COVID-19: ANALYTICS OF CONTAGION ON INHOMOGENEOUS RANDOM SOCIAL NETWORKS

T. R. HURD

Mathematics & Statistics, McMaster University, 1280 Main St. West
Hamilton, Ontario, L8S 4L8, Canada
hurdt@mcmaster.ca

April 7, 2020

Abstract

Motivated by the need for novel robust approaches to modelling the Covid-19 epidemic, this paper treats a population of N individuals as an inhomogeneous random social network (IRSN). The nodes of the network represent different types of individuals and the edges represent significant social relationships. An epidemic is pictured as a contagion process that changes daily, triggered on day 0 by a seed infection introduced into the population. Individuals' social behaviour and health status are assumed to be random, with probability distributions that vary with their type. First a formulation and analysis is given for the basic SI ("susceptible-infective") network contagion model, which focusses on the cumulative number of people that have been infected. The main result is an analytical formula valid in the large N limit for the state of the system on day t in terms of the initial conditions. The formula involves only one-dimensional integration. Next, more realistic SIR and SEIR network models, including "removed" (R) and "exposed" (E) classes, are formulated. These models also lead to analytical formulas that generalize the results for the SI network model. The framework can be easily adapted for analysis of different kinds of public health interventions, including vaccination, social distancing and quarantine. The formulas can be implemented numerically by an algorithm that efficiently incorporates the fast Fourier transform. Finally a number of open questions and avenues of investigation are suggested, such as the framework's relation to ordinary differential equation SIR models and agent based contagion models that are more commonly used in real world epidemic modelling.

Key words: Social network, epidemic models, random network, Watts network model, social systemic risk, Poisson random graphs, locally tree-like, SIR models.

MSC: 05C80, 91B74, 91G40, 91G50

1 Introduction

This paper provides a framework for analyzing the propagation of an infectious disease such as COVID-19 on an *inhomogeneous random social network (IRSN)*. The framework starts with a so-called *inhomogeneous random graph (IRG)* (henceforth called the *skeleton*) whose nodes represent people with a finite number of types, interconnected by edges representing random social contacts. On this skeleton is defined a random collection of individual immunity buffers (that measure individuals' resistance to the disease) and bilateral (person-to-person) exposures that quantify the viral load transmitted by infected individuals to their social contacts. An infection introduced randomly into the population of susceptible individuals will cause a sequence of contagion shocks that will be thought of as iterations of a *cascade mapping* or *cascade mechanism*.

The main contributions of this paper are:

1. Introduction of the *inhomogeneous random social network (IRSN) framework* that provides a flexible and scalable architecture for modelling complex network characteristics. In particular, we will develop infection cascade models for networks of individuals classified by arbitrary types.

2. The large N asymptotics for infection cascades in IRSN models is developed, yielding explicit and efficiently computable recursive probabilistic formulas for the daily update of the state of the system, in particular, the day-by-day changes in the fraction of the individuals of each type that are susceptible, infected or recovered.

3. We provide details of how the contagion analytics can in principle be used to provide large scale investigations into potential policy interventions that one might invoke to mitigate or suppress the progress of the contagion.

4. Overall, the intent of this new framework is to provide a *purely analytical toolkit* for networks, capable of handling thousands of different types of individuals, that can run on a laptop. The network framework is capable of providing much faster results, with a similar degree of accuracy, than is possible with large-scale agent-based epidemic models normally used for informing health policy.

Studying the spread of infectious diseases using the tools of network science has a substantial literature, reviewed for example in Keeling and Eames (2005). The book Newman (2010) provides a broad overview of networks in all areas of science, including applications to epidemic modelling, while Pellis et al. (2015) explores current challenges in network epidemic models. The framework developed here is a variation of the network cascade model of Watts (2002), generalized to allow for random edge weights as in Hurd and Gleeson (2013).

IRSN models for any value of N can always be explored by pure simulation, as in agent-based modelling. Alternatively, sequences of IRGs parametrized by increasing N have an important property called *locally tree-like (LT)*. As described by Aldous and Steele (2004) and others, this property means that the random graph sequence is "locally weakly convergent" as $N \to \infty$ to a connected Galton-Watson random tree, leading to a host of simplifications that are described by *percolation theory on random graphs*. The LT property of IRGs implies for example that for any $k > 1$, the density of cycles of length k in the graph goes to zero as N goes to infinity. The heuristic large N cascade arguments at the core of this paper are related to rigorous results in the literature surveyed by van der Hofstad (2016) that connect percolation properties on random graphs to properties of Galton-Watson processes.

Section 2 of the paper introduces *inhomogeneous random social networks* (IRSNs) and defines the basic SI infection cascade mechanism on such networks. Section 3 explores the large N analytical properties of the model, including the large scale "locally tree-like" structure of the skeleton. It then focusses on the recursive characterization of the stochastic infection cascade mapping in the $N = \infty$ limit. Section 4 discusses how the network approach to epidemiology fits in with two other approaches to infectious disease modelling: SI compartment models and agent based models. Section 5 explores how the basic SI network model can easily be extended to analogues of the SIR and SEIR compartment models. Section 6 provides a numerical implementation of the SI model, showing the flop count for computing a daily update to be $O(M^2 \times \text{Nfft})$ where M is the number of types and Nfft is the number of lattice points in each one-dimensional integration. Section 7 addresses the issue of calibrating ISRN models to real health and social data. In Section 8, we explore a simple illustration of how the method can be used to understand potential policy interventions to protect the residents of a seniors' centre while a pandemic rages in the community outside. Finally, a concluding

section discusses some possible next steps for having a better understanding of contagion risk in inhomogeneous random social networks.

Notation:

1. For a positive integer N, $[N]$ denotes the set $\{1, 2, \ldots, N\}$.

2. For a random variable X, its cumulative distribution function (CDF), probability density function (PDF) and characteristic function (CF) will be denoted $F_X, \rho_X = F'_X$, and $\hat{f}_X$ respectively. Note that $\hat{f}_X = \mathcal{F}(\rho_X)$ where $\mathcal{F}$ denotes the Fourier transform.

3. For any event A, $\mathbf{1}(A)$ denotes the indicator random variable, taking values in $\{0, 1\}$.

4. Any collection of random variables $X = (X_1, X_2, \ldots)$ generates a sigma-algebra (or informally "information set") denoted by $\sigma(X)$.

5. Landau's "big O" notation $f^{(N)} = O(N^\alpha)$ for some $\alpha \in \mathbb{R}$ is used for a sequence $f^{(N)}, N = 1, 2, \ldots$ to mean that $f^{(N)} N^{-\alpha}$ is bounded as $N \to \infty$.

6. The L^2 Hermitian inner product of two complex valued functions $f(x), g(x)$ on a domain D is defined to be $\left\langle f, g \right\rangle_{L^2(D)} = \int_D \; f^*(x) \; g(x) \; \mathrm{d}x$. The L^2 norm of a function $f(x)$ on a domain D is defined to be $\|f\|_{L^2(D)} = \langle f, f \rangle^{1/2}_{L^2(D)}$.

