

First Annual Progression Report

Ethan Kelly

Annual Progression Review date: Thursday 24th June, 2021

ETHAN KELLY

`e.kelly.1@research.gla.ac.uk`

PRINCIPAL SUPERVISOR: JESSICA ENRIGHT

`Jessica.Enright@glasgow.ac.uk`

SECOND SUPERVISOR: ALICE MILLER

`Alice.Miller@glasgow.ac.uk`

Contents

1	Introduction	2
1.1	Thesis Statement	2
1.2	Report Roadmap	2
2	Literature Review	2
2.1	Context	2
2.2	The Firefighter Problem	3
2.3	Markovian <i>SIR</i> epidemics on networks	3
2.3.1	Standard <i>SIR</i> model with fixed population	4
2.3.2	Extending the <i>SIR</i> model to graphs	4
2.4	Percolation Theory	6
3	Work done so far	6
3.1	Agency-oriented modelling	6
3.1.1	Implementation of agency-based models	6
3.1.2	Experimental results	8
3.1.3	Computational Biology Conference	10
3.2	Graph-based compartmental models	10
3.2.1	Adding a new state to the model	10
3.2.2	Implementation	11
3.3	Uses of Percolation	11
3.3.1	Better than random	11
3.3.2	Reproduction rate	12
4	Research Plan	12
4.1	Goals	12
4.2	Timeline	13

1 Introduction

This Annual Progress Report, the first of our PhD programme, serves to detail the work we have done in the first year of our research. We first outline the research title we have worked under and provide an overview of the discussion to follow.

1.1 Thesis Statement

Our working research title is *game-theoretic and probabilistic methods applied to spatial network models of contagion*. Over the past year, we have spent time reading and developing ideas related to graph models of contagion, such as FIREFIGHTER and similar games on graphs. We have explored optimal strategies in classic formulations and applied probabilistic methods in order to adapt these games to stochastic contexts. Such a broad approach has meant that we now have a wide range of tools available to us in understanding and developing models of disease and other contagions.

1.2 Report Roadmap

Our research over the past year has focused into two main strands: a formal, mathematical approach and an experimental approach. Our motivation for this is to ultimately develop a model for disease spread formally, using dynamical system and graph theoretic approaches, which we then implement experimentally to explore different approaches to containment. One key benefit of this approach is that experimental data can be fed into a graph model of disease very quickly, contextualising the results of the model and informing the expected outcomes of approaches to containing the disease very rapidly, which is especially important in epidemic scenarios [3].

To explain the work we have done in detail, we will begin by discussing the literature examined in the past year, with particular focus on papers that have proved foundational, to provide a background to the work we have done in extending these results and concepts. We then discuss the work we have done in our research, including some mathematical results and some key experimental implementations and relevant results. We will then discuss the direction we hope to take this work moving forward and detail milestones we hope to reach in the coming year.

2 Literature Review

2.1 Context

To better understand the relevance and significance of this literature review, we first provide a brief overview of where the concepts that will be outlined fit into our work so far. We begin with the Firefighter Problem and how it can be adapted to a rudimentary model of disease: a huge amount of our research (particularly the experimental strand) has started with this problem as its foundation; in Section 3, we explain the ways we have extended the classic problem in stochastic and game-theoretic directions.

Further, we discuss compartmental graph models of disease, examining the differential equations required to describe such models exactly. This has formed the basis of our more formal work over the past year, so again a discussion of the results and knowledge in this field that pre-dates our work provides important background for Section 3. While this work appears fairly distinct in character to the more experimental prong of our approach, in principle both methods aim to take existing approaches to disease modelling and introduce a new way of representing protection from disease, both as an external defence strategy and as an internal inclination in individuals.

We end this section with a discussion of Percolation Theory and its use in a graph-theoretic context. We have identified this as an important tool for disease modelling using graphs, but we have not yet employed it to a significant extent. We will, nonetheless, discuss the areas that we believe it will prove useful in Section 3.

2.2 The Firefighter Problem

The Firefighter Problem, which we refer to as simply **FIREFIGHTER** models on a given graph an outbreak of fire with a firefighter strategically blocking its path [6]. We formalise **FIREFIGHTER** as follows: at $t = 0$, a fire breaks out at a vertex v_0 of graph G . The firefighter then ‘protects’ another vertex of G . A protected vertex is protected for the remainder of the game and a vertex that is on fire is ‘burning’ for the rest of the game. The fire now spreads to any adjacent vertices that are neither protected nor burnt. After that, the firefighter protects another vertex, the fire spreads again and so on. The following is a decision formulation given by Finbow and MacGillivray for the problem [2]:

FIREFIGHTER
INSTANCE: A rooted graph (G, r) and an integer $k \geq 1$.
QUESTION: Is there a finite sequence d_1, d_2, \dots, d_t of vertices of the graph G such that:

- i d_i is neither burned nor defended at time i ,
- ii At time t , no undefended vertex is adjacent to a burning vertex, and
- iii At least k vertices are saved at the end of time t ?

Common problems in classic **FIREFIGHTER** often involve minimising the number of vertices that will be burnt and, in a given class of trees, determining the average number of burnt vertices. There are several results available to us in classic **FIREFIGHTER** - for one, the decision problem is NP-complete, even if restricted to trees with maximum degree three. Further, the problem is solvable in polynomial time for graphs of maximum degree three, so long as the fire starts at a vertex of degree two [2]. We also have some results regarding game strategies: for instance, the greedy algorithm is a $1/2$ -approximation for the problem on trees [2].

