Mean-field-like approximations for stochastic processes on weighted and dynamic networks

Prapanporn Rattana

A thesis submitted for the degree of Doctor of Philosophy

University of Sussex
Department of Mathematics

April 2015

Declaration

I hereby declare that this thesis has not been, and will not be, submitted in whole or in part to another University for any other academic award. Except where indicated by specific stated in the text, this thesis was composed by myself and the work contained therein in my own.

signature .....Prapanporn Rattana.....

University of Sussex

PRAPANPORN RATTANA

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

MEAN-FIELD-LIKE APPROXIMATIONS FOR STOCHASTIC PROCESSES ON WEIGHTED AND DYNAMIC NETWORKS

SUMMARY

The explicit use of networks in modelling stochastic processes such as epidemic dynamics has revolutionised research into understanding the impact of contact pattern properties, such as degree heterogeneity, preferential mixing, clustering, weighted and dynamic linkages, on how epidemics invade, spread and how to best control them. In this thesis, I worked on mean-field approximations of stochastic processes on networks with particular focus on weighted and dynamic networks. I mostly used low dimensional ordinary differential equation (ODE) models and explicit network-based stochastic simulations to model and analyse how epidemics become established and spread in weighted and dynamic networks.

I begin with a paper presenting the susceptible-infected-susceptible/recovered (SIS, SIR) epidemic models on static weighted networks with different link weight distributions. This work extends the pairwise model paradigm to weighted networks and gives excellent agreement with simulations. The basic reproductive ratio, R_0, is formulated for SIR dynamics. The effects of link weight distribution on R_0 and on the spread of the disease are investigated in detail. This work is followed by a second paper, which considers weighted networks in which the nodal degree and weights are not independent. Moreover, two approximate models are explored: (i) the pairwise model and (ii) the edge-based compartmental model. These are used to derive important epidemic descriptors, including early growth rate, final epidemic size, basic reproductive ratio and epidemic dynamics. Whilst the first two papers concentrate on static networks, the third paper focuses on dynamic networks, where links can be activated and/or deleted and this process can evolve together with the epidemic dynamics. We consider an adaptive network with a link rewiring process constrained by spatial proximity. This model couples SIS dynamics with that of the network and it investigates the impact of

rewiring on the network structure and disease die-out induced by the rewiring process. The fourth paper shows that the generalised master equations approach works well for networks with low degree heterogeneity but it fails to capture networks with modest or high degree heterogeneity. In particular, we show that a recently proposed generalisation performs poorly, except for networks with low heterogeneity and high average degree.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Dr Istvan Kiss, for his continuous guidance, support, and encouragement for my study and research. He is an exemplary supervisor. I have come to understand the true definition of a professional supervisor through his expansive knowledge, enthusiasm, patience, and timely wisdom over the past four years. Without these, this thesis would not exist.

I would like to thank Konstantin Blyuss, Ken T.D. Eames, Joel C. Miller, and Luc Berthouze, all of whom I have had the privilege to collaborate with.

The Ministry of Science and Technology of the Royal Thai Government, Thailand and my home affiliation, King Mongkut's University of Technology Thonburi, Thailand have kindly provided the financial support that enabled me to not only to undertake doctoral research, but also expand my knowledge and research horizon in the process.

I have been fortunate to come to know a great many of local, international and Thai doctoral students. Their geniality, companionship, and succour have greatly contributed to the comforting atmosphere that allowed me to live and work unrelentingly during the years leading up to the completion of my thesis.

Most importantly, I would like to thank my family - - - that is to say, my parents, sister, and brother - - - for their unconditional love and encouragement during the course of my study far away from home and their embrace.

Last but absolutely not least, I would like to thank my husband and son, who have always been there beside me. My studies would not have been as pleasurable without their close presence, fresh smiles, big hugs, and pure love.

List of publications and author contributions

1. A Class of Pairwise Models for Epidemic Dynamics on Weighted Networks

P. Rattana, K.B. Blyuss, K.T.D. Eames and I.Z. Kiss (2013).
Bulletin of Mathematical Biology, Vol. 75, Issue 3, pp. 466-490. ISSN 1522-9602.

- P. Rattana conceived the overall goals of the study and the analysis, derived the pairwise equations, derived R_0 and the maximum R_0 value calculations, implemented the numerical solution of the ODEs, performed all relevant simulations and wrote the first draft of the paper and carried out subsequent revisions.
- K.B. Blyuss conceived the overall goals of the study and the analysis, checked some of the calculations using Maple and contributed to writing/revising the paper.
- K.T.D. Eames conceived the overall goals of the study and the analysis and contributed to revising the paper.
- I.Z. Kiss conceived the overall goals of the study and the analysis, derived the next generation R_0 calculations, helped derive the maximum R_0 values, contributed to writing/revising the paper and closely supervised the work of P. Rattana.

2. Pairwise and Edge-based Models of Epidemic Dynamics on Correlated Weighted Networks

P. Rattana, J.C. Miller and I.Z. Kiss (2014).
Mathematical Modelling of Natural Phenomena, Vol. 9, Issue 2, pp. 58-81.

- P. Rattana conceived the overall goals of the study and the analysis, derived the pairwise models, derived R_0, carried out the early growth rate and final epidemic size calculations, implemented the numerical solution of the ODEs, performed all relevant simulations and wrote the first draft of the paper and carried out subsequent revisions.
- J.C. Miller derived and proved the edge-based models and wrote some of the paper.

- I.Z. Kiss conceived the overall goals of the study and the analysis, derived the average weight and the comparison of R_0 values, contributed to writing/revising the paper and closely supervised the work of P. Rattana.

3. The impact of constrained rewiring on network structure and node dynamics

P. Rattana, L. Berthouze and I.Z. Kiss (2014).
Physical review E, Vol. 90, Issue 2, pp. 052806

- P. Rattana conceived the overall goals of the study and the analysis, derived the degree distribution of networks, helped derive clustering calculations, wrote most of the paper and performed all relevant simulations.
- L. Berthouze conceived the overall goals of the study and the analysis.
- I.Z. Kiss conceived the overall goals of the study and the analysis, derived clustering calculations, wrote some of the paper and supervised the work of P. Rattana.

4. Comment on "A BINOMIAL MOMENT APPROXIMATION SCHEME FOR EPIDEMIC SPREADING IN NETWORKS" in U.P.B. Sci. Bull., Series A, Vol. 76, Iss. 2, 2014

I.Z. Kiss and P. Rattana (2014).
Accepted for publication in U.P.B. Sci. Bull., Series A (July 2014),

- P. Rattana conceived the overall goal of the study, performed all simulations and tests, and wrote the first draft and carried out subsequent revision and editing.
- I.Z. Kiss conceived the overall goals of the study and the analysis, contributed to writing/revision and supervised the work of P. Rattana.

Contents

Chapter 1

Introduction

1.1 Mathematical epidemiology background

Throughout history, we have learned a great deal about infectious diseases and the catastrophic impact they can have. In past times, the focus was on understanding how people became infected and how to treat infected people with appropriate medicines. Attempts to prevent disease spread were based solely on experience and intuition. One of the most disastrous epidemics in historical times is the plague epidemic. Over half of the Roman Empire's population died from plague in AD 161-180. Various sources refer to the physician Galen, who wrote detailed accounts of the signs, symptoms and treatments of the plague which became invaluable in successive years [71, 97]. These techniques were still useful in the middle of the 14th century, when more than thirty percent of the European population perished during an outbreak of the Black Death epidemic, another type of plague [107].

Another famous example of rapid disease spread was the smallpox epidemic in the 16th century. In 1518, smallpox was present in the island of Hispaniola and it quickly spread to Cuba and Aztec Mexico. In 1519, approximately thirty percent of Indians on Hispanola had died from smallpox and it had spread to Puerto Rico where it killed over half of the native population in a few months [51]. By 1530, between 3.5 and 15 million out of 25 to 30 million of the Aztec population succumbed to smallpox in less than 6 months [83].

With modern-day advances, preventing disease spread has been made possible by the development of vaccines, a greater understanding of microorganisms, and increasing knowledge in the fields of biology and epidemiology. Further advances have been

made due to the emergence of high power computing techniques and a greater wealth of knowledge in modelling techniques proposed by researchers from disciplines such as mathematics, physics and computer science. Examples of research published on epidemics include diseases such as Influenza [49, 80], Foot and Mouth disease [61, 92] and HIV [1, 98]. There have also been many mathematical models proposed to investigate the spread and control of infectious diseases, such as Bernoulli's model, the well-known mathematical model of a smallpox epidemic proposed by Daniel Bernoulli in 1760 [28] and the system of differential equations modelling a deterministic general epidemic proposed by Kermack and McKendrick in 1927 [62], see also Anderson and May [3].

In this thesis, we consider mathematical modelling and analysis aspects of the spread of disease on networks. This is done to gain better understanding of how various network properties impact on epidemic dynamics including outbreak threshold, early growth rate, final epidemic size or endemic prevalence. A network is normally used to encode connections or links between individuals and it allows for a higher resolution in describing contact when compared to classic compartmental models, where usually one assumes that everyone mixes with everyone else. In this introductory section we will (i) introduce networks together with network properties or metrics used to describe these, (ii) outline the main mathematical network-based models and (iii) give a detailed overview of the research presented in this thesis.