2 SI Infection Cascades on IRSNs

This section provides the core modelling assumptions of the framework, in a simplified *susceptible-infected (SI)* setting in which infected individuals never recover from the disease, and continue to infect susceptible individuals indefinitely. More realistic infection mechanisms will be considered in later sections.

A social network represents a population of individuals as *nodes* of a graph, whose *undirected edges* represent the existence of a substantial social connection at a moment in time. Our network setting for the spread of an infectious disease begins with a number of preliminary assumptions:

1. The population is classified into a large but finite disjoint collection of "types" that represent people's important attributes, such as age, gender, living arrangement, profession, country and location.

2. In this paper, the network of social contacts, initially random, is taken to be constant during the epidemic; the mathematical methods adapt to time-varying networks.

3. At the start of the outbreak on day 0, most of the population is susceptible ("S"), but a small "seed" of infected individuals is placed in the network.

4. Infected individuals pass on a random viral load to each of their social contacts.

5. An individual's state of health is represented by a random "immunity buffer". Over time they experience an accumulation of viral load; if the total viral load exceeds their buffer, they become infected "I".

The social system at any moment in time will be represented as an inhomogeneous random social network, or IRSN. This is the specification of a multidimensional random variable that captures two levels of structure. The primary level of the IRSN, called the *skeleton graph*, is an undirected random graph with N nodes labelled by $v \in [N]$, representing people of a wide variety of types, and where an un directed edge labelled by $(vw) \in [N] \times [N]$ represents the existence of a significant social interconnection between w and v, such as a family relationship. The secondary layer specifies the *health and mutual exposures* of people, conditioned on knowledge of the skeleton graph.

Inhomogeneity in the IRSN model arises through classifying people by a finite (possibly very large) number of types that can include a wide range of attributes. The collection of random types $\mathcal{T} := \{T_v\}_{v\in[N]}$ will be assumed to completely determine the dependence structure of the remaining random variables. In other words, the remaining random variables will exhibit conditional independence with respect to the sigma-algebra or "information set" $\sigma(\mathcal{T}) := \sigma(T_v, v \in [N])$.

2.1 Skeleton Graph

The skeleton graph is modelled as an undirected inhomogeneous random graph (IRG), generalizing Erdös-Renyi random graphs, in which edges are drawn independently between unordered pairs of nodes, not with equal likelihood but with likelihood that depends on their types. This class has its origins in Chung and Lu (2002) and has been studied in generality in Bollobás et al. (2007) and the textbook van der Hofstad (2016). The IRG structure arises by the assumption that edge indicators are Bernoulli random variables $I_{vw} = I_{wv}$ defined for unordered pairs of individuals (v, w), that are independent conditioned on the assignment of node types.

Assumption 1 (Skeleton Graph)**.** *The primary layer of an IRSN, namely the skeleton graph* $\mathrm{IRG}(\mathbb{P}, \kappa, N)$*, is an inhomogeneous random graph with* N *nodes labelled by* $v \in [N]$*. It can be defined by two collections* $\mathcal{T}, \mathcal{I}$ *of random variables* $T_v, v \in [N]$ *and* $I_{vw}, v, w \in [N]$*, with sigma-algebras ("information sets")* $\sigma(\mathcal{T}), \sigma(\mathcal{I})$*.*

1. *Each node* $v \in [N]$*, representing a person, has type* T_v *drawn independently with probability* $\mathbb{P}(T)$ *from a finite list of* types $[M]$ *of cardinality* $M \geq 1$*.*

2. *Each edge* $(v, w) \in [N] \times [N]$ *corresponds to a non-zero entry of the symmetric* incidence matrix I*. For each pair* (v, w)*,* $I_{vw} = I_{wv}$ *is the indicator for* w *to be (significantly) socially connected to* v*. Conditioned on the vector of all types* $\mathcal{T}$*, the collection of edge indicators* $\mathcal{I} := \{I_{vw}\}$ *is an independent family of Bernoulli random variables with probabilities*

$$\mathbb{P}[I_{vw} = 1 \mid \sigma(\mathcal{T})] := \mathbb{P}[I_{vw} = 1 \mid T_v, T_w] = (N-1)^{-1} \kappa(T_v, T_w) \mathbf{1}(v \neq w) \ . \qquad (1)$$

Here $\kappa : [M]^2 \to [0, \infty)$, the *probability mapping kernel*, determines the likelihood that two people v, w of the given types have a social connection, or edge. The assumed N dependence assures sparseness of the graph for large N, and for consistency we require that $N - 1 \geq \max_{T,T'} \kappa(T, T')$.

2.2 Health, Transmission and the Epidemic Trigger

The additional fundamental assumption of the IRSN modeling framework is that the relevant health attributes of all people are summarized by an independent collection of multivariate random variables, *conditioned on the skeleton*. Essentially, we will assume: (i) that each individual has a random "immunity buffer", and (ii) in case they are infected, a random viral load will be transmitted to each of their social contacts. Finally, we assume that as soon as a person's cumulative viral exposure exceeds their buffer, then they become infected and infective.

Definition 1. 1. *The* initial immunity buffer $\bar{\Delta}_v$ *of node* v *prior to the crisis is a non-negative value that represents the resistance of that person to the virus.*

2. *The* nominal exposure pair *between* w *and* v *is a pair denoted by* $(\Omega_{vw}, \Omega_{wv})$ *of positive values:* Ω_{vw} *represents the total viral load transmitted from* v *to* w *should* v, w *be connected (i.e. if* $I_{vw} = 1$*), and if* v *is infected.*

3. *The health state of the network before the onset of the outbreak is determined by the collection of conditionally independent random variables* Ω_{vw} *and* $\bar{\Delta}_v$.

4. *Each person becomes infected at the first time* t *that* $\Delta_v = 0$.

Initial values are considered just prior to the outbreak. The epidemic trigger on day 0 introduces a number of infected individuals in the population:

Definition 2. *An* epidemic trigger *at a moment in time, which we label by day* $t = 0$*, occurs when for each* T*, a specified* seed fraction $\Pi^{(0)}(T) \in [0, 1]$ *of all type* T *individuals are infected.*

Now we make some pragmatic probabilistic assumptions about the initial buffer and exposure random variables, conditioned on the vector of individual types $\mathcal{T} = T_{vv\in[N]}$.

Assumption 2 (Immunity Buffers and Exposures)**.** *The secondary layer of an IRSN, the collection of initial immunity buffers and potential exposures* $\bar{\Delta}_v, \Omega_{vw}$ *are continuous non-negative random variables that are mutually independent, and independent of* $\mathcal{I} = \{I_{vw}\}$*, conditioned on* $\mathcal{T} = T_v$.

1. *For each individual* v*,* $\bar{\Delta}_v$ *conditioned on* $T_v = T \in [M]$ *has a continuous density* $\bar{\rho}_\Delta(x|T) = \bar{F}'_\Delta(x|T)$ *supported on* $\mathbb{R}_+$*. Thus the cumulative distribution function (CDF) is*

$$\bar{F}_\Delta(x|T_v) := \mathbb{P}(\bar{\Delta}_v \leq x \mid T) = \int_0^x \bar{\rho}_\Delta(y|T)\mathrm{d}y \quad , \forall x \in [0, \infty) \cup \{\infty\} \ . \tag{2}$$

2. *Immediately after the trigger with initial seed probabilities* $\Pi^{(0)}(T)$*,*

$$F^{(0)}_\Delta(x|T_v) := \mathbb{P}(\Delta^{(0)}_v \leq x \mid T_v) = \Pi^{(0)}(T) + (1 - \Pi^{(0)}(T))\bar{F}_\Delta(x|T_v) \ . \tag{3}$$

3. *The* initially infected individuals *are those with* $\Delta^{(0)}_v = 0$.