If we let each vertex represent an individual and edges between them represent social contact, we obtain a simple model for disease infection. There are many natural contextualisations of **FIREFIGHTER** beyond modelling disease spread, which we may choose to apply our work to in future - for instance, we may think of edges representing virtual contact between individuals on social media, yielding a model for the spread of viral internet memes [11].

2.3 Markovian *SIR* epidemics on networks

Definition 2.1. Compartmental models. Mathematical ways of simulating how sections of a population interact, usually involving systems of equations describing the behaviour of each compartment and then studying how compartments interact with each other.

The states in compartmental models can describe many states an individual (typically a person or an animal) can be in. For instance, predator/prey compartments can simulate a hunting situation and susceptible/infected/recovered compartments may model disease spread. In this discussion, we begin by detailing the latter of these examples - the *SIR* model - allowing us to later explain extension of this model to a graph-theoretic context and understand the generation of systems of equations describing the model.

The work detailed here on Markovian *SIR* epidemic graph models from [9] is hugely foundational for our formal work so far. The authors of that work note that there are generally three approaches to compartmental models of disease: we can take averages at population level, maintain a probabilistic view by considering the full state space, or begin modelling at the level of vertices and build up to larger structures from there. The second of these three describes the approach we have taken to our work so far, as will be seen in Section 3 and the latter of these three approaches is the one used in the work on Markovian *SIR* graph epidemics we are now focusing on. Broadly, their approach begins by considering equations for single vertices, then dependencies on pairs, then triples and so on until we reach the full system size. Such equations are well defined and consistent, which is not difficult to see [9].

2.3.1 Standard *SIR* model with fixed population

The *SIR* Model is a compartmental epidemiological model with the following three states.

Definition 2.2. Susceptible. An individual who does not currently have but could contract the contagion.

Definition 2.3. Infected. An individual who currently has the contagion and is infectious.

Definition 2.4. Recovered. An individual who previously had the contagion and subsequently recovered, possibly granting them some level of immunity in certain contexts.

We define $S(t)$, $I(t)$ and $R(t)$ as the numbers of susceptible, infected and recovered people (all at time t) respectively. For a fixed¹ population N , we specify that $S(t) + I(t) + R(t) = N$. Then, for β and γ the rates of infection and recovery respectively, the standard *SIR* model without vital dynamics (birth and death rates) is given as follows [7]:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{SI}{N}, \\ \frac{dI}{dt} &= \beta \frac{SI}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I - \mu R.\end{aligned}$$

2.3.2 Extending the *SIR* model to graphs

In order to extend our *SIR* model to graphs, we first examine the probability of an agent being in a given class. Let $\langle A_i \rangle$ represent the time-independent probability of vertex i being in state A , so $\langle A_i B_j \rangle$ represents the probability of vertices i and j being in states A and B respectively [9]. We begin with a contact network, where vertices represent individuals and the edges between them represent social contact which may serve as an infection pathway. Then, the adjacency matrix G of is constructed by assigning $G_{ij} = 1$ when i and j share an edge and $G_{ij} = 0$ otherwise.² Then, we extend this contact network to a transmission network: let β_i represent the per-link infection rate for individual i and γ_i represent the recovery rate for i . For the transmission matrix T , we assign $T_{ij} = \beta_i$ if there is a single-edge route of infection between i and j and $T_{ij} = 0$ otherwise. Often, we will consider unweighted and undirected graphs, but in general T_{ij} may not equal T_{ji} .

We can replace β_i with a term involving such a transmission matrix in order to begin extending the usual *SIR* model into a network realm. Using the substitution $\beta_i \frac{SI}{N} = \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle$, the equations become

$$\begin{aligned}\langle \dot{S}_i \rangle &= - \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle \\ \langle \dot{I}_i \rangle &= \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I_i \rangle \\ \langle \dot{R}_i \rangle &= \gamma_i \langle I_i \rangle,\end{aligned}$$

which are the evolution equations given in [9].

The work done in [9] has two goals. Firstly, the authors aim to provide an exact, deterministic representations of Markovian *SIR* epidemics on graphs with and without loops. Secondly, they seek to identify a link between the structural properties of graphs and the viability of closures that can be used to write down exact systems of equations. In particular, the authors show this structural link is founded on cut-vertices. Cut-vertices are vertices that, if removed from a connected graph, result in the formation of at least two disconnected sub-graphs [9].

¹A fixed population is a reasonable assumption where an epidemic is short-lived.

²Because we are infrequently interested in self-transmission, we often set $G_{ii} = 0$.

The significance of cut-vertices is that they permit us to create closures in the systems of equations that describe a compartmental model. To better explain closures, consider the equations below for calculating singles and pairs [9]:

$$\begin{aligned}
\langle \dot{S}_i \rangle &= - \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle, \\
\langle \dot{I}_i \rangle &= \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I_i \rangle, \\
\langle \dot{S}_i I_j \rangle &= \sum_{k=1, k \neq i}^N T_{jk} \langle S_i S_j I_k \rangle - \sum_{k=1, k \neq j}^N T_{ik} \langle I_k S_i I_j \rangle \\
&\quad - T_{ij} \langle S_i I_j \rangle - \gamma_i \langle S_i I_j \rangle, \\
\langle \dot{S}_i S_j \rangle &= - \sum_{k=1, k \neq j}^N T_{ik} \langle I_k S_i S_j \rangle - \sum_{k=1, k \neq i}^N T_{jk} \langle S_i S_j I_k \rangle,
\end{aligned}$$

This is *not* a closed system - we require equations for triples, although we have only sought to express singles and doubles. Obtaining closures here would involve substituting triple terms for lower-order terms (singles and doubles).