1.2 Overview of networks, stochastic simulations and model types used in network epidemics

While many early models concentrated on compartmental ordinary differential equation models [3, 26, 60], the last 15 to 20 years have seen the refinement of these models by the inclusion of explicit contact structures via graphs or networks. This has revolusionised and revitalised the field of mathematical epidemiology and has lead to a number of modelling techniques on networks [33, 45, 58, 59, 70]. Below, we provide an introduction to the concept of networks, and discuss their main characteristics and metrics, as well as touch on the impact of various network properties on disease spread. This is complemented by a description of how Markovian SIS and SIR epidemics can be simulated on networks using the Gillespie algorithm [42, 43]. In fact, we start with networks since even the most basic or classical models operate on some assumptions about the contact structure, e.g., models may assume either fully mixed populations

or dynamic contacts, which in terms of networks translates to either a fully connected network or a network where links switch at a very fast rate between different node pairs. The introductory section on networks is followed by a review of some of the mainstream model types, starting from exact and classical mean-field to dynamic network models. Moreover, across the introduction we highlight where and how we (i) generalised models, (ii) provided new insight into the impact of network properties on epidemics or (iii) developed new and more realistic models.

1.2.1 Networks and stochastic simulation

Before we discuss a range of models, let us introduce the concept of networks, since all models can be viewed from this perspective.

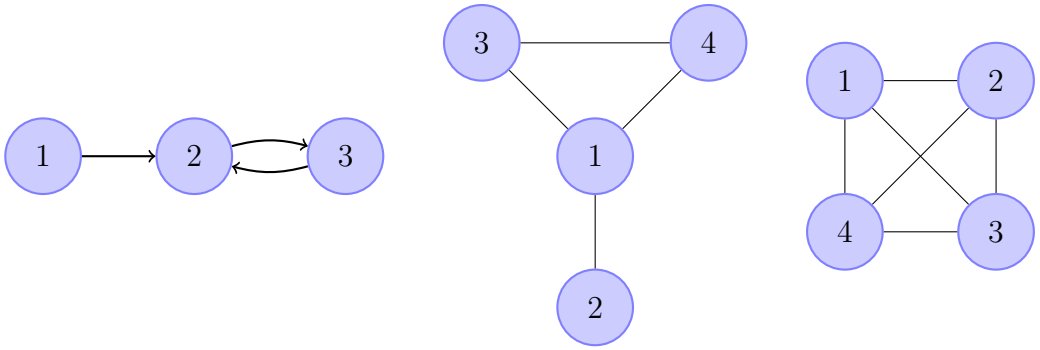
Definition and properties

A network can be defined as $G = \{V, E\}$, where V is a set of vertices/nodes and E is a set of edges/links connecting pairs of vertices. Each edge contained in E corresponds to a pair (u, w) where $u, w \in V$. If $(u, w) \neq (w, u)$ the graph is a directed graph, and if $(u, w) = (w, u)$ for all $u, w \in V$, the graph is undirected. A visual representation of networks is given in Fig. 1.1, where nodes are represented by circles, while links can be undirected (with no arrows present) or directed (with arrows present). For example, a directed graph with three nodes $V = \{1, 2, 3\}$ and edges $E = \{(1, 2), (2, 3), (3, 2)\}$ can be seen in Fig. 1.1(a). Figure 1.1(c) depicts a complete and undirected graph with four nodes $V = \{1, 2, 3, 4\}$ and edges $E = \{(1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$, and we acknowledge that these edges apply in both directions.

A more useful or practical definition is to think of the network in terms of an adjacency matrix, where the entries describe which pairs of nodes in the network share a connection. A network with N nodes is described by an $N \times N$ adjacency matrix, $G = [g_{ij}]$ where $i, j \in 1, 2, \ldots, N$, and where

$$g_{ij} = \begin{cases} 1 & \text{if there is a connection going from node } i \text{ to node } j, \\ 0 & \text{otherwise.} \end{cases}$$

In our case, it is assumed that networks have no self connections, hence $g_{ii} = 0$ for all $i \in 1, 2, \ldots, N$, and the network is undirected, i.e. $g_{ij} = g_{ji}$ for all $i, j \in 1, 2, \ldots, N$. However, we note that using the adjacency matrix definition to describe a network also allows one to account for directed and weighted networks by simply using the



(a) (b) (c)

Figure 1.1: Examples of networks (a) a directed network, (b) an undirected network, (c) a complete and undirected network containing four nodes.

appropriate values for the entries of G. In this thesis we are concerned with undirected networks; henceforth we will focus solely on these network types. Now we turn our attention to properties of the network. The degree of a node, defined as the number of links that it participates in, is an important indicator of the risk of this node becoming infected. Mathematically, the degree of a node i can be defined as $k_i = \sum_j g_{ij}$, for all $i = 1, 2, \ldots, N$. The average degree of a network can be computed as $\langle k \rangle = \sum_{ij} g_{ij}/N$.

The degree distribution, $P(k)$, gives the probability that a randomly chosen node will have degree k. Figure 1.2 depicts two different examples of degree distributions on networks with $N = 1000$ nodes. From a disease transmission viewpoint, the degree distribution is one of the most important characteristics of a network, as it can describe many features of a population's contact structure, such as the prevalence of high-degree nodes, and the extent of homogeneity or heterogeneity across the degrees. Typically, and especially if the network is mixed at random, meaning that the probability of two nodes being connected is proportional to their degrees, highly connected nodes tend to connect to other highly connected nodes and once infection has hit such a node, it is likely that other highly connected nodes will soon follow. Also, the more nodes with higher degrees a network has, the higher the level of infection in the network. For example, in [3, 48, 73, 90, 91] it has been shown that highly heterogenous networks (i.e. $P(k)$ has high variance $Var(X) = E(X^2) - E^2(X)$) are more prone to disease transmission and will experience faster initial spread of infection, even for small transmission rates. It has also been shown that the initial growth of infection is inversely proportional to

the second moment of the degree distribution. In addition, it is useful to compute the moments of the degree distribution, $\langle k^n \rangle = \sum_k k^n P(k)$; the average degree is equivalent to the first moment, $\langle k \rangle = \sum_k kP(k)$.

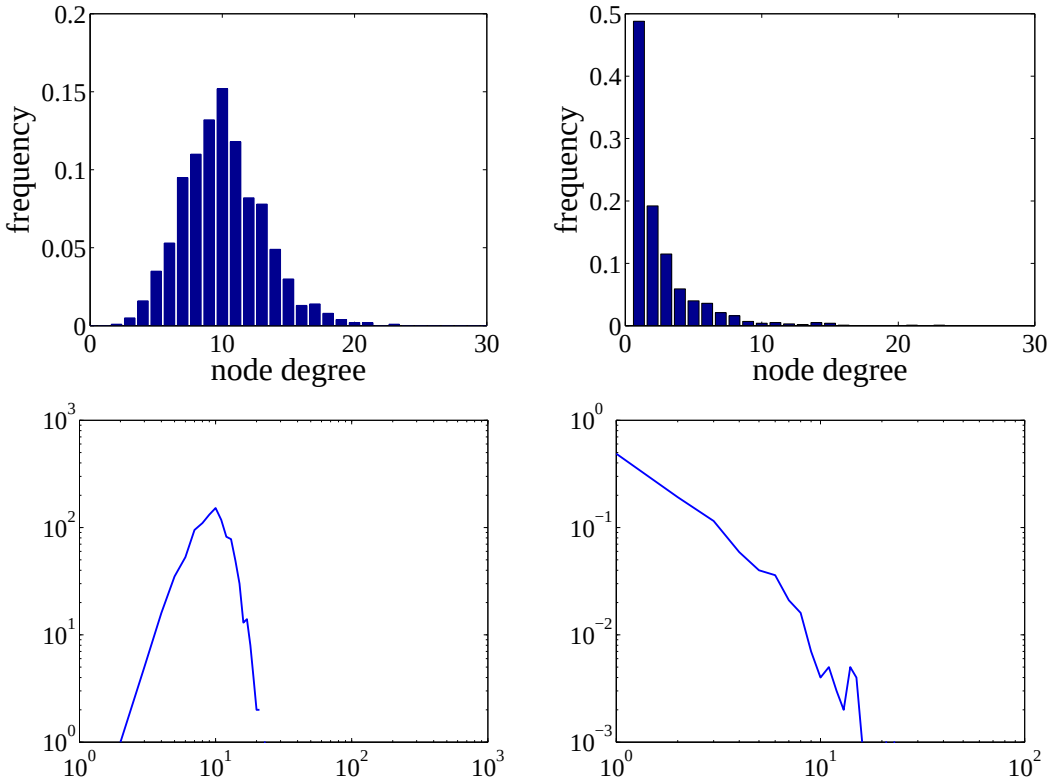


Figure 1.2: Examples of degree distributions of networks with $N = 1000$: random network (left panel) and truncated power law distribution network (right panel). The top histograms are on a linear scale, while the bottom graphs depict the same data on a log scale.