4. *For each edge* vw*,* Ω_{vw} *and* Ω_{wv} *are a pair of random variables. Conditioned on* $T_v = T, T_w = T'$*,* Ω_{vw} *has a continuous marginal density* $\rho_\Omega(x \mid T_v, T_w)$ *supported on* $\mathbb{R}_+$.

In summary, a finite IRSN representing the system after a crisis trigger amounts to a collection of random variables $\{T, I, \Delta^{(0)}, \Omega\}$ satisfying Assumptions 1 and 2.

2.3 SI Transmission Mechanism

We now consider how such IRSNs will evolve on a day-to-day basis when a *trigger infection* occurs at time $t = 0$. Recall that each person becomes infected at the first time t that $\Delta_v = 0$.

The infection state of each individual at day t will be identified by the *infection indicator* random variable defined by

$$\mathcal{D}_v^{(t)} = \mathbf{1}(\Delta_v^{(t)} \leq 0) \ , \tag{4}$$

that takes values either 0 ("susceptible") and 1 ("infected"). The infection state of individual w at day t now influences the infection shock transmitted to another individual v:

$$S_{wv}^{(t)} := I_{wv}\Omega_{wv}\mathcal{D}_w^{(t)} \ . \tag{5}$$

The aggregated infection shock transmitted to v is:

$$S_v^{(t)} := \sum_{w \neq v} S_{wv}^{(t)} \ , \tag{6}$$

and the impacted immunity buffer of v on day $t+1$ is:

$$\Delta_v^{(t+1)} = \Delta_v^{(0)} - \sum_{w \neq v} S_{wv}^{(t)} \ . \tag{7}$$

Putting (4, 5, 6, 7) together gives the complete infection mapping at day $t \geq 0$.

3 Analytics of IRSN models

The IRSN framework just introduced specifies the joint distributions of the random variables $\{T, I, \Delta^{(0)}, \Omega^{(0)}\}$, thereby providing a compact stochastic representation of the state of a given real world network of N individuals at the moment an outbreak is triggered. The same distributional data defines a sequence of random networks with varying N. As we will see in this section, the so-called *locally tree-like* property of the IRG skeleton has very important analytical implications in the limit $N \to \infty$.

3.1 Degree Distribution of the Skeleton Graph

The distribution of the number of social contacts of nodes in IRGs, in other words their *degree distributions*, has a natural Poisson mixture structure in the large N limit. By permutation

symmetry, one only needs to consider individual 1 with arbitrary type $T_1 = T$, whose degree is defined as $\mathrm{d}_1 = \sum_{w=2}^{N} I_{w1}$, a sum of conditionally IID random variables. Since $e^{ikI_{w1}} = 1 + I_{w1}(e^{ik} - 1)$, each term has the identical conditional characteristic function (CF)

$$\mathbb{E}^{(N)}[e^{ikI_{w1}} \mid T_1 = T] = \sum_{T' \in [M]} \mathbb{P}(T') \left(1 + (N-1)^{-1} \kappa(T, T')(e^{ik} - 1)\right) \ . \tag{8}$$

The conditional CF of d_1 is the $N-1$ power of this function, and can be written

$$\mathbb{E}^{(N)}[e^{ik\mathrm{d}_1} \mid T] = \left[1 + \frac{1}{N-1} \sum_{T'} \mathbb{P}(T') \kappa(T, T')(e^{ik} - 1) + O(N^{-2})\right]^{N-1} \ , \tag{9}$$

to display its asymptotic structure as $N \to \infty$.

Proposition 1. *The characteristic function of the degree* d_v *of an individual* v*, conditioned on its type* $T \in [M]$*, is* 2π*-periodic on* $\mathbb{R}$ *and has the* $N \to \infty$ *limiting behaviour:*

$$\hat{f}^{(N)}(k \mid T) \ = \ \hat{f}(k \mid T) \ \left(1 + O(N^{-1})\right) \ , \tag{10}$$

$$\hat{f}(k \mid T) \ := \ \exp\left[\lambda(T)(e^{ik} - 1)\right] \ , \tag{11}$$

where $\lambda(T) = \sum_{T'} \mathbb{P}(T') \kappa(T, T')$*. Here, convergence of the logarithm of* (10) *is in* $L^2([0, 2\pi])$*.*

This type of limit can be handled by Lemma 2, stated and proved in the Appendix.

Proposition 1 tells us that for different values of T, the conditional degree distribution is asymptotic to a Poisson distribution with mean parameter $\lambda(T) = \sum_{T'} \lambda(T, T')$ where $\lambda(T, T') = \mathbb{P}(T') \kappa(T', T)$. Now, recall that a *finite mixture* of a collection of probability distribution functions is the probability distribution formed by a convex combination. Thus the asymptotic *unconditional* degree distribution of any individual is a finite mixture with characteristic function:

$$\hat{f}(k) = \sum_{T \in [M]} \mathbb{P}(T) \hat{f}(k \mid T) \ . \tag{12}$$

Each mixture component has a Poisson distribution with Poisson parameters $\lambda(T)$ and the mixing variable is the individual-type T with mixing weight $\mathbb{P}(T)$.

3.2 Asymptotic Properties of the Infection Cascade

This section provides the most important formula of the paper, namely a characterization given in Section 3.2.2 of the stochastic dynamics of the tth day of the infection cascade defined by equations (4, 5, 6, 7). The formula remains conjectural in the sense that it depends on

the asymptotic independence of shocks hitting a given node, an unproven property that nevertheless we expect should result in any cascade setting such as ours where a "tree independent" transmission mechanism acts on a locally tree-like random social network.

3.2.1 The First Cascade Step

Consider for $t = 0$ the single shock transmitted from 2 to 1 for two typical individuals $1, 2$, that is, $S_{21}^{(0)} = I_{21}\Omega_{21}\mathcal{D}_2^{(0)}$ as defined by (5). Since $e^{ikI_{21}\Omega_{21}\mathcal{D}_2^{(0)}} = 1 + I_{21}\mathcal{D}_2^{(0)}(e^{ik\Omega_{21}} - 1)$, the characteristic function of $S_{21}^{(0)}$ conditioned on the type $T_1 = T$ is given for finite N by a sum over the possible types of node 2, $T_2 = T'$:

$$\begin{aligned}\mathbb{E}^{(N)}[e^{ikS_{21}^{(0)}} \mid T] &= \sum_{T'} \mathbb{P}(T')\left(1 + \frac{\kappa(T', T)\Pi^{(0)}(T')}{N-1}\mathbb{E}^{(N)}[e^{ik\Omega_{21}} - 1 \mid T_1 = T, T_2 = T']\right) \\ &= 1 + \sum_{T'} \frac{\mathbb{P}(T')\kappa(T', T)\Pi^{(0)}(T')}{N-1}(\hat{f}_\Omega(k|T, T') - 1) \ . \end{aligned} \tag{13}$$