The main result of the work reveals an important relation between the structure of the graph used in the epidemic model and the types of closures that are feasible. Let $G = \{V, E\}$ be a graph on N vertices with a set of edges, E . Consider a connected subset of vertices $F = \{v_1, v_2, \dots, v_k\} \subset V$ and assume that $\exists v_{i*} \in F$, a cut-vertex in G such that $F \setminus \{v_{i*}\}$ is partitioned into at least two disjointed components with vertices $F_1 = \{v_1, v_2, \dots, v_{i-1}\}$ and $F_2 = \{v_{i+1}, v_{i+2}, \dots, v_k\}$ belonging to any such two, distinct and disjointed components or subnetworks. Then the following equation holds:

$$\langle Z_{v_1} Z_{v_2} \dots Z_{v_{i-1}} S_{v_{i*}} Z_{v_{i+2}} \dots Z_{v_k} \rangle(t) = \frac{\langle Z_{v_1} Z_{v_2} \dots Z_{v_{i-1}} S_{v_{i*}} \rangle(t) \langle S_{v_{i*}} Z_{v_{i+2}} \dots Z_{v_k} \rangle(t)}{\langle S_{v_{i*}} \rangle(t)},$$

where in each Z_{v_i} term, Z may stand for any state symbol in our model (such as S for susceptible, P for protected and so on) for all vertices v_i that are not cut-vertices [9].

The authors further prove an impressively general result: if a graph with N vertices and E edges has T triangles and no larger loops than size 3 (meaning also that triangles cannot have overlapping edges), an upper bound on the size of the system of equations describing the system dynamics can be calculated [9]:

$$2N + 3E + 7T \leq 10N$$

The authors also provide a “recipe-like” approach to establish the feasibility of writing down an exact representation for a given graph even more generally. They use this to provide an upper bound for the number of equations required to describe epidemic dynamics exactly [9]:

$$N_{EQ}(G) = \sum_{i=1}^P m_i f_i - 2 \sum_{j=1}^L (\text{Ind}(v_{i_j}) - 1).$$

where P is the number of distinct sub-graphs produced when the original graph is spliced into sub-graphs through cut-vertices, m_i represents the number of equations required to describe the corresponding sub-graph i , f_i is the frequency or count of the sub-graph G_i and $\text{Ind}(v_{i_j})$ is the number of sub-graphs to which the cut-vertex v_{i_j} belongs [9].

In their concluding remarks, the authors recommend the implementation of an algorithmic approach to this equations generation with closure detection as the examples provided in their work were done by hand [9]. This is something we have been working on for several months and will discuss in Section 3.2.2.

2.4 Percolation Theory

Widely known and used in physics, statistics and mathematics, Percolation theory involves modelling scenarios as n -dimensional graphs, so our interest from a graph-theoretic perspective is not entirely unexpected. We now outline the relevant details of percolation theory and later, in Section 3.3, explain the utility we have found it may yield in extending and expanding modelling work from FIREFIGHTER.

In percolation, the edges between vertices can be either ‘open’ or ‘closed’ with probability p or $1 - p$ respectively. This is known as ‘bond’ percolation, as edges correspond to bonds in many of its applications - ‘site’ percolation is roughly analogous, where vertices (rather than edges) can be either ‘occupied’ or ‘unoccupied.’ We can think of percolation problems as liquid being poured onto a porous material and whether there is a path from hole to hole along open paths through the material. Note that removing (or closing) more and more edges moves us towards a critical point at which removing further edges would ensure the graph must fall apart into smaller clusters of vertices and edges that have no access to each other [4] - this is known as the *percolation threshold* or *critical probability*.

Formally, we consider a point lattice \mathbb{L} and denote the open cluster as $C(x)$, where $x \in \mathbb{L}$ is the local origin of the cluster. This cluster $C(x)$ is defined as the set of all vertices that can be reached from open paths beginning at the nucleation site, x . Then, we are particularly interested in the *percolation probability*:

$$\theta(p) = \mathbb{P}_p(|C(0)| = \infty),$$

and the *percolation threshold*:

$$p_c = \sup\{p \mid \theta(p) = 0\}.$$

where \mathbb{P}_p is the product measure given by $\mathbb{P}_p = \prod_{v \in \mathbb{L}^d} \mu_v$ and μ_v is the *Bernoulli measure*, which returns p when v is open and $1 - p$ when v is closed [10]. Analytically, others have shown that in the case of a two-dimensional regular point lattice, the critical probability is $p_c = 1/2$ [8].

3 Work done so far

3.1 Agency-oriented modelling

One of the key gaps in existing approaches to compartmental models of disease, as described in Section 2, is the lack of agency in individuals. By agency, we mean the following:

Definition 3.1. Agency. The ability of an individual to make decisions and choose to pursue them

In our context of disease modelling, agency might manifest in a modelled individual’s capacity to make decisions about mitigating their own risk of contracting the infection (given the extent to which they are able to do so). To address this, we have designed and implemented a compartmental graph model of infectious disease, which notably includes associating with each individual a ‘defence rating,’ which is a probability corresponding to their internal inclination towards self-protection (which also accounts for external constraints, for instance whether their circumstances mandate increased social contact). In this section, we first describe the implementation carried out so far before discussing the initial results obtained. We will then discuss the conference at which we presented some of these results and the significance of the approach employed. We end this section by discussing the implications to equations generation when we add a new state to a compartmental graph model and then the uses of Percolation Theory we have identified for usage in our current work over the next year.