Another important network property is clustering, which represents the propensity that neighbours of the same node are also connected. This property can be captured via the clustering coefficient, C. The clustering coefficient can be computed by counting the total number of triangles contained in the network, and dividing this by the number of connected triples of nodes, open or closed, in the network. For example, the clustering coefficient of the network depicted in Fig. 1.1(b) is 6/10=0.6. When the network is large, the clustering coefficient can simply be calculated using the adjacency matrix G [58]

$$C = \frac{\text{number of triangles}}{\text{number of triples}} = \frac{\text{trace}(G^3)}{\|G^2\| - \text{trace}(G^2)},$$

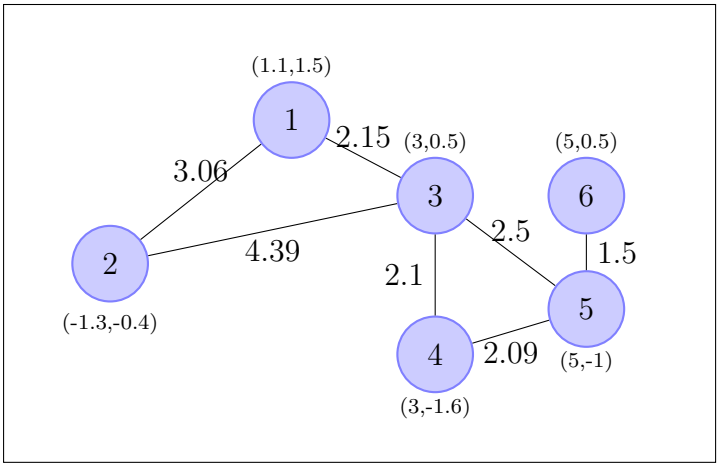


Figure 1.3: A weighted network where the weight of the links are equal to the distance between nodes.

where $\|G\| = \sum_{ij} g_{ij}$. Obviously, for larger networks where this may not be feasible to compute, one can simply investigate each node and its neighbourhood to work out clustering. It is important to note that two networks with identical clustering coefficient and degree distribution can be comprised of completely different building blocks, such as triangles, two triangles with a common edge or fully connected motifs such as squares, pentagons or hexagons. This is important since despite clustering being the same, the outcome of the spread of infection can be non-negligibly different on such networks [32, 58, 102, 109].

A simple network can be modified by the addition of link weights [6, 8, 110]. Weights can encode many network properties, such as the spatial distance between nodes or the contact frequency between nodes. The value of the distance between nodes i and j, w_{ij}, is allocated to each link, as shown in Fig. 1.3. In addition to this, the weighted adjacency matrix may be additionally considered as a new set of topological measures of the contact network characteristics. An important element of this extension is the relation between weights and their distribution and how these impact on disease transmission. [16, 21, 22, 25, 34].

In this thesis, we first consider weighted networks and see these as a more refined model when compared to models defined on unweighted networks. While the analysis is simpler on unweighted networks, many examples suggest that not all links are equal and can reflect true processes that make the interactions uneven, e.g. household or local *versus* global links. In Chapter 2, we consider models on undirected, homogeneous

networks. Different weights w_{ij} are associated with the strength of links, and we encode all of this information within a weights matrix W. We consider two different methods of allocating weights to network links, namely random and fixed allocation, and we experiment with various distributions of link weights. In Chapter 3, we extend the weighted networks model to undirected, heterogeneous networks. This time, the weights of links can be related to the degree of the nodes they connect. In particular, the weight $\hat{w}_{ij}$ is associated with a link between two nodes having degree i and j, respectively. In other words, weights and nodal degrees are not independent; weights are a function of i and j.

Generating networks

In the previous section, we discussed how all network properties, such as average degree, degree distribution and clustering coefficient, can be computed from an adjacency matrix G. In this section, we describe a method for generating networks, and creating the associated adjacency matrix G.

We first discuss a method which will generate the most basic network, namely a classical random network or Erdős-Rényi network [36]. The algorithm starts with N nodes, and an empty graph, i.e, $G = [g_{ij}] = 0$ for all $i, j \in 1, 2, \ldots, N$. Then, a graph is constructed by connecting nodes randomly. This is done by considering each possible edge and activating it with probability p or discarding it otherwise. It is worth noting that each link is considered uniquely, and we consider an undirected network. Using this algorithm, the resulting network has average degree $\langle k \rangle = p(N-1)$. Since a node has $N-1$ potential neighbours, the probability that k of these will be connected to it is simply

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}.$$

The degree distribution of such a network of large is well described by a Poisson distribution $P(k) = \langle k \rangle^k \frac{e^{-\langle k \rangle}}{k!}$ as $N \to \infty$.

Next, we discuss the configuration model; a simple and widely used method which constructs a network with a given degree distribution. We select a degree for each node from the degree distribution, and create a dynamic list containing the node label i, repeated as many times as the degree of that node. For example, if we desire to construct a network with N nodes as in Fig. 1.4, the given degree of node 1 is $k_1 = 4$, $k_2 = 2$, $k_3 = 6$,..., $k_{N-1} = 3$ and $k_N = 3$. So, the dynamic list can be created to yield

1	1	1	1	2	2	3	3	3	3	3	3	...	$N-1$	$N-1$	$N-1$	N	N	N

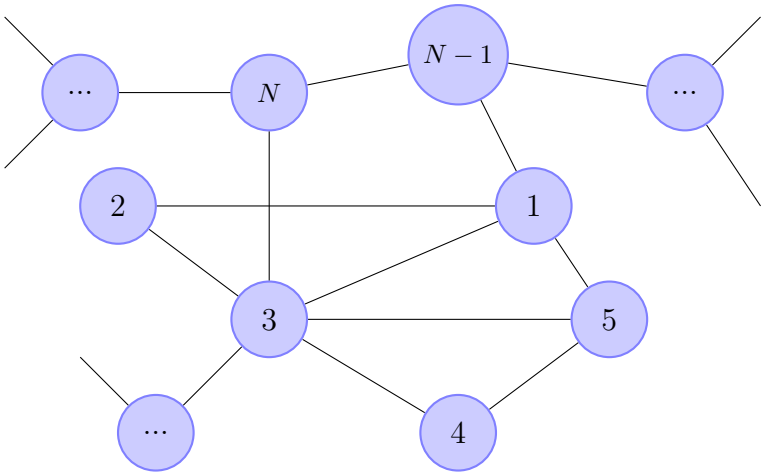


Figure 1.4: The example of a network with N nodes.

Each element in the list denotes one half of an edge, also called a stub. We choose two stubs from the list at random and connect the associated nodes, before removing the two stubs from the list. When a link between two nodes is created, we set $g_{ij} = g_{ji} = 1$. The process is continued until the dynamic list is empty. Clearly, this process needs to deal with the occurrence of double links and self loops. However, we know that the number of such events is approximately $1/N$. As the network size increases, the probability of double links and self loops occurring tends towards zero, provided that the average degree is finite. We note that we also need to impose conditions, such as the sums of all degrees must be even. The algorithm can stall, in which case it can no longer successfully allocate any further links. When this occurs, the degrees of all nodes are reallocated and the process is restarted.

Another approach to generate networks could be to account for more details or properties of the nodes. In many cases, nodes are embedded in a 2D space or higher dimensions. Generating networks based on the nodes' spatial proximity leads to random geometric graphs [93]. This is a specific approach which we used in the context of dynamic networks, as shown in Chapter 4. Here, nodes are placed randomly on a square of size L with periodic boundary conditions. Connectivity is determined based on spatial proximity; if a node is within a certain radius of the target node, they are said to be in the same local area, and can become connected during a rewiring process,

whereby one existing link is removed and a new link is generated simultaneously. The connectivity of the network can be calculated by using an extra matrix to store the node locations. This matrix can be used to compute the distance between nodes and to determine which nodes are in the same local area. Throughout this thesis, we use various techniques to generate our networks.

Epidemic dynamics

In this part, we describe the basic model ingredients of a stochastic epidemic model on networks. We assume that individuals are represented by nodes in the network and the contact pattern amongst these is coded by the links of the network. Births and deaths are ignored and only fixed-size populations are considered. Epidemic transmission on networks needs a classification of individuals according to their status from a disease view point. For all cases considered in this thesis it is enough to consider the following states: susceptible S, infected and infectious I, and recovered/immune/removed R. This results in all nodes being labeled according to the dynamics, and the dynamics itself describes the rates of change of these labels on the nodes.

In order to understand and describe the spread of infectious diseases across networks, we focus on the two most common and widely used epidemic models, namely Susceptible-Infectious-Susceptible (SIS) and Susceptible-Infectious-Recovered (SIR) models. The first model, SIS, represents epidemics where individuals simply alternate between two possible states S and I. Namely, susceptible individuals can be infected by their neighbours, and will remain in the infectious state until they recover. Once the individual recovers, they return to the susceptible state and are ready to catch the disease again. Therefore, the cycle $S \to I \to S$, can be observed many times for various nodes. This model is a good representation of diseases where treatment is possible but being infected does not result in immunity. For example, sexually transmitted infections such as chlamydia or gonorrhoea [40, 48] can be modelled in this way.

On the contrary, SIR represents diseases where after being infected, individuals move into a recovered state. After a node has been infected and his/her infectious period has elapsed, it plays no further role and remains in the recovered state permanently. The recovered state can be used to represent individuals who have died as a result of the disease, individuals who have developed lifelong immunity to the disease, or individuals who have been removed from the process and are therefore isolated from the rest of the population. Therefore, in this model, the transition $I \to S$ of the SIS process

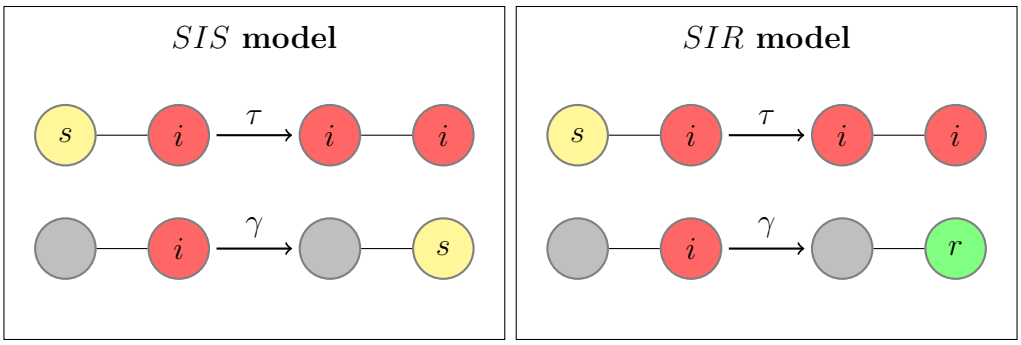


Figure 1.5: A susceptible individual can be infected by an infected neighbour at rate τ. An infected individual recovers with rate γ, independently of the state or number of his/her contacts. After recovering, for the SIS model, infected individuals become susceptible, but for the SIR model, recovery results in immunity. After recovery, individuals play no further role in the dynamics.

changes to $I \to R$. The SIR process can be used to model diseases such as measles or chickenpox [15, 68].