Next consider the asymptotic distribution of the total infection shock $S_1^{(0)} := \sum_{w\neq 1} S_{w1}^{(0)}$ transmitted to individual 1 in day 0. For any N, its characteristic function conditioned on the type $T_1 = T$, is

$$\mathbb{E}^{(N)}[e^{ikS_1^{(0)}} \mid T] = \mathbb{E}^{(N)}\left[\prod_{w\neq 1} e^{ikS_{w1}^{(0)}} \mid T\right] \ . \tag{14}$$

One can prove that any *finite collection* of shocks $\{S_{w1}^{(0)}\}_{w\neq 1}$ are identical and asymptotically independent, conditioned on the type $T_1 = T$. However, this fact cannot prove the following stronger statement:

$$\begin{aligned}\mathbb{E}^{(N)}[e^{ikS_1^{(0)}} \mid T] &= \mathbb{E}^{(N)}\left[\prod_{w\neq 1} e^{ikS_{w1}^{(0)}} \mid T\right] \\ &\sim \prod_{w\neq 1} \mathbb{E}^{(N)}[e^{ikS_{w1}^{(0)}} \mid T](1 + O(N^{-1})) \ . \end{aligned} \tag{15}$$

where $\sim$ represents an unproven step. If this unproven step is accepted as a conjecture, then from (13) and the argument proving Proposition 1, the characteristic function of the *total infection shock* $S_1^{(0)} = \sum_{w\neq 1} S_{w1}^{(0)}$ transmitted to individual 1 in day 0, conditioned on the type $T_1 = T$ must be:

$$\mathbb{E}^{(N)}[e^{ikS_1^{(0)}} \mid T] = \hat{f}_S^{(0)}(k \mid T)(1 + O(N^{-1})) \ , \tag{16}$$

$$\hat{f}_S^{(0)}(k \mid T) := \exp\left(\sum_{T'} \mathbb{P}(T')\kappa(T', T)\Pi^{(0)}(T')\Big(\hat{f}_\Omega(k \mid T', T) - 1\Big)\right) \ . \tag{17}$$

Finally, the *impacted immunity buffer* $\Delta_1^{(1)} = \Delta_1^{(0)} - S_1^{(0)}$ at the end of day 0 is given by (7). We can see directly that $S_1^{(0)}$ and $\Delta_1^{(0)}$ share no common health random variables, and are therefore independent conditionally on the type T of individual 1. From the multiplicative property of characteristic functions of sums of independent random variables, the impacted immunity buffer $\Delta_1^{(1)}$ has the product conditional characteristic function

$$\hat{f}_\Delta^{(1)}(k \mid T) = \hat{f}_\Delta^{(0)}(k \mid T)\hat{f}_S^{(0)}(-k \mid T) \ . \tag{18}$$

By the Fourier Inversion Theorem, one can compute the CDF by taking an L^2-inner product of the kernel $Z(k,x) := \frac{e^{ikx}}{2\pi ik}$ with the CF of $\Delta_1^{(1)}$

$$F_\Delta^{(1)}(x|T) = \mathbb{P}(\Delta_1^{(1)} \leq x|T) = \int_{\mathbb{R}} Z^*(k,x)\hat{f}_\Delta^{(1)}(k \mid T)\ \mathrm{d}k := \left\langle Z(\cdot,x), \hat{f}_\Delta^{(1)}(\cdot \mid T) \right\rangle \ . \tag{19}$$

The conditional infection probability is got from (19) by taking $x = 0$:

$$\Pi^{(1)}(T) = F_\Delta^{(1)}(0|T) = \left\langle Z(\cdot,0), \hat{f}_\Delta^{(1)}(\cdot \mid T) \right\rangle \ . \tag{20}$$

Remark 1. *To handle the singularity of Z at $k = 0$, one can show that under analyticity assumptions, it is sufficient to shift the k-integration slightly into the upper half complex plane. We assume enough regularity that we can do this throughout the paper.*

In summary, day 0 of the infection cascade mapping has been broken down into three substeps that capture the probabilistic implications of equations (4, 5, 6, 7). Each of these substeps depends on the initial conditional distributional data for the collection $\{T_v, I_{vw}, \Omega_{vw}, \Delta_v^{(0)}\}$, combined with a conditional independence assumption. The result of the mapping is full conditional univariate distributional data for the collection $\{\Delta_v^{(1)}\}_{v\in[N]}$.

3.2.2 Higher Order Cascade Steps

In its most reduced form, the proposed infection cascade dynamics is assumed to be given by iterates $t = 0, 1, 2, \ldots$ of the mapping from $\Pi^{(0)}$ to $\Pi^{(1)}$ defined above. This dynamics also leads to formulas mapping the probability distributions for the collection $\{\Delta_v^{(t)}\}$ to probability distribution data for the collection $\{\Delta_v^{(t+1)}\}$. Given the distributional initial data for the collection $\{T_v, I_{vw}, \Omega_{vw}, \Delta_v^{(0)}\}$, day t of the cascade is generated by the following algorithm.

Stochastic Infection Cascade Mapping: Starting from the characteristic functions $\hat{f}_\Delta^{(0)}(k,T)$ and infection probabilities $\Pi^{(0)}(T)$ derived from the initial immunity buffers $\Delta^{(0)}$, iterate the following three steps for $t = 0, 1, 2, \ldots$:

1. Compute the univariate characteristic function $\hat{f}_S^{(t)}(k \mid T) = \mathbb{E}[e^{ikS_1^{(t)}} \mid T]$ of the total infection shock $S_1^{(t)}$ using (17) and (14) with $\Pi^{(0)}$ replaced by $\Pi^{(t)}$:

$$\hat{f}_S^{(t)}(k \mid T) := \exp\left(\sum_{T'} \mathbb{P}(T')\kappa(T', T)\Pi^{(t)}(T')\Big(\hat{f}_\Omega(k \mid T', T) - 1\Big)\right) . \tag{21}$$

2. Compute the univariate distribution of the impacted immunity buffer $\Delta_1^{(t+1)} = \Delta_1^{(0)} - S_1^{(t)}$ using the formula (18):

$$\hat{f}_\Delta^{(t+1)}(k \mid T) = \hat{f}_\Delta^{(0)}(k \mid T)\hat{f}_S^{(t)}(-k \mid T) . \tag{22}$$

3. Compute the conditional infection probability using formula (20):

$$\Pi^{(t+1)}(T) = \Big\langle Z(\cdot, 0), \hat{f}_\Delta^{(t+1)}(\cdot \mid T)\Big\rangle . \tag{23}$$

3.3 Incremental Version of Infection Mapping

The previous derivations lead to recursive formulas (21)-(23) for the cumulative infection probabilities $\Pi^{(t+1)}(T)$, with no need to actually compute $\hat{f}_\Delta^{(t+1)}$ using (22). This formulation is however too restrictive in general. In particular, it will be necessary to determine the fraction of *new type T infectives* on day $t+1$, that is $\pi^{(t+1)}(T) := \Pi^{(t+1)}(T) - \Pi^{(t)}(T)$.

$$\hat{f}_\Delta^{(t+1)}(k \mid T) = \hat{f}_\Delta^{(t)}(k \mid T)\exp\Big(\sum_{T'} \mathbb{P}(T')\kappa(T', T)\pi^{(t)}(T')\big(\hat{f}_\Omega(-k \mid T', T) - 1\big)\Big) \tag{24}$$

$$\pi^{(t+1)}(T) = \Big\langle Z(\cdot, 0), \hat{f}_\Delta^{(t+1)}(\cdot \mid T) - \hat{f}_\Delta^{(t)}(\cdot \mid T)\Big\rangle . \tag{25}$$

Note that we can recursively compute the quantities $\hat{f}_\Delta^{(t+1)}, \pi^{(t+1)}$ from $\hat{f}_\Delta^{(t)}, \pi^{(t)}$.