3.1.1 Implementation of agency-based models

In our implemented model, we begin with a graph on a particular number of vertices and edges. This can be either given by the user or generated using code we have written for several random graph types.

We then assign an agent to each vertex. Agents can be in a number of states, such as ‘susceptible’ (could contract), ‘infected’ (currently has the infection and is infectious), ‘recovered’ (previously had

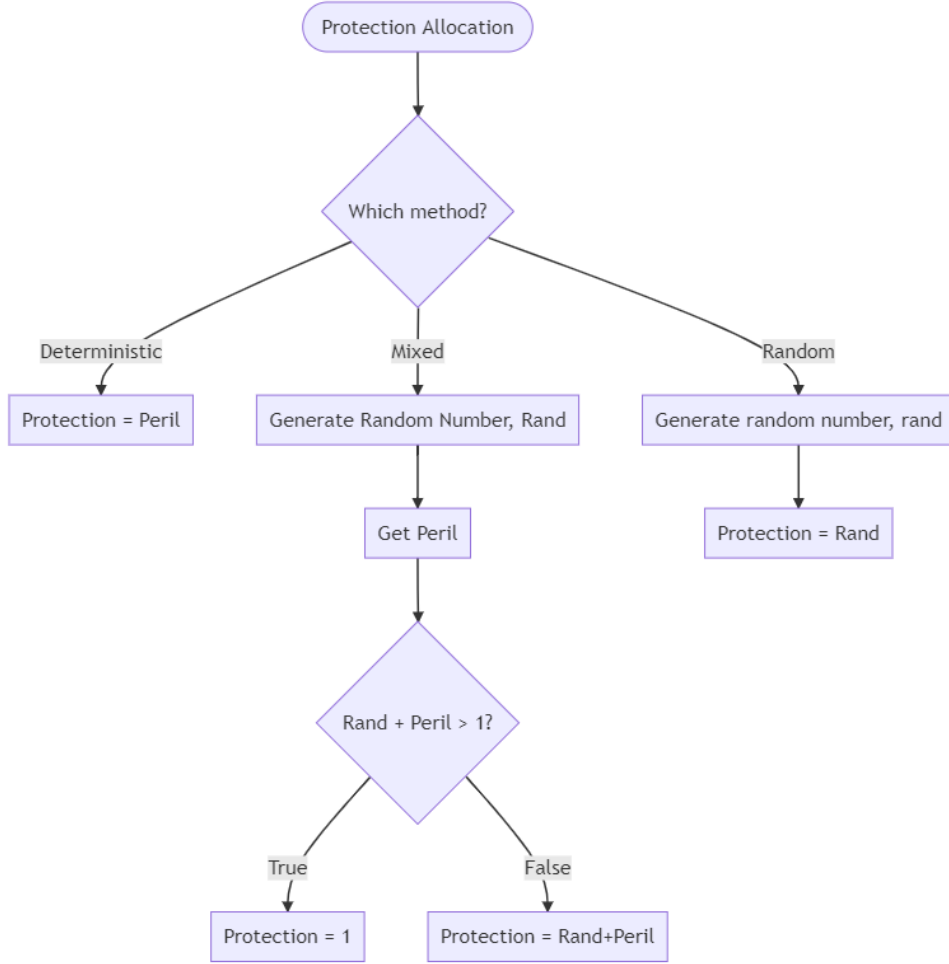


Figure 1: Visualisation of protection allocation - this is allocated either based on proximity to infection and updated at each turn, uniformly at random at initialisation or by combination of both.

the infection) and ‘protected’ (cannot contract the infection). The latter of these state is where much of our work has been focused: we have been studying the dynamics of the system with this state using the ‘protection ratings’ of vertices. We currently assign this rating in one of three ways: purely randomly, based upon proximity to infection or a combination of the two approaches.

In Figure 1, we see that the ‘random’ allocation currently entails assigning protection ratings uniformly at random once at the onset of the disease. These ratings do not change as the model progresses, which is in contrast to assignment based on proximity to infection where defence ratings are recalculated after each turn. Of course, this seems fairly simplistic but stands for something profound - this rating represents inclination towards self-protection, it is a reflection of simulated personality in individuals. If we wished to make this a more interesting assignment to better reflect a given population and context, we might look into assigning based on, for instance, normal or Poisson distributions. This is something that could be easily implemented based on the requirements of a given epidemic context as and when required.

Once these agents are initialised, a defensive move is made that is adapted to account for the inherent individual protection ratings. The current existing defence strategies are detailed in Figure 2. Currently, there are three main defence strategies that are deployed.

The defence strategies are represented visually in Figure 2. These strategies are:

- Defend based on highest degree, breaking ties on greatest proximity to infection;
- Defend based on greatest proximity to infection, break ties on highest degree; and
- Defend the agents who currently have the highest protection rating.