Hence, both epidemics are driven by two events: (a) infection and (b) recovery, see Fig. 1.5. The rate of transmission from an infected to susceptible individual is assumed to be constant, denoted by τ, and it assumed to happen according to a Poisson process with this rate. An infected individual recovers independently of the network and the state of its contacts, and this is also modelled as a Poisson process at rate γ. All these events are considered to be independent and thus a susceptible node with k infectious contacts, becomes infected according to a Poisson process at rate $k\tau$, as given by the theory of pooled Poisson processes [87]. As a result of assuming Poisson processes, it follows that time to infection and time to recovery are exponentially distributed. This may not be realistic and can be changed but model formulation and analytical tractability will be much more difficult [15, 68].

Stochastic simulation

Such a stochastic process can be rigorously simulated by keeping track of all possible events in the network and the rate at which these happen. An efficient method for implementing a stochastic simulation is given by the well-known Gillespie algorithm [42, 43]. This algorithm is based on two simple steps: (a) working out the rate of all events and computing the overall rate of a change occurring, followed by (b) selecting the event to happen at random, but proportionally to the events' rates relative to each

other. Firstly, the total rate of all possible transitions, denoted by T, is calculated from the current status of all individuals across the whole network. This is done by determining the rate of infection for all susceptible nodes and the rate of recovery for all infectious nodes. Let the rates of these events be denoted by r_i where $i = 1, 2, 3, ..., N$. For example, rate r_1 describes the speed or timescale of node number 1 changing its state. The infection of a susceptible node depends on how many infected neighbours it has, but the recovery rate of an infected node is independent of the network and status of neighbours. For example, in Fig. 1.6 we illustrate a possible situation with some transition rates being computed. Node 1 has 2 infected neighbours, so $r_1 = 2\tau$, and the 3 infected neighbours of node 3 yield $r_3 = 3\tau$. While all infected nodes, e.g. nodes 2, 5 and N have $r_2 = r_5 = r_N = \gamma$. Therefore, $T = \sum_{i=1}^{N} r_i$, and the time to the next event, t_{next}, is chosen from an exponential distribution with rate T. This can be chosen by computing

$$t_{next} = \frac{-\ln(u)}{T},$$

where $u \sim U(0, 1)$.

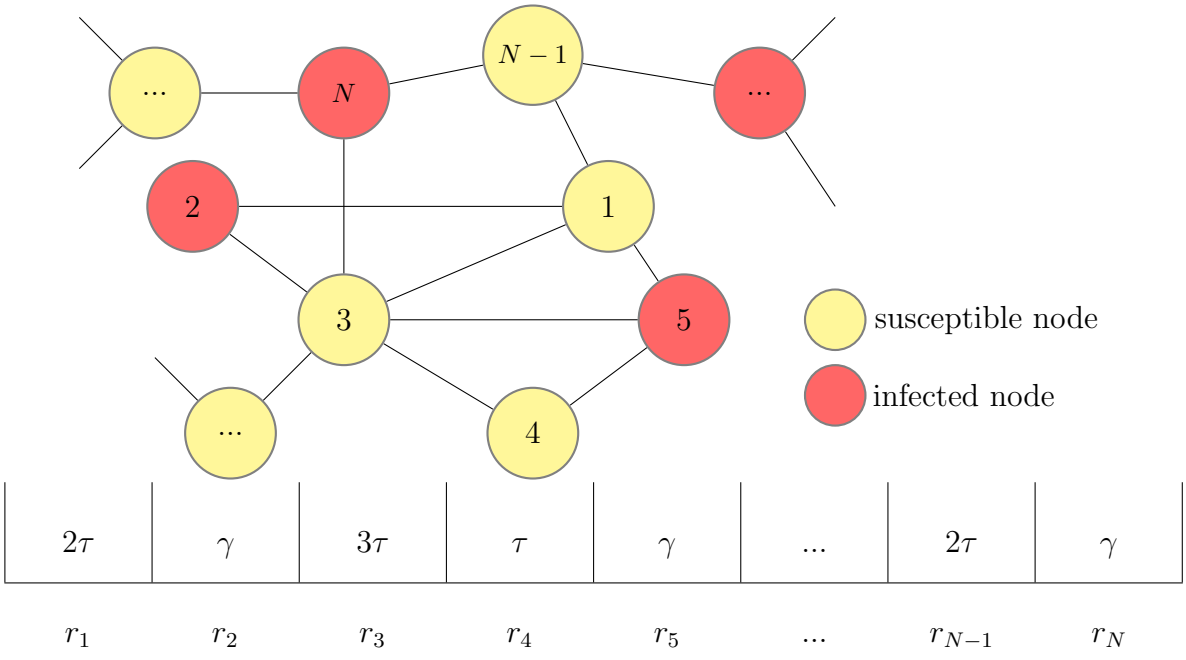


Figure 1.6: Illustrating the relation between the network, status of nodes and events rate vector used by the Gillespie algorithm.

Next, a single event is chosen at random but proportionally to its rate. This can be done by computing $T_i = \sum_{j=1}^{i} r_j$, for all $i = 1, 2, 3, ..., N$. Then, the associated event occurs at node i, where i is the smallest integer satisfying $T_i \geq T \times u$. Once the time to next event and the event itself has been found, necessary rate updates are performed, and the process begins again. The Gillespie algorithm is efficient since following a new event rates only need to be checked and recomputed in the neighbourhood of the node where the change has happened and there is a simple one-to-one correspondence between the node number and events. The process is continued until a time specified by the user or until an absorbing state has been reached. It is worth noting that the times in each individual realisation are at un-even time points and care has to be taken when averaging between different simulations (see the right panel of Fig. 1.7). In Fig. 1.7, we chose values for the transmission and recovery rates which were likely to lead to an epidemic and we selected simulation which did not die out.

We now present the main mathematical models that can be used to capture either the true probabilities of various states arising in time or mean-field models that are capable of capturing the topology of the network and disease dynamics and typically give rise to results that are in good agreement with the expected values resulting from multiple realisations of the simulations. In fact the models presented below form part of the current tool kit used by researchers focusing on modelling stochastic process on networks [24, 59, 89].

1.2.2 Exact models

We first discuss a common approach to describe SIS disease dynamics on a fully connected network with N nodes, given by the Kolmogorov forward equations or master equations [56, 57, 87]. The model describes the probability of the population being in all possible states at time t. The system states are given by $\{S, I\}^N$, where S and I are susceptible and infected states, respectively. However, the state space containing 2^N elements can be reduced to a state space of size $(N + 1)$ by noting that only the number of infectious nodes matters and not their position, due to the network being fully connected. Hence, the resulting Kolmogorov equations can be written as

$$\dot{p_n}(t) = a_{n-1}p_{n-1}(t) - (a_n + c_n)p_n(t) + c_{n+1}p_{n+1}(t), \qquad \text{(KE)}$$

where $p_n(t)$ is the probability that the system has n infectious nodes at time $t \geq 0$, with $n = 0, 1, 2, \ldots, N$, a_n is the birth or infection rate at which the system moves from