The importance of the incremental formulation is that it makes it clear how to introduce flexibility to adjust the dynamics in different ways, as we shall explore in subsequent sections.

4 Network Models versus Agent Based Models versus Compartment Models

In a nutshell, the network approach to modelling infectious disease is intermediate in complexity between compartment models, the most popular framework and reviewed in Brauer (2008), and the more complex agent-based models (ABMs) that underpin much public policy as reported in Ferguson and Ghani (2020).

Drawing a random sample of the underlying IRSN for a fixed size N, following assumptions 1 and 2, can be thought of as setting the initial conditions for an agent-based SI contagion model (ABM). Equations (4, 5, 6, 7) give the behavioural rules these agents follow to up-date their immunity buffers and decide to become infected. Our large N cascade mapping formulas provide an approximation to day-by-day rates of infection realized on the finite N sample. It is important to understand the key simplifications that underlie this approximation.

The most important simplification is the washing out of correlations between different parts of the network, as exemplified by the terms in the sum (6). A heuristic argument relates the information not accounted for in the approximation to the information lost if we "homogenize" the ABM as follows. Given any realization of node types $T_v, v \in [N]$, one has a group of permutations τ of the labels v that preserve the node types, i.e. $T_{\tau(v)} = T_v$. Such a τ effectively "rewires" the finite IRG, by mapping any sample of indicators $I_{vw}, v, w \in [N]$ to $\tilde{I}_{vw} = I_{\tau(v)\tau(w)}$. This rewiring preserves the statistical properties of the skeleton, but breaks all social connections, for example mother-child relations. We can "homogenize" the agent-based model by applying a randomly chosen τ to the skeleton each day of the contagion.

Clearly homogenization leads to an "exchangability" symmetry amongst the nodes within each type $T \in [M]$ that is likely not present in the original random sample. We expect that the original ABM will be well approximated by the homogenized ABM if N is large, and this in turn will be well approximated by the large N asymptotic formulas. Furthermore, this line of thinking suggests that more fine-grained type decompositions reduce the effect of homogenization, and hence lead to more accurate approximations.

We should also consider the relation between the M type IRSN model of contagion with a more conventional compartment SI model with M types. In this setting, the population is modelled by an infinite collection of agents falling into disjoint compartments $\mathcal{S}_T, \mathcal{I}_T, T \in [M]$ representing susceptibles and infectives of type T. The standard multi-type SI model follows the ODEs (ordinary differential equations) for the fractional amounts $s(t|T), i(t|T)$ subject to the constraints $s(t|T) + i(t|T) = \mathbb{P}(T)$ for all $t \geq 0$:

$$\frac{\mathrm{d}s(t|T)}{\mathrm{d}t} = -\sum_{T'} k(T,T')s(t|T)i(t|T') \tag{26}$$

$$\frac{\mathrm{d}i(t|T)}{\mathrm{d}t} = \sum_{T'} k(T,T')s(t|T)i(t|T') \ . \tag{27}$$

Each constant transmission coefficient $k(T,T')$ of the compartment model represents

some average rate that type T' infectives infect type T susceptibles. This will be some approximation of the detailed network type-to-type transmission mechanism on an IRSN that intertwines the quantities $\mathbb{P}(T'), \kappa(T,T')$, $\rho_\Omega(\cdot|T,T')$ and the time and type dependent buffer PDF $\rho_\Delta^{(t)}(\cdot|T)$.

5 SIR and SEIR Models: Dynamics with Recovery

The simple SI contagion formulation above assumes that infected individuals never recover, and continue indefinitely to infect other susceptibles. This may be reasonable during the early phase of a contagion, but it is not reasonable over longer periods.

5.1 SIR Model

An SIR (susceptible-infected-removed) model arises when we assume that each day a constant fraction $\beta(T) \in [0,1)$ of infected type T individuals recover or die. We now define $I^{(t)}(T)$ to represent the total fraction of infectious individuals at the end of day t, while $R^{(t)}(T)$ represents the removed (recovered or dead) fraction. If on day t the fraction of new infectives is $\pi^{(t)}(T)$, then on day $t+1$ we have

$$\pi^{(t+1)}(T) = \Big\langle Z(\cdot,0), \hat{f}_\Delta^{(t+1)}(\cdot \mid T) - \hat{f}_\Delta^{(t)}(\cdot \mid T) \Big\rangle \tag{28}$$

where

$$\hat{f}_\Delta^{(t+1)}(\cdot \mid T) = \hat{f}_\Delta^{(t)}(\cdot \mid T) \exp\Big(\sum_{T'} \mathbb{P}(T') \kappa(T',T) \pi^{(t)}(T') \big(\hat{f}_\Omega(-k \mid T',T) - 1 \big) \Big) \ . \tag{29}$$

The remaining fractions satisfy recursions

$$\begin{aligned} I^{(t+1)}(T) &= (1-\beta(T)) I^{(t)}(T) + \pi^{(t+1)}(T) & (30) \\ R^{(t+1)}(T) &= R^{(t)}(T) + \beta(T) I^{(t)}(T) \ . & (31) \end{aligned}$$

5.2 SEIR Model

When infected by COVID-19, as for other infectious diseases, there is a short period averaging $T_e \sim 5.1$ days, called the *exposed period* during which the infected person is not contagious. As in compartment models, it is straightforward to model this additional effect by supposing that all new infections are in the *exposed class (E)*, and each day a fraction $\gamma(T)$ of the type T exposed class becomes contagious, moving into the *infectious class (I)*. Of individuals in

the I-class, a fraction $\beta(T)$ recovers into the R-class. Let $E^{(t)}(T), I^{(t)}(T), R^{(t)}(T)$ denote the fraction of type T individuals in each class at the end of day t. Note that $\gamma(T)E^{(t-1)}(T)$ is the fraction of new type T infectives on day $t \geq 1$, and thus the type T newly exposed fraction on day $t+1$ is

$$\pi^{(t+1)}(T) = \left\langle Z(\cdot, 0), \hat{f}_\Delta^{(t+1)}(\cdot \mid T) - \hat{f}_\Delta^{(t)}(\cdot \mid T) \right\rangle \tag{32}$$

where

$$\hat{f}_\Delta^{(t+1)}(\cdot \mid T) = \begin{cases} \hat{f}_\Delta^{(t)}(\cdot \mid T) \exp\Big(\sum_{T'} \mathbb{P}(T') \kappa(T', T) \gamma(T') E^{(t-1)}(T') (\hat{f}_\Omega(-k \mid T', T) - 1) \Big) & t \geq 1 \\ \hat{f}_\Delta^{(t)}(\cdot \mid T) \exp\Big(\sum_{T'} \mathbb{P}(T') \kappa(T', T) \pi^{(0)}(T') (\hat{f}_\Omega(-k \mid T', T) - 1) \Big) & t = 0 \end{cases} . \tag{33}$$