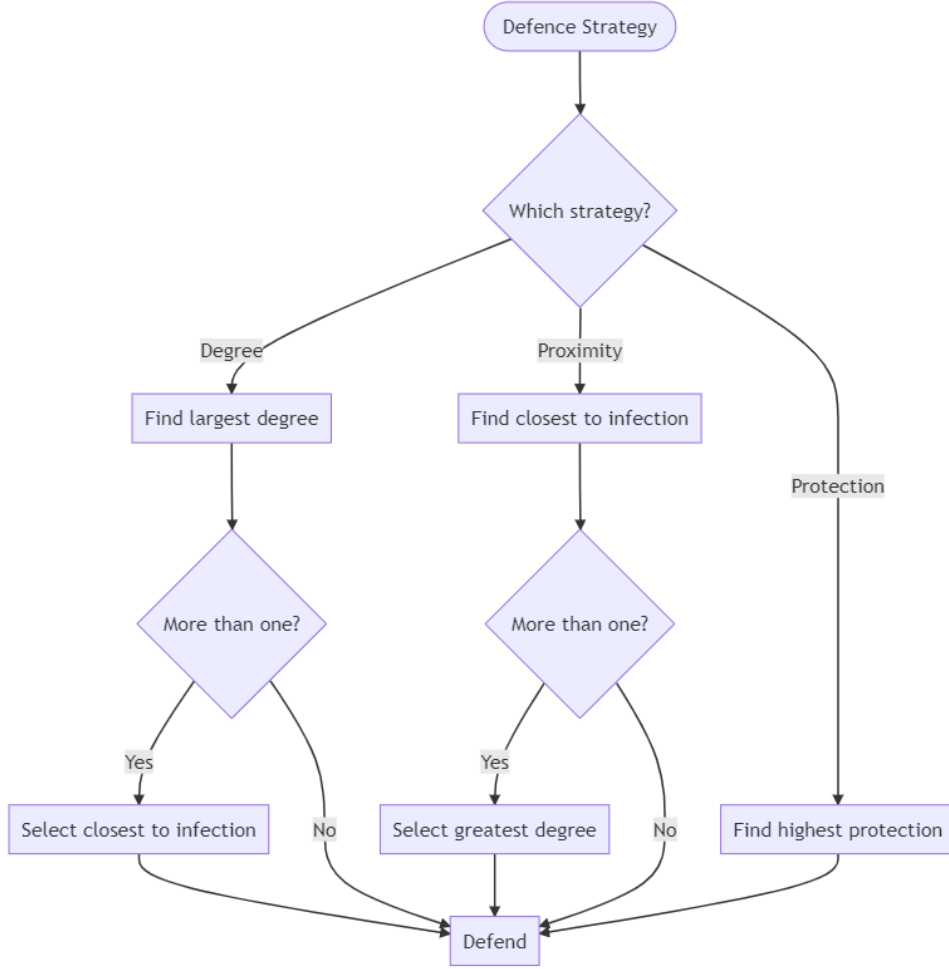


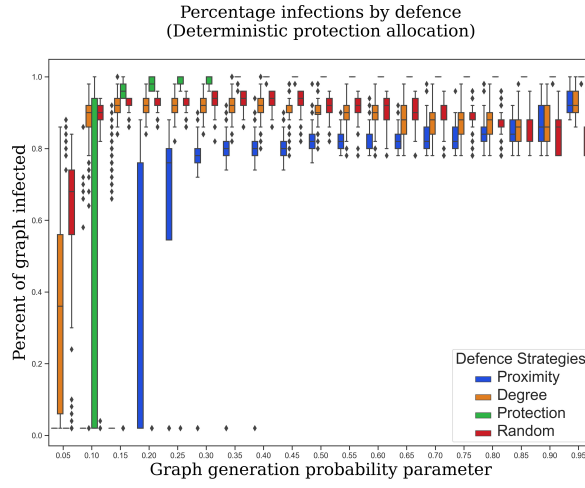
Figure 2: Representation of three implemented defence strategies.

The first two of these strategies are common in classic FIREFIGHTER - defending based on degree is effective in fairly dense graphs and defending based on proximity to fire is effective in sparse and tree-like graphs [2]. We note that all three of these strategies are heuristic methods and are effectively all greedy algorithms, the difference-maker is what exactly each strategy is greedy about (protection rating, proximity to infection or degree).

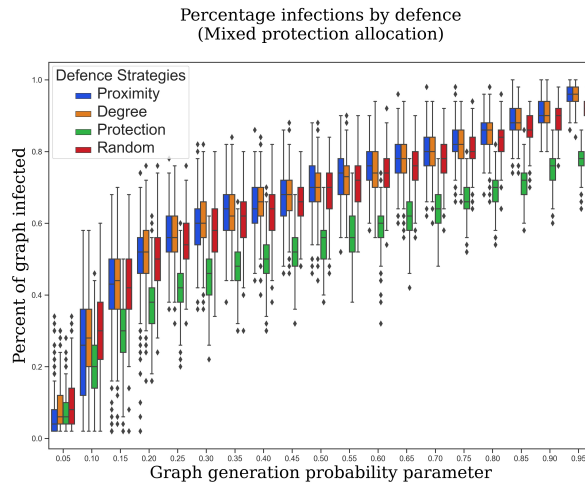
3.1.2 Experimental results

To gather data about defence strategy performance, we ran multi-graph experiments on various graph classes: in Figure 3, we see the results of such experiments on Erdős Rényi random graphs. For each value of p (the probability parameter for Erdős Rényi graph generation) between 0 and 1 in increments of 0.05, we generated 50 graphs on 50 vertices each. For each of these graphs, we started the contagion at each vertex of the graph in turn and stored the results of these simulations. We used this data to plot the percent of each graph infected under each defence strategy. We also include the results for a purely random defence strategy for comparison, which was also run on each graph. The plots produced are box plots, with a single box for each defence strategy in each value of p used. The box of each plot shows the interquartile range (from the 25th to the 75th percentile) with the median value indicated by a line in the box. The ‘whiskers’ of the boxes indicate the minimum and maximum in the range, with any outliers indicated by filled-in rhombi, where relevant.