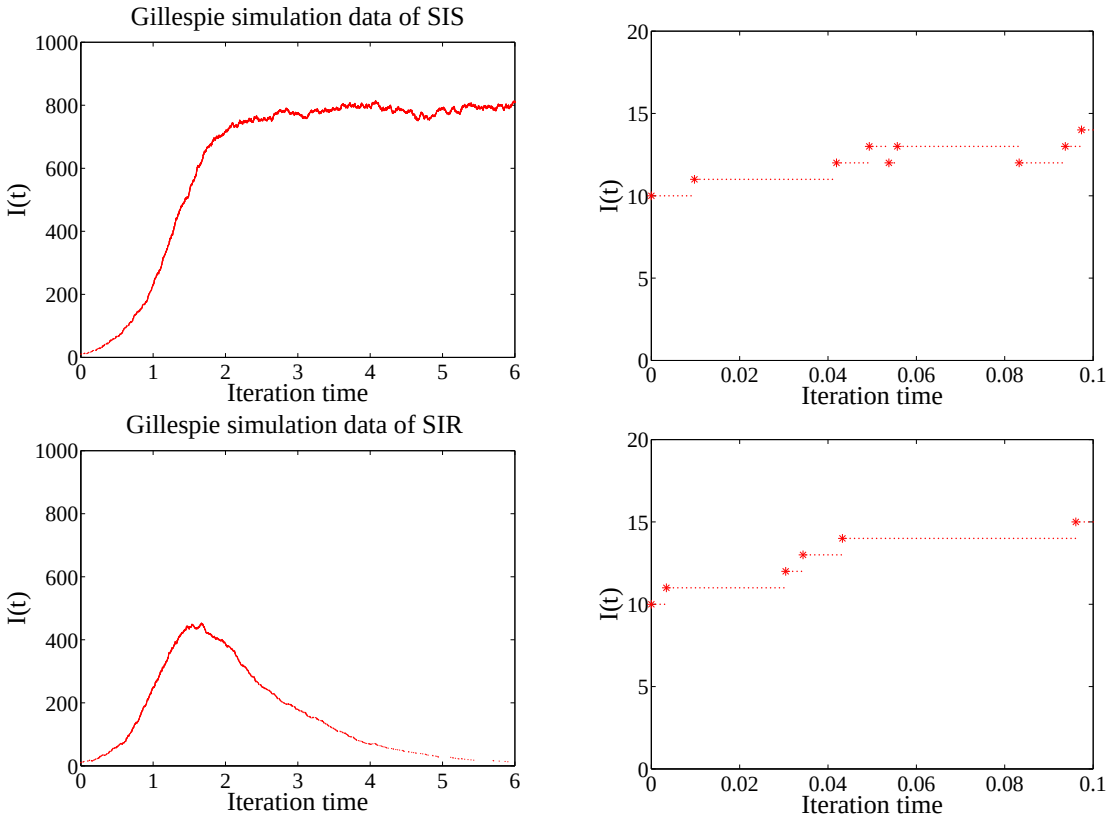


Figure 1.7: The plots show the prevalence of infection for SIS (the top panel) and SIR (the bottom panel) dynamics from the network with $N = 1000$, $\gamma = 1, \tau = 0.5$, and 10 initial nodes at $t = 0$ are infected. For the right panel, the markers ($\star$) correspond to the data based on the Gillespie simulation from the left panel, and the markers ($\cdot$) correspond to the interpolate data.

state n to $n+1$, c_n is the death or recovery rate at which the system moves from state n to $n-1$ and $a_{-1} = c_{N+1} = 0$. Solving these master equations leads to an exact solution of the SIS dynamics on fully connected networks. However, this approach is limited as fully connected networks are not often encountered in real situations. For a fully connected network, $a_k = \tau k(N-k)$ and $c_k = \gamma k$. As expected, $k(N-k)$ represents the number of edges between infected and susceptible individuals when k infectious nodes are present. The master equations can be written down in general for any network but the system will be 2^N-dimensional and thus unfeasible to work with, even for small networks.

Kiss & Simon [66] have shown that the master equations can be reduced to a low dimensional approximation model, with fewer than $(N+1)$ equations. By making the assumption that the number of infected individuals is binomially distributed and using an a priori binomial distribution, the Kolmogorov equations can be presented in the form of two differential equations.

The major challenge however, is to find a correct functional form for the infection rates a_k for any network in general. Based on the random mixing argument, Kiss and Simon [66] have also shown that for homogeneous random networks, a_k can be written as

$$a_k = \tau(N-k)\langle k\rangle \frac{k}{N-1},$$

where $\langle k\rangle$ is the average degree of nodes. This model works well for SIS dynamics on fully connected networks and for networks with low degree heterogeneity. Furthermore, Shang [104] claims to generalise this approach of Kiss and Simon [66]. Shang proposed that a_k in general could be written as

$$a_k = \frac{\tau k(N-k)\langle k^2\rangle}{\langle k\rangle(N-1)},$$

where the network is given in terms of the degree distribution $P(k)$, $k = 0, 1, 2, \ldots, N$, with $\langle k\rangle = \sum kP(k)$ and $\langle k^2\rangle = \sum k^2P(k)$. However, our tests, as shown in Chapter 5, show that this proposed generalisation performs poorly for all networks proposed by Shang [104], except for heterogenous networks with relatively high average degree.

1.2.3 Classic compartmental models

Classic compartmental models are mathematical models which are based on the homogeneous random mixing assumption, where it is assumed that every individual in the system is connected to every other individual, or links between individuals change

or rewire at an infinitely fast rate. An early and well-known model of this type was proposed by Kermack and McKendrick [62], and more recently, the more mainstream models were summarised by Anderson & May, Diekmann & Heesterbeek and Keeling & Rohani [3, 26, 60]. Compartmental models focus on evolution equations for the number of individuals in the various possible states, e.g. susceptible, $S(t)$, infected, $I(t)$, and/or recovered, $R(t)$. This results in a compartmentalisation of the population depending on disease state of individuals. The infection and recovery process then dictates the explicit form of the differential equations for the various classes. The two fundamental compartmental epidemic models, given as sets of ordinary differential equations (ODEs), are as follows:

the Susceptible-Infectious-Recovered model, or the SIR model

$$\begin{aligned} \frac{dS}{dt} &= -\beta I \frac{S}{N}, \\ \frac{dI}{dt} &= \beta I \frac{S}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I, \end{aligned} \tag{1.1}$$

the Susceptible-Infectious-Susceptible model, or the SIS model

$$\begin{aligned} \frac{dS}{dt} &= \gamma I - \beta I \frac{S}{N}, \\ \frac{dI}{dt} &= \beta I \frac{S}{N} - \gamma I. \end{aligned} \tag{1.2}$$

In both the SIR and SIS models, it is assumed that no births or deaths occur in the population; γ is the rate at which an infected individual recovers, and β is the rate at which an individual makes contact with random members of the population. A typical output from these models is shown in Fig. 1.9. The models above are frequency dependent and tell us that during a unit time, only a proportion S/N of the βI infectious contacts are made with susceptible nodes. Such models can be analysed using classic tools from dynamical systems and bifurcation theory which often involves finding steady states and their stability. Using such approaches, it is well known that for the SIR-type epidemics the following statements hold:

- For t close to zero, the number of infected individuals is $I(t) \approx I(0)e^{\left(\beta \frac{S(0)}{N} - \gamma\right)t}$. This is calculated by integrating $\dfrac{dI}{dt}$, Eq. (1.1), and assuming that the number

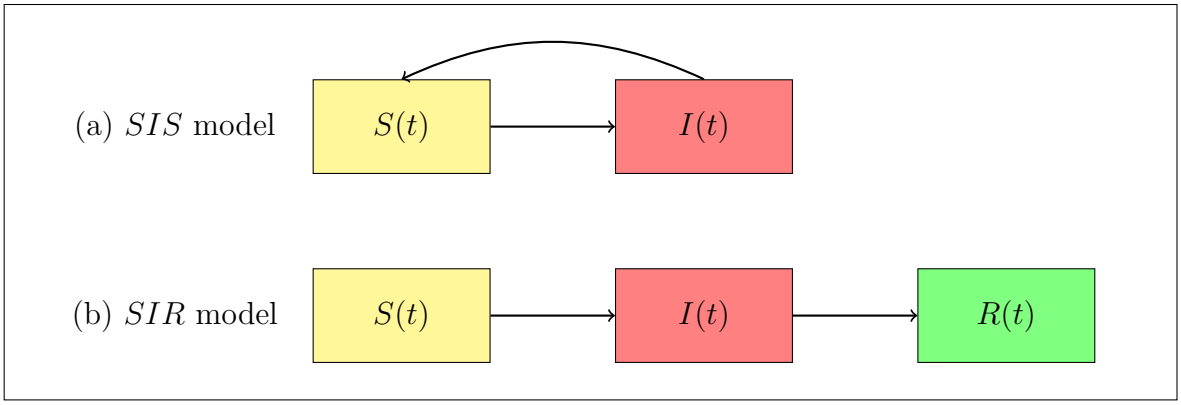


Figure 1.8: Compartmental model for SIS and SIR dynamics.

of susceptible and infectious individuals when t is close to zero are equal to the initial conditions for the number of susceptible and infectious individuals, given by $S(0)$ and $I(0)$, respectively.

- The basic reproductive number is given by $R_0 = \dfrac{\beta}{\gamma}$, which describes the average number of secondary infections produced by a single infectious individual in an otherwise susceptible population [3, 27]. This results from stability analysis of the disease free steady state or by more biologically relevant arguments.

- The final epidemic size, satisfies $R(\infty) = 1 - S(\infty) = 1 - e^{-R_0 R(\infty)}$, where $I(\infty) = 0$.

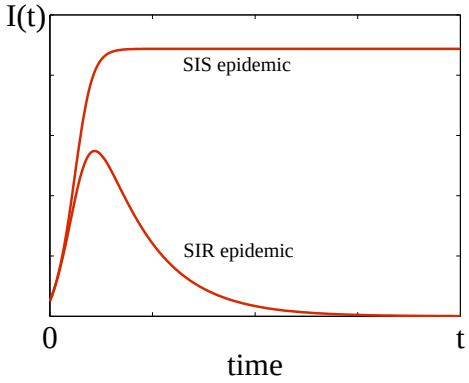


Figure 1.9: The evolution of disease prevalence during an outbreak.

Whilst compartmental models are simply used to analyse and calculate epidemiological details for understanding the epidemics of infectious diseases, they are formulated under the assumption of homogeneous random mixing of the population. This means that every individual has the same probability of coming into contact with another individual. In other words, each susceptible individual is infected by infectious individuals in the population with the same probability. However, in reality contact between individuals is more sophisticated or complex [58, 82, 86, 99] and this will affect the spread of the diseases. Such aspects are difficult to account for and extensions are needed.

1.2.4 Pairwise models

In order to account for more realistic and complex contact patterns, more sophisticated epidemic models on networks have been developed [24, 30, 33, 34, 45, 59, 70, 72].