The remaining fractions satisfy recursions:

$$E^{(t+1)}(T) = (1-\gamma) E^{(t)}(T) + \pi^{(t+1)}(T) \tag{34}$$

$$I^{(t+1)}(T) = (1-\beta) I^{(t)}(T) + \gamma E^{(t)}(T) \tag{35}$$

$$R^{(t+1)}(T) = R^{(t)}(T) + \beta(T) I^{(t)}(T) \ . \tag{36}$$

6 Numerical Implementation

The core of the numerical implementation of the stochastic cascade mapping will be to approximate integrals such as (23) using the Fast Fourier Transform (FFT). The FFT works most effectively on a grid of nonnegative integers we denote by $[\text{Nfft}] := \{0, 1, 2, \ldots, \text{Nfft} - 1\}$ whose log-size $\log_2(\text{Nfft})$ is a small integer, chosen to compromise between precision and computational efficiency. All immunity buffers and exposures will be taken to have integer values on a smaller grid $\{0, 1, 2, \ldots, \text{deltamax} - 1\}$ that represent multiples of a unit of viral dose. That is, we assume that every PDF ρ_X can be replaced by a dimension Nfft probability vector with components $\rho_X(x), x \in [\text{Nfft}]$, such that $\rho_X(x) = 0$ for $x \geq \text{deltamax}$. Here $\text{deltamax} \ll \text{Nfft}$ is a practical upper bound on immunity: anyone with $\Delta \geq \text{deltamax}$ will be assumed likely to resist infection even when all their social contacts get infected.

The characteristic function $\hat{f}_X$ is now replaced by the FFT $\hat{f}_X := \mathcal{F}(\rho_X)$ of ρ_X, defined for each $k \in [\text{Nfft}]$ by

$$\hat{f}_X(k) = \sum_{x \in [\text{Nfft}]} e^{2\pi i k x / \text{Nfft}} \rho_X(x)$$

Then the inverse FFT $\rho_X = \mathcal{F}^{-1}(\hat{f}_X)$ is given by

$$\rho_X(x) = \text{Nfft}^{-1} \sum_{k \in [\text{Nfft}]} e^{-2\pi i k x/\text{Nfft}} \hat{f}_X(k)$$

With the grid [Nfft] set this way, we can implement the incremental SI infection mapping of Section 3.3 with the following steps:

1. Initialize arrays $\mathbb{P}, \kappa, \rho_\Omega, \rho_\Delta^{(0)}$ of sizes $[1, M], [M, M], [\text{Nfft}, M, M], [\text{Nfft}, M]$ respectively. Compute the cumulative distribution $F_\Delta^{(0)}$, of size $[\text{Nfft}, M]$. The initial infection probabilities are $\pi^{(0)}(T) = F_\Delta^{(0)}(0|T)$.

2. Apply the FFT: $\hat{f}_\Omega = \mathcal{F}(\rho_\Omega), \hat{F}_\Delta^{(0)} = \mathcal{F}(F_\Delta^{(0)})$. Then $\pi^{(0)}(T) = \text{Nfft}^{-1} \sum_{k \in [\text{Nfft}]} \hat{F}_\Delta^{(0)}(k, T)$.

3. Precompute the size $[\text{Nfft}, M, M]$ array $R(k, T, T') = \kappa(T, T')\mathbb{P}(T')(\hat{f}_\Omega(-k, T, T') - 1)$.

4. For each day $t = 0, 1, 2, \ldots$ compute recursively the updated arrays by

$$\hat{F}_\Delta^{(t+1)}(k, T) = \hat{F}_\Delta^{(t)}(k, T) \ \exp[\sum_{T' \in [M]} R(k, T, T')\pi^{(t)}(T')] \tag{37}$$

$$\pi^{(t)}(T) = \text{Nfft}^{-1} \sum_{k \in [\text{Nfft}]} (\hat{F}_\Delta^{(t+1)}(k, T) - \hat{F}_\Delta^{(t)}(k, T)) \ . \tag{38}$$

One sees immediately that for day t the computational complexity is dominated by (37) which amounts to $O(\text{Nfft} \times M^2)$ flops for the complex matrix-vector multiplication, followed by $\text{Nfft} \times M$ complex exponentiations. Memory usage is dominated by storing the constant matrix R with $\text{Nfft} \times M^2$ components. Since $\text{Nfft} = 2^{10}$ is a typical value, there is clearly no difficulty in computing the general model with several thousand types on an ordinary laptop.

7 Calibrating IRSNs

This section addresses some of the issues in implementing the infection cascade model on IRSNs, and its generalizations, for a real world network of $\hat{N}$ individuals. The central issue is to construct a sequence of IRSNs of size N increasing to infinity, that is statistically consistent with the real world network when $N = \hat{N}$. Then the statistical model for $N = \infty$ can be subjected to epidemic triggers with any initial infection probabilities $\Pi^{(0)}(T)$, and the resultant infection cascade analytics developed in Section 3 will yield measures of the resilience of the real world network.

The type of network data available to policy makers varies widely from one health jurisdiction to another. Here we imagine a minimal dataset for $\hat{N} = \sum_{T\in[M]} \hat{N}_T$ individuals classified into M types labelled by $T \in [M]$, where $\hat{N}_T$ denotes the number of individuals of type T. Individual type will be assumed not to change over the past N_m months. As a first estimation step, we choose the *empirical type distribution*:

$$\widehat{\mathbb{P}}(T) = \frac{\hat{N}_T}{\hat{N}} \ .$$

Now suppose for illustration that the interconnectivity, exposures and health statistics of the network have been observed monthly for the past N_m months. For any of the monthly observations of the network, edges are drawn between any ordered pair (v, w) of individuals if the exposure of individual w to individual v exceeds a specified threshold (a "significant exposure"). Let $\hat{E} = \sum_{T,T'} \hat{E}_{T,T'}$ be the total number of significant exposures in the network identified in the N_m month historical database, decomposed into a sum over the individual types involved. This data then leads to the *empirical connection kernel*

$$\frac{\widehat{\kappa}(T,T')}{\hat{N}-1} = \frac{\hat{E}_{T,T'}}{N_m \hat{N}_T \hat{N}_{T'}} \ .$$

Recall from the previous section that buffers and exposures are assumed to take values on the integer grid $\{0, 1, 2, \ldots, \text{deltamax}\}$ for some moderately large integer deltamax. For each $T \to T'$ edge $e \in [\hat{E}_{T,T'}]$ we observe the value Ω_e, while for each $v \in [N_m \times \hat{N}_T]$ we also observe samples Δ_v of the type T immunity levels. Then, in view of the intrinsic uncertainties involved, it is reasonable to infer empirical distributions $\rho_\Omega(\cdot, T, T')$ and $\rho_\Delta(\cdot, T)$ from a parametric family of discrete distributions on $\{0, 1, 2, \ldots, \text{deltamax}\}$ that match the sample means and variances $\hat{\mu}_\Omega(T,T'), \hat{\sigma}^2_\Omega(T,T'), \hat{\mu}_\Delta(T), \hat{\sigma}^2_\Delta(T)$.