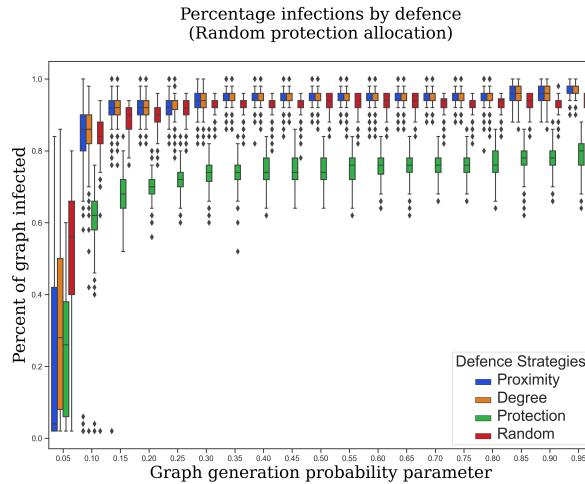
From these plots, we can see for mixed and uniformly random protection ratings, defence based on protection rating consistently outperforms the other strategies. We can see that, when we assign internal protection purely based on proximity to infection for each agent, defence based on proximity outperforms the other strategies. We would otherwise expect this strategy to perform identically to



(a) Protection ratings are allocated based on proximity to closest infection.



(b) Protection ratings allocated based on a baseline random number which is increased based on proximity to closest infection.



(c) Protection ratings allocated uniformly at random.

Figure 3: Charts showing a comparison of percentage of Erdős Rényi graphs infected when defended by three different defence strategies (and a random defence for comparison) with probability parameter p varied between 0 and 1 in increments of 0.05 for three different agent protection rating allocation methods.

defence based on protection: this is not the case since the protection defence does not break ties, which defence on proximity does (here, we break ties on degree). Further, the other strategies sometimes do not even perform as well as random defence - this would suggest these, in our extended formulation of the game, are not as worth pursuing as our novel protection-based strategy, particularly for very dense graphs.

3.1.3 Computational Biology Conference

On 27th May 2021, we presented a talk on agency-oriented modelling to the First University of Glasgow Computational Biology Conference. The title of the talk was *Introducing features of agency into computational models of infectious disease*. The key points of discussion were:

- Current computational approaches to graph models of disease
- Extending existing graph models to better account for individual agency
- Introducing a compartmental modelling approach

Following this talk, we have been contacted by several students and academics interested in the work presented. For some of these individuals, we have arranged further discussion to establish potential areas for collaboration.

3.2 Graph-based compartmental models

In this section, we will detail and explain the work we have done in extending the work outlined in Section 2.3. We begin by exploring the addition of a new compartmental state to a graph-based *SIR* model and then we determine the impact this has on the total system of equations describing these models. The main focus of our work so far in this area has been to develop an algorithmic approach for the work detailed in [9] (discussed in Section 2.3), which the authors of that work left as an open problem.

3.2.1 Adding a new state to the model

Let ζ_i be the probability that we defend individual i and let α_i represent the efficacy of the protection measure for individual i , which may decay over time and vary from person to person. Using these rates of protection and effectiveness, for fixed population size the differential equations become:

$$\begin{aligned}\langle \dot{S}_i \rangle &= \alpha_i \langle P_i \rangle - \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle - \zeta_i \langle S_i \rangle \\ \langle \dot{I} \rangle &= \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I \rangle \\ \langle \dot{R}_i \rangle &= \gamma_i \langle I \rangle \\ \langle \dot{P}_i \rangle &= \zeta_i \langle S_i \rangle - \alpha_i \langle P_i \rangle.\end{aligned}$$

Importantly, we now need to examine the effect that addition of this state has on the overall system of equations. As an example, we consider the ‘triangle network’ - a loop of three nodes. The equations required to precisely express the system *SIR* dynamics of this network are as follows [9]:

$$6 \text{ singles: } \langle \dot{S}_1 \rangle, \langle \dot{S}_2 \rangle, \langle \dot{S}_3 \rangle, \langle \dot{I}_1 \rangle, \langle \dot{I}_2 \rangle, \langle \dot{I}_3 \rangle. \quad (1)$$

$$6 \text{ doubles: } \langle \dot{S}_1 I_2 \rangle, \langle \dot{I}_1 S_2 \rangle, \langle \dot{S}_1 I_3 \rangle, \langle \dot{I}_1 S_3 \rangle, \langle \dot{S}_2 I_3 \rangle, \langle \dot{I}_2 S_3 \rangle. \quad (2)$$

$$6 \text{ triples: } \langle \dot{S}_1 I_2 I_3 \rangle, \langle \dot{S}_1 I_2 S_3 \rangle, \langle \dot{S}_1 S_2 I_3 \rangle, \langle \dot{I}_1 S_2 S_3 \rangle, \langle \dot{I}_1 I_2 S_3 \rangle, \langle \dot{I}_1 S_2 I_3 \rangle. \quad (3)$$

Now, using the equations for the *SIRP* model, we have the following equation requirements:

$$\begin{aligned}
& 9 \text{ singles: } (1) \text{ and } \langle \dot{P}_1 \rangle, \langle \dot{P}_2 \rangle, \langle \dot{P}_3 \rangle. \\
& 18 \text{ doubles: } (2) \text{ and } \langle S_1 \dot{P}_2 \rangle, \langle P_1 \dot{S}_2 \rangle, \langle I_1 \dot{P}_2 \rangle, \langle P_1 \dot{I}_2 \rangle, \langle S_1 \dot{P}_3 \rangle, \langle P_1 \dot{S}_3 \rangle, \\
& \quad \langle I_1 \dot{P}_3 \rangle, \langle P_1 \dot{I}_3 \rangle, \langle S_2 \dot{P}_3 \rangle, \langle P_2 \dot{S}_3 \rangle, \langle I_2 \dot{P}_3 \rangle, \langle P_2 \dot{I}_3 \rangle. \\
& 24 \text{ triples: } (3) \text{ and } \langle S_1 \dot{S}_2 P_3 \rangle, \langle S_1 \dot{P}_2 S_3 \rangle, \langle S_1 \dot{I}_2 P_3 \rangle, \langle S_1 \dot{P}_2 I_3 \rangle, \langle S_1 \dot{P}_2 P_3 \rangle, \langle I_1 \dot{S}_2 P_3 \rangle, \\
& \quad \langle I_1 \dot{P}_2 S_3 \rangle, \langle I_1 \dot{I}_2 P_3 \rangle, \langle I_1 \dot{P}_2 I_3 \rangle, \langle I_1 \dot{P}_2 P_2 \rangle, \langle P_1 \dot{S}_2 S_3 \rangle, \langle P_1 \dot{S}_2 I_3 \rangle, \\
& \quad \langle P_1 \dot{I}_2 S_3 \rangle, \langle P_1 \dot{I}_2 I_3 \rangle, \langle P_1 \dot{P}_2 I_3 \rangle, \langle P_1 \dot{P}_2 S_3 \rangle, \langle P_1 \dot{S}_2 P_3 \rangle, \langle P_1 \dot{I}_2 P_3 \rangle.
\end{aligned}$$

Note that the reason we dispense with the cases of all three vertices being in the same state is that this would not result in any dynamics - no vertices would ever change state in this case - although this may be required in the use of closures in some circumstances. We hope to prove, in the coming year, that the addition of a singular state to the model results in a linear increase in the required number of equations.

3.2.2 Implementation

We have been working on an implementation of the above work in equations generation. Our goal is to produce code that can accept a particular graph, for instance as a CSV file, and determine the number of equations that could be required to fully describe a compartmental model (for instance, *SIR* or *SIRP*) on that graph. This could be by providing upper and lower bounds if the exact graph structure is unknown, or by providing the exact number of equations if the graph is known. With this information, the user can request the software returns the full list of differential equations that exactly describe the epidemic dynamics using an algorithmic approach to generation.

3.3 Uses of Percolation

We have identified two main avenues that we would like to explore further regarding the use of Percolation Theory in *FIREFIGHTER*: the firefighter may use percolation in order to defend the graph, or the fire might spread with percolation probability p . We will now discuss these two possible avenues and how we hope to use them in our research.

3.3.1 Better than random

One potential use of percolation is a baseline test. In most scenarios, a method for obtaining defence strategies should be at least as effective as a random defence sequence. We could find such a random sequence by allowing the firefighter to defend in a percolation-like strategy for comparative purposes. Consider a sequence of vertices in graph G , written as d_1, d_2, \dots, d_t . An optimal defence sequence could be found using integer programming as provided by Finbow and MacGillivray [2]:

$$\begin{aligned}
& \text{Maximise} && \sum_{v \in V(G)} d_v w(v) && \text{for each level } i \\
& \text{subject to} && d_v + \sum_{\text{level}(v)=i} d_v \leq 1 && \text{for each level } i \\
& && d_v + \sum_{u \succ v} d_u \leq 1 && \text{for every outer vertex } v \text{ of } T, \\
& && d_v \in \{0, 1\}.
\end{aligned}$$

where $u \succ v$ indicates that u is an ancestor of v . The optimal strategies provided for different classes and densities of graphs here will provide an upper bound (which may indeed be impossible to attain in some cases) for success of a given strategy. We can find a lower bound using percolation, and so we have a range of success values as a starting point: if some strategy is better than random percolation, then it is worth considering, but below the particular expected optimal solution from integer programming and we can, in principle, improve or find a better strategy.

We conjecture that, at the lowest graph densities, the random strategy will be close to the optimal strategy and thus finding an improvement is at once difficult and lacking in great utility. At the very highest graph densities, random strategies will have a very low expected best-case scenario but so will most strategies, since the constraint on the firefighter that they have only one vertex to save per turn does not go as far towards fire containment per turn compared to sparser graphs.

3.3.2 Reproduction rate

We might also consider the fire spread being determined by percolation (rather than the firefighter's defence sequence). Diseases, when there is a large enough sample size, have a basic reproduction rate associated with them, denoted R_0 : for instance, measles has a basic reproduction rate $12 \leq R_0 \leq 18$ [5] and the influenza strain responsible for the 1918 pandemic has a basic reproduction rate of $1.4 \leq R_0 \leq 2.8$ [1]. These baseline, theoretical values can be implemented as an internal probability to a propagating fire: to formulate a stochastic version of FIREFIGHTER, we let the fire propagate with some probability (which could be determined by reproduction rate of a real infectious disease) in a percolation-like process and examine the change to optimal defence.

Where we wish to consider vertices as individuals and edges as the connections between them, percolation may give us a more useful model for disease spread when we do not assume the population is well mixed and instead introduce probability functions to correspond to the likelihood one vertex is connected to another.

4 Research Plan

4.1 Goals

Over the next year, we hope to extend the two strands of work we have been pursuing with a number of goals and milestones in mind. In the more formal, mathematical side of our work, we aim to achieve the following:

- Formalise a probabilistic, agency-based graph model of disease that can be implemented in the PRISM model checking language.
- Complete and verify the equations generation work that we have started over the last year and write down the algorithmic approach used.
- Explore further formal approaches to agency-based modelling that could be used to formulate models or understand better our existing models (for instance, integer programming approaches or model-checking using PRISM).