We will now discuss pairwise models, which are well known epidemic models on networks with both SIS and SIR dynamics [32, 33, 47, 54, 53, 58, 105]. Pairwise models successfully interpolate between classic compartmental ODE models and full individual-based stochastic network simulations with the added advantage of high transparency and a good degree of analytical tractability. The aim of this approach is to take into account more details of the network structure by considering not only nodes in the network but also separately accounting for edges. This is natural as the status of the node depends on the links to its neighbours.

Let us introduce some notations which are commonly used in classic pairwise models for a network with N nodes. First, for all nodes $i \in N$, A_i represents the following logical statement: if node i is in state A then $A_i = 1$, otherwise $A_i = 0$. The notation $[A]$ represents the number of nodes across the whole network in state A, so $[A] = \sum_i A_i$. Next, a pair of type $A - B$ describes the connection between a node in state A and a node in state B. Thus, $[AB] = \sum_{i,j} A_i B_j G_{ij}$ denotes the number of $A - B$ pairs across the whole network, where G is the adjacency matrix of the network, and where $A_i B_j G_{ij}$ is the logical statement that node i is in state A and node j is in state B and there is a link between node i and node j, and $[AB] = [BA]$. We note that $[AA]$ is equal to twice the number of uniquely counted edges with nodes at both ends in state A. A triple of type $A - B - C$ denotes a group of three connected nodes, where the centre node is in state B and it is connected to a node in state A and a node in state C. The logical statement still holds true for a triple; $[ABC] = \sum_{i,j,k} A_i B_j C_k G_{ij} G_{jk}$.

As discussed previously, our studies are focused on SIS and SIR dynamics. So,

in line with the above, $A, B, C \in \{S, I, R\}$. We now reflect on how nodes, pairs and triples are related. For example, the expected number of infectious nodes, $[I]$, increases proportional to the expected number of $S - I$ links, $[SI]$, with proportionality rate τ. Similarly, $[SI]$ decrease due to within pair infection of the S node or the recovery of the I node. Equally, $[SI]$ decrease due to infection of the S node by a node outside the pair, and this is captured by the $-\tau[ISI]$ term. In the same way, $S - S$ links can be destroyed by infection from outside the pair and this is proportional to $-\tau[SSI]$, and this provides a positive contribution to the SI class, since $S - S - I$ turns into $S - I - I$. With careful bookkeeping and taking into account these hierarchies, the pairwise models for the SIR [58] and SIS dynamics for homogenous networks are:

$$
\begin{aligned}
[\dot{S}] &= -\tau[SI], \\
[\dot{I}] &= \tau[SI] - \gamma[I], \\
[\dot{R}] &= \gamma[I], \\
[\dot{SS}] &= -2\tau[SSI], \\
[\dot{SI}] &= \tau([SSI] - [ISI] - [SI]) - \gamma[SI], \\
[\dot{SR}] &= -\tau[ISR] + \gamma[SI], \\
[\dot{II}] &= 2\tau([ISI] + [SI]) - 2\gamma[II], \\
[\dot{IR}] &= \tau[ISR] + \gamma([II] - [IR]), \\
[\dot{RR}] &= \gamma[IR],
\end{aligned}
$$

and

$$
[\dot{S}] = \gamma[I] - \tau[SI],
$$

$$
\begin{aligned}
[\dot{I}] &= \tau[SI] - \gamma[I], \\
[\dot{SI}] &= \gamma([II] - [SI]) + \tau([SSI] - [ISI] - [SI]), \\
[\dot{II}] &= -2\gamma[II] + 2\tau([ISI] + [SI]), \\
[\dot{SS}] &= 2\gamma[SI] - 2\tau[SSI],
\end{aligned}
$$

where τ is the rate of disease transmission across an edge between an infected and a susceptible individual, and infected individuals recover independently of each other at rate γ. The following conservation relations hold: $[S] + [I] + [R] = N$ and $[S] + [I] = N$. Moreover, we have that $[SS] + 2[SI] + [II] + 2[SR] + 2[IR] + [RR] = \langle k \rangle N$ and $[SS] + 2[SI] + [II] = \langle k \rangle N$, where $\langle k \rangle$ is the average degree. By specifying the initial conditions we in fact feed information about the network to the ODE model.

Both systems are not closed, as equations for the pairs require knowledge of triples, and thus, equations for triples are needed. This dependency on higher-order moments can be broken via approximating triples in terms of singles and pairs [58]. The agreement of the results from the closed system with simulation results depends on how well the closure captures essential features of network structure. A classic closure for a homogeneous network is given by

$$[ABC] \approx \frac{\langle k \rangle - 1}{\langle k \rangle} \frac{[AB][BC]}{[B]},$$

where $\langle k \rangle$ is the average degree. Moreover, under the assumption that a connection between the node in state A and the node in state C may be present, a different closure may be more appropriate. This can take the following form

$$[ABC] \approx \frac{\langle k \rangle - 1}{\langle k \rangle} \frac{[AB][BC]}{[B]}((1 - \phi) + \phi \mathcal{C}_{AC}),$$

where ϕ is the clustering coefficient, and $\mathcal{C}_{AC} = N[AC]/\langle k \rangle [A][C]$ is the correlation between nodes in A and C, respectively.

When looking at the basic reproductive number R_0 in an SIR pairwise model, any $R_0 > 1$ means that the disease is expected to spread within the population, while $R_0 < 1$ will lead to the extinction of the disease. We consider the condition under the

initial growth rate $[\dot{I}] > 0$, so we have

$$\frac{\tau[SI]}{\gamma[I]} > 1.$$

Thus, Keeling [58] proposed that R_0 should depend on the correlation between susceptible and infected individuals which is $\mathcal{C}_{SI} = N[SI]/\langle k\rangle[S][I]$, and it was also shown that if the network has susceptible-infective correlation, ignoring clustering, $R_0 = ((\langle k\rangle - 2)\tau)/\gamma$, and $R_0 = \langle k\rangle\tau/\gamma$ if susceptible and infected individuals are uncorrelated and thus we are lead back to the compartmental model.

Eames & Kelling [59] extended this approach to networks with heterogeneous degree distributions, where $[A_k]$ is the expected number of nodes in state A with degree k. The system of equations remains challenging for both SIR and SIS dynamics. For example, the expected number of $S-I$ pairs in an SIS model is given by:

$$\dot{[S_k I_l]} = \tau \sum_m ([S_k S_l I_m] - [I_m S_k I_l]) - \tau[S_k I_l] - \gamma[S_k I_l] + \gamma[I_k I_l],$$

with the closure,

$$[A_l B_m C_n] \approx \frac{m-1}{m} \frac{[A_l B_m][B_m C_n]}{[B_m]}.$$

In this thesis, pairwise models are studied in Chapters 2 and 3. We develop new pairwise models for both SIS and SIR dynamics on weighted networks and explore how pairwise models perform. We also investigate how R_0-like thresholds for the SIR epidemic models depend on the weights and their distribution.

1.2.5 Edge-based compartmental model

Another technique which models SIR dynamics on a random network with arbitrary degree distribution is an edge-based compartmental model [76, 78, 79]. This approach uses the probability generating function of the degree distribution, $P(k)$, and it focuses on working out the probability at time t that a random node has not yet been infected by any of its neighbours. Then, the dynamics of the number of susceptible individuals in the network can be explored.

Here, we consider the edge-based compartmental model on a network generated by the configuration model. A random target node, denoted by u, is considered before the model can be formulated. $P(k)$ is the probability that u has degree k. Let $\theta(t)$ be the probability that u has not yet been infected by its neighbours at time t, and $\theta = \phi_S + \phi_I + \phi_R$ where ϕ_x is the probability that a neighbour of u is in state x at time

t, and has not transmitted infection to u. Then, the probability that u with degree k is still susceptible at time t is $\theta(t)^k$. Thus, the number of susceptible nodes in a network at time t can be calculated as

$$S(t) = \sum_k P(k)\theta(t)^k = \psi(\theta(t)).$$

Note that $\sum_k P(k)\theta(t)^k$ is the probability generating function (PGF) [111] of the degree distribution.

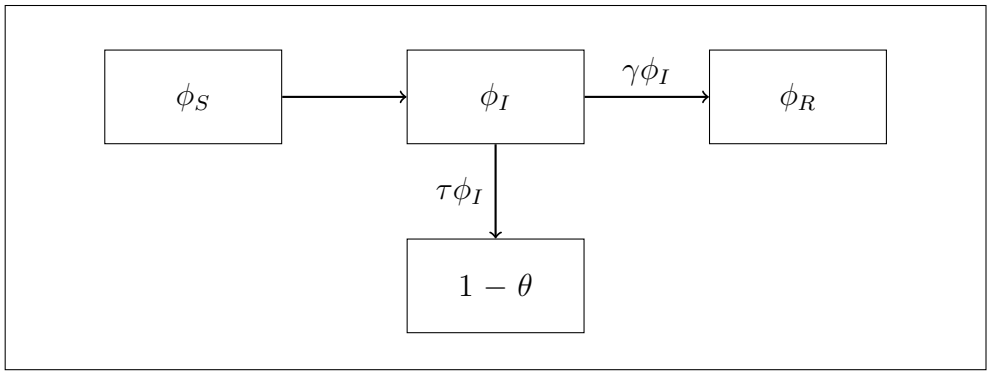


Figure 1.10: Flow diagram of SIR edge-based compartmental model for a configuration network.

