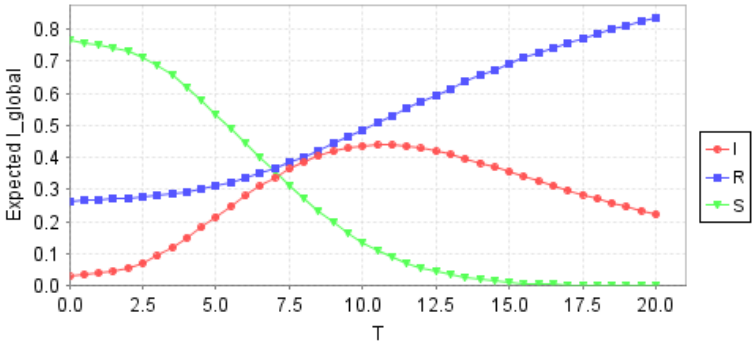
We choose to use a more complex contact network for this stage, shown in Figure 9. We first run the initial non-vaccinated simulation to develop a baseline of the dynamics of the network. We then apply two vaccination approaches in turn: we first apply a uniform vaccination to 25% of the entire population on a per-node basis. That is to say, that each node of the contact network has 25% of its population vaccinated immediately. As a second strategy, we vaccinate 50% of key populations. The three sets of results are shown in Figure 10.

Vaccination seems to help in general, as the red infectious curve becomes suppressed throughout its lifetime. Currently, the uniform 25% vaccination outperforms the targeted 50% version. This currently is the case for two reasons: (1) because the overall number of vaccinations in the 25% is drastically higher. Given that there are 18 nodes in the graph and uniform populations in each node, and that only three nodes were vaccinated in the 50% targeted case, while the 25% uniform strategy requires vaccinations of a quarter of the population, the targeted case only requires $\frac{0.5 \cdot 3}{18} = \%18.75$. Additionally, the "key" locations chosen were not optimal in the sense that they were not direct neighbors of the initially infected node; this could suggest that a more blanket uniform strategy may be needed if issuing a delayed response to a rampant disease where the

No vacciation



Uniform 25%



Targeted 50%

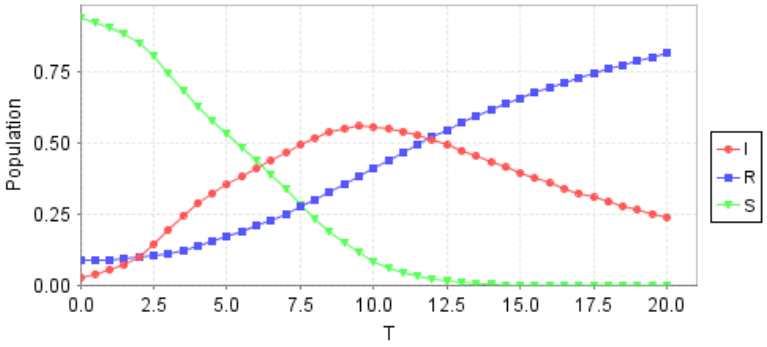


Figure 10: The effects of vaccination on the larger contact network.

disease has already spread past a single node.

5 Conclusion & Future Work

We have considered the modeling of disease spread over populations, and shown how one can readily encode existing classes of models (compartmental, true contact networks, and intermediate variants) into CTMCs which allow for queries to be posed over the models using temporal logics, which are a natural fit for many types of queries that one would desire for the purposes of disease prevention and control. Control measures such as vaccinations were considered, and it was shown that, as one would expect, vaccinating the correct target group(s) of a contact network prevents a larger number of infections that blindly administering the same (or more) vaccinations to different members of the population. This type of prevention measure adjusts the initial population states by shifting many to a "recovered" state. An alternative measure, proposed by Myers, suggests [6] that by restricting the contacts themselves, one can also limit disease spread effectively. This would correspond to deleting edges in the model's contact network (or dropping transmission rates along the edges to 0), and could be investigated further.

References

[1] Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert Brayton. Model-checking continuous-time markov chains. *ACM Trans. Comput. Logic*, 1(1):162–170, 2000.

[2] Christel Baier, Joost-Pieter Katoen, and Holger Hermanns. Approximate symbolic model checking of continuous-time markov chains (extended abstract). page 781. 1999.

[3] Flavio C. Coelho, Oswaldo G. Cruz, and Claudia T. Codeco. Epigrass: a tool to study disease spread in complex networks. *Source code for biology and medicine*, 3(1), February 2008.

[4] A. Hinton, M. Kwiatkowska, G. Norman, and D. Parker. PRISM: A tool for automatic verification of probabilistic systems. In H. Hermanns and J. Palsberg, editors, *Proc. 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06)*, volume 3920 of *LNCS*, pages 441–444. Springer, 2006.

[5] M. Kwiatkowska, G. Norman, and D. Parker. Quantitative analysis with the probabilistic model checker PRISM. *Electronic Notes in Theoretical Computer Science*, 153(2):5–31, 2005.

[6] Lauren Ancel Myers. Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *American Mathematical Society*, 44(1):63–86, January 2007.

[7] Mark E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1), 2002.

[8] Ana Perisic and Chris T. Bauch. Social contact networks and disease eradicability under voluntary vaccination. *PLoS Comput Biol*, 5(2):e1000280+, February 2009.