Moreover, in the experimental side of our work, we aim to achieve the following:

- Further generalise the current implementation to accept further parameters and output requested results.
- Implement, test and explore further defence strategies in the implementation, using this to reason about optimal defence candidates in different contexts.
- Write down a series of (provable) statements regarding defence strategies in this implementation on various graph classes.
- Use real-world data to model different epidemic scenarios and contextualise defence strategies.

Some significant milestones we aim to reach are as follows:

- Submit a conference abstract or poster.
- Write and submit a paper to a journal or conference.

- Request, write and present a talk for the FATA research section in the department in a post-graduate slot in the Autumn of 2021.

Of course, the first two of these milestones are very general - we are in the process of examining potential candidate venues for presentation of this area of work and we would be grateful for any advice on this.

4.2 Timeline

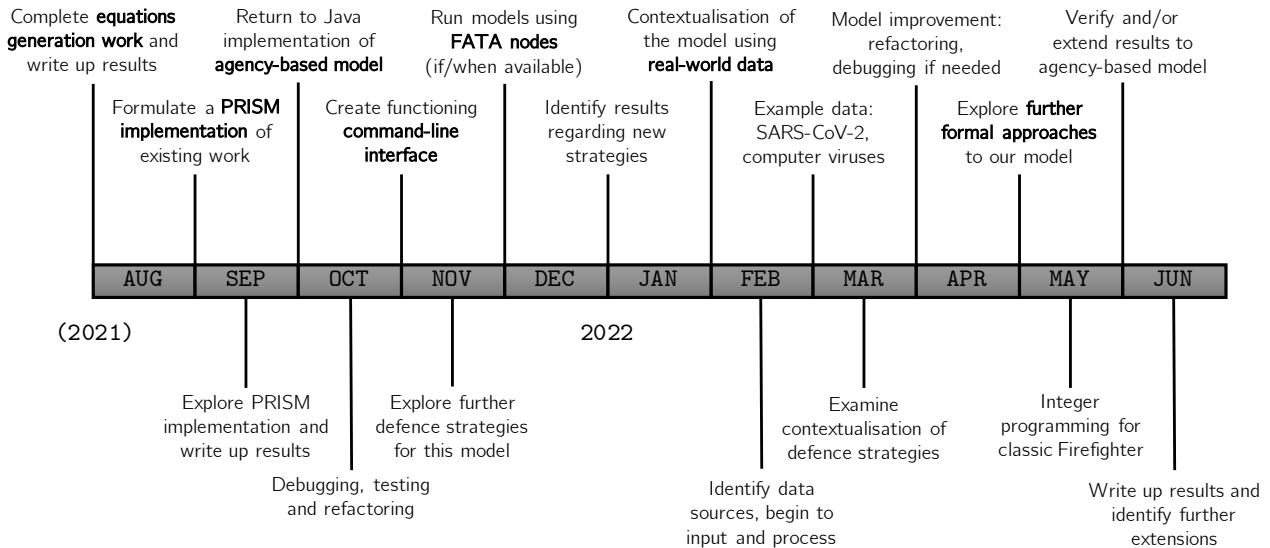


Figure 4: Timeline of work for the next year

References

- [1] N. M. FERGUSON, D. A. CUMMINGS, C. FRASER, J. C. CAJKA, P. C. COOLEY, AND D. S. BURKE, *Strategies for mitigating an influenza pandemic*, Nature, 442 (2006), pp. 448–452.
- [2] S. FINBOW AND G. MACGILLIVRAY, *The firefighter problem: A survey of results, directions and questions*, The Australasian Journal of Combinatorics, 43 (2009).
- [3] S. M. FIRESTONE, Y. HAYAMA, R. BRADHURST, T. YAMAMOTO, T. TSUTSUI, AND M. A. STEVENSON, *Reconstructing foot-and-mouth disease outbreaks: a methods comparison of transmission network models*, Scientific Reports, 9 (2019), p. 1.
- [4] G. GRIMMETT, *Percolation*, Springer-Verlag, Berlin, 2 ed., 1999.
- [5] 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, 17 (2017), pp. 420–428.
- [6] B. L. HARTNELL, *Firefighter! an application of domination*, in 25th Manitoba Conference on Combinatorial Mathematics and Computing, University of Manitoba in Winnipeg, Canada, 1995.
- [7] H. HETHCOTE, *The mathematics of infectious diseases*, SIAM Review, 42 (2000), pp. 599–653.
- [8] H. KERSTEN, *The critical probability of bond percolation on the square lattice equals $1/2$* , Communications in mathematical physics, 74 (1980), pp. 41–59.
- [9] I. Z. KISS, C. G. MORRIS, F. SÉLLEY, P. L. SIMON, AND R. R. WILKINSON, *Exact deterministic representation of markovian SIR epidemics on networks with and without loops*, Journal of Mathematical Biology, 70 (2014), pp. 437–464.
- [10] A. KLENKE, *Probability Theory*, Springer, Berlin Heidelberg, 2 ed., 2014.
- [11] J. D. O'BRIEN, I. K. DASSIOS, AND J. P. GLEESON, *Spreading of memes on multiplex networks*, New Journal of Physics, 21 (2019).

ETHAN KELLY

`e.kelly.1@research.gla.ac.uk`

PRINCIPAL SUPERVISOR: JESSICA ENRIGHT

`Jessica.Enright@glasgow.ac.uk`

SECOND SUPERVISOR: ALICE MILLER

`Alice.Miller@glasgow.ac.uk`

DEPARTMENT OF COMPUTING SCIENCE
SIR ALWYN WILLIAMS BUILDING
UNIVERSITY OF GLASGOW
G12 8QN