Now, we consider the dynamics of the epidemic, and of the function θ. Let τ be the rate of disease transmission across an edge. From Fig. 1.10 it follow immediately that $\dot{\theta}(t) = -\tau\phi_I(t)$, or in other words, the probability that an infectious node has not yet transmitted decays at rate τ. Similarly, $\dot{\phi_R}(t) = \gamma\phi_I(t)$ which accounts for the recovery of an I neighbour of the test node. The two equations above can be combined to give

$$\dot{\phi}_R = -\frac{\gamma}{\tau}\dot{\theta}.$$

Integrating the above from zero to t yields

$$\phi_R = \frac{\gamma}{\tau}(1 - \theta(t)).$$

We also note that $\theta(t_0) = 1$ and $\phi_I = \theta - \phi_S - \phi_R$.

Next, we need to calculate ϕ_S, where ϕ_S denotes the probability that a neighbour of u is susceptible. Let $\langle k \rangle$ be the average nodal degree of the network. The probability that a random neighbour of u has degree k is $kP(k)/\langle k \rangle$ [19], and at time t, this random

neighbour is susceptible with probability $\theta(t)^{k-1}$, as it has $k-1$ other connections. Therefore,

$$\phi_S = \frac{\sum_k kP(k)\theta^{k-1}}{\langle k \rangle}.$$

Using the first derivative of the PGF and $\sum_k kP(k) = \langle k \rangle$, we have

$$\phi_S = \frac{\psi'(\theta)}{\psi'(1)}.$$

Thus, taking all of the above into account, an SIR edge-based compartmental model [78] is given by

$$\begin{aligned} \dot{\theta} &= -\tau\theta + \tau\frac{\psi'(\theta)}{\psi'(1)} + \gamma(1-\theta), \\ \dot{R} &= \gamma I, \qquad S = \psi(\theta), \qquad I = 1 - S - R. \end{aligned}$$

The system of equations above means the model is suitable for analysing important epidemic descriptors, such as the early growth rate and final epidemic size. For example, finding the expected final size, $R(\infty)$, which is equivalent with $I = 0$, we need to find the solution of

$$R(\infty) = 1 - S(\infty) = 1 - \psi(\theta(\infty)),$$

where $\theta(\infty)$ is the solution of $\dot{\theta} = 0$. Using this framework, we develop more complex SIR edge-based compartmental models for weighted networks, as shown in Chapter 3. We also derive analytic calculations of epidemic descriptors and show that this modelling approach agrees well with results from stochastic simulations and pairwise models.

1.2.6 Dynamic networks

In the previous section, we focused on static networks. In a static network, the set of nodes and edges is fixed and does not change over time. This also means that network properties such as the average degree, degree distribution and clustering coefficient remain the same over time. A dynamic network is such that its structure and properties can change over time. For example, links may be added, deleted or rewired. Due to the changes or evolution of the network structure, network properties of a dynamic network will also be changing over time.

From a disease transmission viewpoint, topological properties of nodes and edges can be exploited in order to minimise the impact of epidemics. More recently,

many epidemic studies concentrated on dynamic and time evolving network models [44, 45, 46, 101, 103, 106]. It is widely accepted that during an epidemic, the risk of becoming infected leads to social distancing, with individuals either losing links or simply rewiring [18, 38, 45, 47].

The original model which combines dynamics of the network structure with dynamics on the network in the form of a simple SIS model was proposed by Gross et al. [45]. The epidemic dynamics is specified in terms of infection and recovery events. The rate of transmission across an SI link is denoted by τ. Infected individuals recover independently of each other at rate γ. The network dynamics are specified in terms of rewiring events which affect $S-I$ links. This is usually implemented by susceptible individuals breaking high risk contacts and rewiring to other susceptible individuals chosen at random, as depicted in Fig. 1.11. Here, an $S-I$ link is broken with a rate w and the susceptible individual rewires to another randomly chosen susceptible individual in the network.

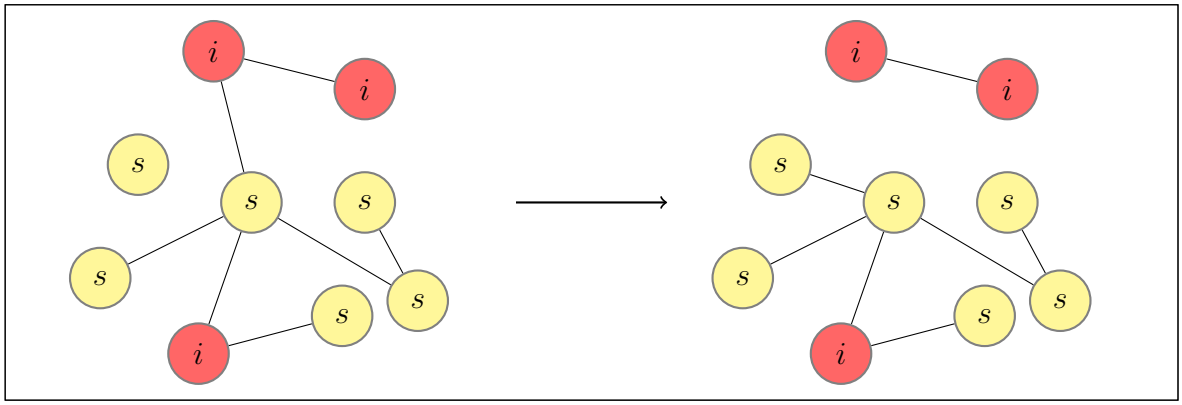


Figure 1.11: An example of a dynamic network where a contact between a susceptible individual and an infectious individual is broken and rewiring occurs between the susceptible individual and another susceptible individual chosen at random from the network.

Following the notations and counting procedures of pair approximation models for SIS dynamics, the original adaptive model can be described by

$$[\dot{S}] = \gamma[I] - \tau[SI],$$

$$[\dot{I}] = \tau[SI] - \gamma[I],$$

$$
\begin{aligned}
[\dot{SI}] &= \gamma([II] - [SI]) + \tau([SSI] - [ISI] - [SI]) - w[SI], \\
[\dot{II}] &= -2\gamma[II] + 2\tau([ISI] + [SI]), \\
[\dot{SS}] &= 2(\gamma + w)[SI] - 2\tau[SSI],
\end{aligned}
$$

where a closure approximation, i.e. $[ABC] \approx \frac{[AB][BC]}{[B]}$, is used to close the model [45, 58]. In the above system of equations, the term describing the destruction of $S - I$ links, denoted by $w[SI]$, is included in the equations describing the rate of change of the expected number of $S - S$ links, and thus describes the instantaneous reconnection/rewiring process. The analysis of this model shows that link rewiring, if high enough, can curtail an epidemic and the model displays a richer spectrum of behaviour including oscillations and bi-stability.

To study the effects of adaptive networks in more depth, various rewiring mechanisms have been explored with a range of assumptions. For example, a study by Risau-Gusman & Zanette [101] presents an SIS model where susceptible nodes reconnect to any node chosen at random, regardless of its state. Furthermore, infected nodes which have links with susceptible nodes broken may also rewire to a new node using the same mechanism. Kiss et al. [63] propose a model in which the connections between nodes are destroyed and rewired depending on the pair type, i.e. $S - I$, $S - S$ and $I - I$, with an associated rate of activation and deletion. Again, both adaptive models described here are derived using the pairwise model framework.

In Chapter 4, we investigate an SIS epidemic spreading on adaptive networks. We make the assumption that susceptible nodes break links with infected nodes independently of distance, and reconnect at random to susceptible nodes available within a given radius. Nodes are placed at random on a square of size $L \times L$ with periodic boundary conditions. We then investigate the impact of rewiring on characteristics of the epidemic and on the network properties, such as degree distribution and clustering coefficient.

1.3 Thesis overview

This thesis is based on four published research papers (3 published and one accepted for publication) focusing on developing epidemic models on networks. Each chapter, apart from the Introduction, corresponds to one of these papers. The thesis is concluded with a discussion of the findings and how this work can be extended for future research in the field of epidemics on networks.

In Chapter 2, we begin by looking at both SIS and SIR disease dynamics on weighted networks. We illustrate how two different methods for choosing link weights can be formulated to model infectious disease spread on networks; (i) random weight distribution and (ii) deterministic weight distribution. We manage to successfully extend the classic pairwise model to weighted networks. We show that our weighted pairwise ODEs for both SIS and SIR epidemics reduce to the original pairwise models under appropriate conditions. A fundamental quantity for epidemic models is the basic reproductive ratio R_0. This is considered both based on the individual or network perspective by using the next generation matrix approach [5] and investigating the R_0-like quantity from the pairwise model by using the approaches proposed by Keeling [58] and Eames [32]. We show that (i) for both network models R_0 is maximised if all weights are equal, and (ii) when the two models are "equally-matched", the networks with a random weight distribution give rise to a higher R_0 value. We illustrate the accuracy of the pairwise approximation models compared to simulations for both SIS and SIR disease dynamics using a variety of different weight distributions.We also explain how disagreements can arise in extreme scenarios of weight distributions.

In Chapter 3, we build on the work in Chapter 2 and consider epidemic dynamics on heterogenous weighted networks. This time, we focus on heterogeneous networks where link weights and node degree are not independent. We construct two network types which depend on how link/edge weights are assigned; (i) network with randomly-distributed edge weights and (ii) network with degree-dependent weights. We develop and analyse the pairwise and edge-based compartmental (EBCM) models, as well as simulation, for SIR-type dynamics to investigate the impact of different weight distributions and of correlation between link weight and degree for both networks. We show that the pairwise, EBCM and simulation demonstrate excellent agreement in describing the evolution of the disease for both networks and for different weight functions. Furthermore, we employ the edge-base modelling approach to derive important epidemic descriptors, such as early growth rate and final epidemic size, and the results are in ex-

cellent agreement with simulations. We also present an analytical calculation of R_0 for both models and discuss the implication of random and correlated weight distributions on this as well as on the time evolution and final outcome of epidemics. Finally, we illustrate that the two seemingly different modelling approaches, the pairwise and the EBCM models, operate on similar assumptions and it is possible to formally link the two.