The data described above leads to a natural calibration of the pre-trigger IRSN model for any value of $N \geq \hat{N}$ (including $N = \infty$) at any time in the near future. The increasing sequence of random IRSN models based on these empirical probability distributions is hoped to capture essential aspects of systemic risk in our specific real world network of size $\hat{N}$. This hope can be realized if it turns out that the $N = \infty$ infection cascade analytics provide a reasonably accurate approximation to simulation results for finite $\hat{N}$.

8 Illustrative Example: Seniors' Residential Centre

The purpose of this example is to provide an easy-to-visualize context for the IRSN framework, namely the setting of a seniors residence with 100 residents (type $T = 1$), 50 trained

staff workers (type $T = 2$) within a town of total population $N_0 = 10000$. We also consider the same IRSN specification scaled up by a multiplier $N = kN_0$. In anticipation of an oncoming contagion, the workers have been trained to high standards of hygiene and care and the residents (who are elderly but healthy) have been instructed in social-distancing and hygiene. The townspeople ("outsiders", with type $T = 3$) on the other hand have only average ability to social distance, and so the contagion hits the town before the centre. The goal of this example is to investigate the vulnerability of the centre to internal contagion starting in the outside town. The benchmark network parameters are given in Table 1, together with numerical implementation parameters $\text{deltamax} = 30, \text{Nfft} = 256$.

The upper left plot of Figure 1 shows the daily infective and removed fractions for the three types, in the benchmark SIR model without further policy interventions. We see that the contagion starts in the outside community, but rapidly invades the centre, resulting in similar infection rates, with a time delay of about 2 days. One can see that the strategy failed for two reasons: first, the contagion was allowed to gain a foothold in the centre and infect a resident; second, the hygiene within the centre was not adequate to contain the resulting seed infection.

What further policy improvements implemented by the management might lead to a better result? The remaining plots in Figure 1 show the results for several combinations of policy interventions. Strategy A is to improve internal hygiene by quarantining all residents and dramatically reducing contacts between workers: $\lambda(1, 1)$ changes from 4 to 0.5 and $\lambda(2, 2)$ changes from 5 to 1. Strategy B is to dramatically reduce the connectivity between the centre and the outside: $\lambda(2, 3)$ changes from 4 to 0.5. We observe that neither A nor B succeeds. Strategy A manages to reduce the contagion to about 37% of the residents, but fails because there is a continual reintroduction of infection from outside. Strategy B also fails: reducing the connections to outsiders simply delays the onset of contagion within the centre by about 10 days. However, the combination of both strategies A and B led to a success in keeping 97% of the residents healthy.

These policy interventions target the social connectivity in the network through social distancing and quarantine. Another important channel would be to reduce the mean viral exposures entering in the exposure PDFs, by measures such as encouraging more cleanliness and the use of masks. Yet another channel is to improve individual immunity buffers by vaccination or other health improvements.

Large N networks typically exhibit "resilient" states that are intrinsically resistant to contagion and "susceptible" states that amplify any introduced infection. Moreover they can be made to transition discontinuously from a resilient state to a susceptible state by

Table 1: Benchmark Parameters

	Resident $T=1$	Worker $T=2$	Outsider $T=3$
$\beta(T)$	0.05	0.05	0.05
$\mathbb{P}(T)$	0.01	0.005	0.985
$\lambda(1,T)=\kappa(1,T)\mathbb{P}(T)$	4	5	0
$\lambda(2,T)=\kappa(2,T)\mathbb{P}(T)$	10	5	4
$\lambda(3,T)=\kappa(3,T)\mathbb{P}(T)$	0	0.020	14
$\mu_\Omega(1,T)$	2	3	1
$\sigma_\Omega(1,T)$	2	2	1
$\mu_\Omega(2,T)$	2	3	1
$\sigma_\Omega(2,T)$	1	2	1
$\mu_\Omega(3,T)$	1	3	3
$\sigma_\Omega(3,T)$	1	3	3
$\mu_\Delta(T)$	8	12	10
$\sigma_\Delta(T)$	3	3	4

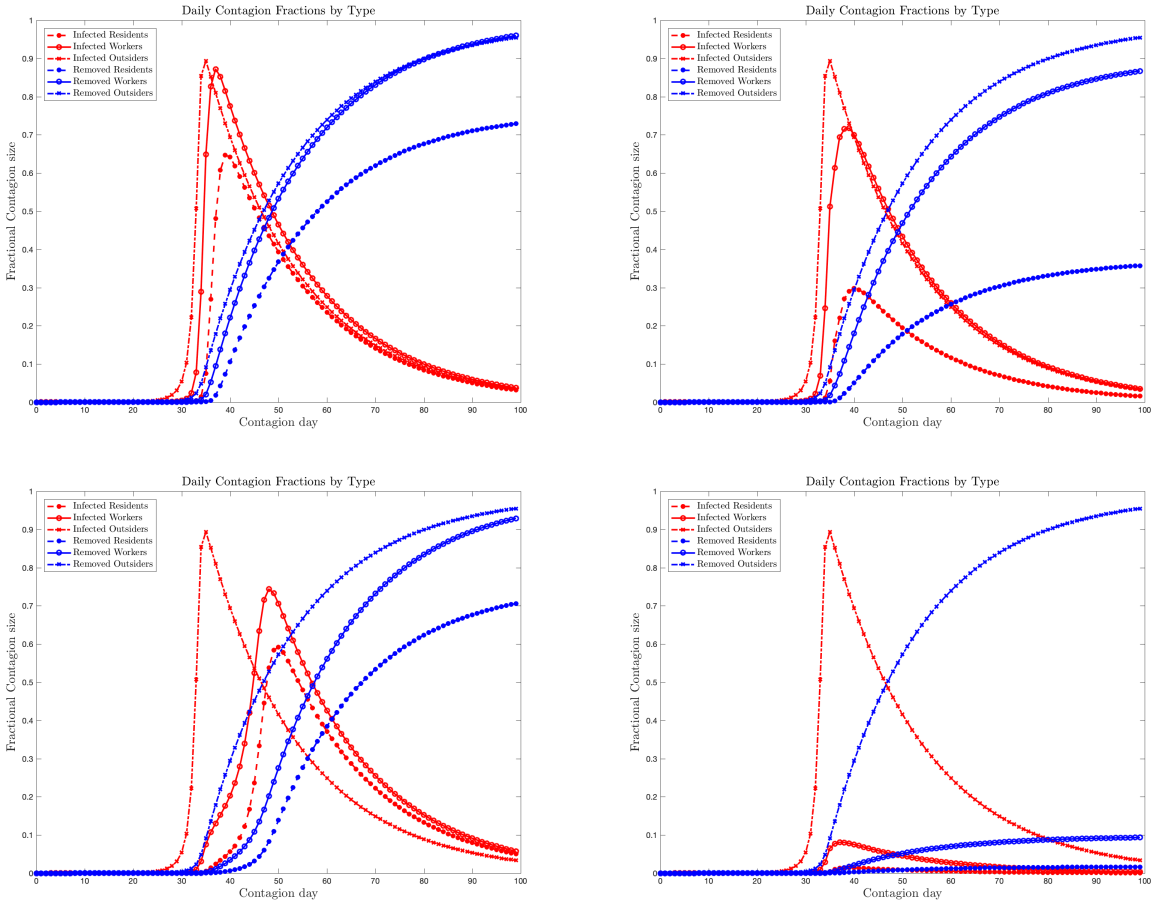

Figure 1: Contagion in the Senior's Residential Centre Model of Section 8. Top Left: Benchmark strategy; top right: Strategy A; bottom left: Strategy B; bottom right: both Strategies A and B

varying a key parameter. Figure 2 shows the long-time values of the removed fractions, as functions of a multiplier z that rescales the benchmark probability mapping kernel $\kappa \to z\kappa$. One sees the remarkable transition from resilient to susceptible at a critical value $z^* \sim 0.70$. This single graph shows clearly the general principle that any contagion can be prevented at the outset by sufficiently strong restrictions on social interactions.

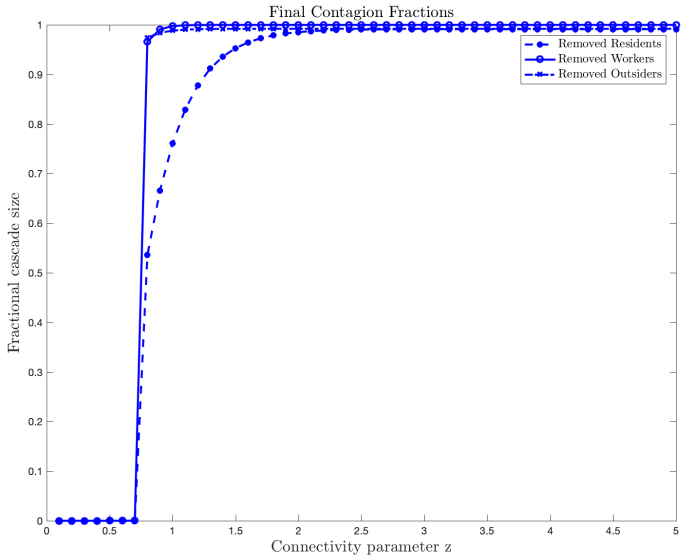

Figure 2: Final removed fractions as a function of z for the benchmark Senior's Residential Centre Model of Section 8, with rescaled probability mapping kernel $\kappa \to z\kappa$.

9 Open Questions

This paper is intended to provide a road map for future research using IRSNs as a tool in understanding aspects of epidemic risk. We end with a brief discussion of three interesting areas of exploration that have so far been left unaddressed.

One line of inquiry asks about the accuracy of the large N approximation to real world models of this type. A first step in this direction is to investigate "synthetic models" to compare the large N asymptotic formulas to simulation studies for finite N. An optimistic hope is that $N = \infty$ formulas will prove to be an effective tool for explaining the systemic resilience of moderately large networks.

A second line of inquiry focusses on calibrating IRSN models of this type to real world social systems. Here the critical issue is the availability of data along the lines discussed in Section 7. Where a suitable representation of a real world network can be found, it will then be of interest to investigate the multiple dimensions of vulnerability exhibited by the calibrated cascade model.

A third avenue of investigation is how to design network models that can be used as a tool to explore and understand further social risk effects. Examples of interesting effects include: the impact of exceptional "superspreader" nodes; overlapping contagions such as influenza

and coronavirus; more diverse types of nodes; country wide networks and the global network.

References

D. Aldous and J. M. Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.

B. Bollobás, S. Janson, and O. Riordan. The phase transition in inhomogeneous random graphs. *Random Struct. Algorithms*, 31(1):3–122, Aug. 2007.

F. Brauer. Compartmental models in epidemiology. In F. Brauer, P. van den Driessche, and J. Wu, editors, *Mathematical Epidemiology*, pages 19–79. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-78911-6.

F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002. ISSN 0218-0006.

N. Ferguson and A. C. Ghani. The global impact of covid-19 and strategies for mitigation and suppression. Technical Report 12, Imperial College COVID-19 Response Team, March 2020.

T. R. Hurd and J. P. Gleeson. On Watts cascade model with random link weights. *Journal of Complex Networks*, 1(1):25–43, 2013.

M. J. Keeling and K. T. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, 2005.

M. Newman. *Networks: An Introduction*. Oxford University Press, Oxford/New York, 2010.

L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, and P. Trapman. Eight challenges for network epidemic models. *Epidemics*, 10:58 – 62, 2015. ISSN 1755-4365. Challenges in Modelling Infectious Disease Dynamics.

R. van der Hofstad. Random Graphs and Complex Networks: Volumes I and II. Book, to be published, 2016. URL `http://www.win.tue.nl/~rhofstad/NotesRGCN.html`.

D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.

A Lemmas and Proofs

Lemma 2. *Let I be any hyperinterval in $\mathbb{R}^d$ and $\bar{y} > 0$. Suppose $g(x, y) : I \times [0, \bar{y}] \to \mathbb{C}$ is a bivariate function such that $g(\cdot, y), \partial_y g(\cdot, y), \partial_y^2 g(\cdot, y)$ are pointwise bounded and in $L^2(I)$ for each value $y \in [0, \bar{y}]$. Then*

$$\lim_{y \to 0} \left\| \frac{1}{y} \log(1 + y g(x, y)) - g(x, 0) \right\|_{L^2} = O(y) \ .$$

Proof of Proposition 1. Verify the hypotheses of Lemma 2 with $N - 1 = y^{-1}$ and

$$g(k, y) = \sum_{T' \in [M]} \mathbb{P}(T') \Big[\kappa(T', T)(e^{ik} - 1) \Big] \tag{39}$$

and apply the Lemma to the logarithm of (9). □

Proof of Lemma 2. Under the assumptions, one can show directly that $f(x, y) := \log(1 + yg(x, y))] - yg(x, 0)$ satisfies $\lim_{y \to 0} f(x, y) = \lim_{y \to 0} \partial_y f(x, y) = 0$ and hence by Taylor's remainder theorem

$$f(x, y) = \int_0^y (y - v) \partial_y^2 f(x, v) \mathrm{d}v$$

One can also show that $\partial_y^2 f(x, v)$ is in $L^2(I)$ for each value $v \in [0, \bar{y}]$ provided $\bar{y} > 0$ is small enough. Then, by Fubini's Theorem, for $y \in [0, \bar{y}]$

$$\|\log(1 + yg(x, y))] - g(x, 0)\|^2 \leq \Big(\int_0^y (y - v) \mathrm{d}v \Big)^2 \max_{v \in [0, \bar{y}]} \left\| \partial_y^2 f(x, v) \right\|^2 \leq M y^4$$

for some constant M, from which the result follows. □