In Chapter 4, we consider a coupled model of disease and network dynamics. We consider an SIS-type dynamics on an adaptive spatial network with a link or contact rewiring process constrained by spatial proximity. We use two different initial starting networks (i) homogeneous and (ii) heterogeneous Erdős-Rényi networks, where nodes are placed at random on a square of size $L \times L$ with periodic boundary conditions, and we define the local area in terms of circles of certain radii around nodes. We assume that susceptible nodes break links with infected nodes independently of distance, and reconnect at random to susceptible nodes available within a given radius. By systematically manipulating this radius we investigate the impact of rewiring on the structure of the network and characteristics of the epidemic. We adopt a step-by-step approach whereby we first study the impact of rewiring on the network structure in the absence of an epidemic. In this step, the average degree distribution and clustering of both networks at the end of the simulation, or when a steady state has been reached, are explained. We provide both analytic and semi-analytic formulas for the value of clustering achieved in the network. Then, with nodes assigned a disease status but still without disease dynamics, we derive the degree distribution formulas for both networks at time t to explore the impact of the rewiring dynamics, and we show that average degree distributions for susceptible and infectious nodes for both homogeneous and heterogeneous initial network structures display excellent agreement with simulations. Finally we run network and epidemic dynamics simultaneously, and we describe potential outcomes based on the values of the radius. Our results also show that the rewiring radius and the network's initial structure have a pronounced effect on final outcome of the epidemic, with increasingly large rewiring radiuses yielding smaller final endemic equilibria.

In Chapter 5, we take the opportunity to look at mean-field models for the study of SIS type dynamics on networks with heterogenous degree distributions, such as bimodal and truncated power law degree distributions. This paper presents the well-known pairwise [54] and effective degree models [70], and highlights a binomial moment

approximation proposed by Shang in [104]. We show that the pairwise and effective degree models display good agreement with simulations but Shang's model does not. The proposed generalisation performs poorly for all networks proposed by Shang, except for heterogenous networks with relatively high average degree. While the binomial closure gives good results, in that the solution of the full Kolmogorov equations, with the newly proposed infectious rates, agrees well with the closed system, the agreement with simulation is extremely poor. This disagreement invalidates Shang's generalisation and shows that the newly proposed infectious rates do not reflect the true stochastic process unfolding on the network. We conclude that the generalisation proposed by Shang [104] is incorrect and that Shang's simulation method and the excellent agreement with the ODE models is based on flawed or incorrectly implemented simulations.

Chapter 2

Paper 1: A Class of Pairwise Models for Epidemic Dynamics on Weighted Networks

P. Rattana[1], K.B. Blyuss[1], K.T.D. Eames[2] & I.Z. Kiss[1]

[1] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK

[2] The Centre for the Mathematical Modelling of Infectious Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK

2.1 Abstract

In this paper, we study the SIS (susceptible-infected-susceptible) and SIR (susceptible-infected-removed) epidemic models on undirected, weighted networks by deriving pairwise-type approximate models coupled with individual-based network simulation. Two different types of theoretical/synthetic weighted network models are considered. Both start from non-weighted networks with fixed topology followed by the allocation of link weights in either (i) random or (ii) fixed/deterministic way. The pairwise models are formulated for a general discrete distribution of weights, and these models are then used in conjunction with stochastic network simulations to evaluate the impact of different weight distributions on epidemic thresholds and dynamics in general. For the SIR model, the basic reproductive ratio R_0 is computed, and we show that (i) for both network models R_0 is maximised if all weights are equal, and (ii) when the two models are "equally-matched", the networks with a random weight distribution give rise to a higher R_0 value. The models with different weight distributions are also used to explore the agreement between the pairwise and simulation models for different parameter combinations.

2.2 Introduction

Conventional models of epidemic spread consider a host population of identical individuals, each interacting in the same way with each of the others (see [3, 26, 60] and references therein). At the same time, in order to develop more realistic mathematical models for the spread of infectious diseases, it is important to obtain the best possible representation of the transmission mechanism. To achieve this, more recent models have included some of the many complexities that have been observed in mixing patterns. One such approach consists of splitting the population into a set of different subgroups, each with different social behaviours. Even more detail is included within network models that allow differences between individuals to be included. In such models, each individual is represented as a node, and interactions that could permit the transmission of infection appear as edges linking nodes. The last decade has seen a substantial increase in research into how infectious diseases spread over large networks of connected nodes [59, 86], where the networks themselves can represent either small social contact networks [82] or larger scale travel networks [24, 30], including global aviation networks [90, 91]. Importantly, the characteristics of the network, such as the average degree

and the node degree distribution, have a profound effect on the dynamics of infectious disease spread, and hence significant efforts are made to capture properties of realistic contact networks.

One of the common simplifying assumptions of network models is that all links are equally capable of transmitting infection [14, 37, 59, 99]. However, in reality, this is often not the case. Some links will be more likely to transmit infection than others due to closer contacts (e.g. within households [11]) or long-duration interactions [35, 96, 99, 100]. To account for this heterogeneity in properties of social interactions, network models can be adapted, resulting in *weighted contact networks*, where connections between different nodes have different weights. These weights may be associated with the duration, proximity, or social setting of the interaction, and the key point is that they are expected to be correlated with the risk of disease transmission. The precise relationship between the properties of an interaction and its riskiness is hugely complex; here, we will consider a"weight" that is directly proportional to the transmission rate along a link.

A substantial amount of work has been done on the analysis of weighted networks [6, 7, 8, 69] and scale-free networks with different weight distributions [110]. In an epidemiological context, Britton et al. [16] have derived an expression for the basic reproductive ratio in weighted networks with generic distributions of node degree and link weight, and Deijfen [25] has performed a similar analysis to study vaccination in such networks. In terms of practical epidemiological applications, weighted networks have already been effectively used to study control of global pandemics [21, 22, 34] and the spread of animal disease due to cattle movement between farms [41]. Eames et al. [34] have considered an SIR model on an undirected weighted network, where rather than using a theoretical formalism to generate an idealised network, the authors have used social mixing data obtained from questionnaires completed by members of a peer group [96] to construct a realistic weighted network. Having analysed the dynamics of epidemic spread in such a network, they showed how information about node-specific infection risk can be used to develop targeted preventative vaccination strategies. Yang et al. [112] have shown that disease prevalence can be maximized when the edge weights are chosen to be inversely proportional to the degrees of nodes that they link to but, in this case, the transmissibility was not directly proportional to the weights, and weights were also asymmetric. Yang and Zhou [113] have considered SIS epidemics on homogeneous networks with uniform and power law edge weight distributions and

shown how to derive a mean-field description for such models. Furthermore, their simulation results show that the more homogeneous weight distribution leads to higher epidemic prevalence.

In this paper, we consider the dynamics of an infectious disease spreading on weighted networks with different weight distributions. Since we are primarily concerned with the effects of weight distribution on the disease dynamics, the connection matrix will be assumed to be symmetric, representing the situation when the weights can be different for different network edges, but for a given edge the weight is the same irrespective of the direction of infection. From an epidemiological perspective, we consider both the case when the disease confers permanent immunity (represented by an SIR model), and the case when the immunity is short-lived, and upon recovery individuals become susceptible once again (SIS model). For both of these cases, we derive the corresponding ODE-based pairwise models and their closure approximations. Numerical simulations of both the epidemic spread on the network and the pairwise approximations are performed.

The outline of this paper is as follows. In the next section, the construction of specific weighted networks to be used for the analysis of epidemic dynamics is discussed. This is complemented by the derivation of corresponding pairwise models and their closure approximations. Section 3 contains the derivation of the basic reproductive ratio R_0 for the SIR model with different weight distributions, as well as numerical simulations of both stochastic network models and their pairwise ODE counterparts. The paper concludes in Section 4 with discussion of results and possible further extensions of this work.

2.3 Model derivation

2.3.1 Network construction and simulation

There are two conceptually different approaches to constructing weighted networks for modelling infectious disease spread. In the first approach, there is a seed or a primitive motif, and the network is then grown or evolved from this initial seed according to some specific rules. In this method, the topology of the network is co-evolving with the distribution of weights on the edges [7, 8, 9, 69, 112]. Another approach is to consider a weighted network as a superposition of an un-weighted network with a distribution of weights across edges which could be independent of the original network, or it may be

correlated with node metrics, such as their degree [16, 25, 39]. In this paper, we use the second approach in order to investigate the particular role played by the distribution of weights across edges, rather than network topology, in the dynamics of epidemic spread. Besides computational efficiency, this will allow us to make some analytical headway in deriving and analysing low-dimensional pairwise models.

Here, we consider two different methods of assigning weights to network links: a network in which weights are assigned to links at random, and a network in which each node has the same distribution of weighted links connected to it. In reality, there is likely to be a great deal more structure to interaction weights, but in the absence of precise data and also for the purposes of developing models that allow one to explore a number of different assumptions, we make these simplifying approximations.

Random Weight Distribution

First, we consider a simple model of an undirected weighted network with N nodes where the weights of the links can take values w_i with probability p_i, where $i = 1, 2, \ldots, M$. The underlying degree distribution of the corresponding un-weighted network can be chosen to be of the more basic forms, e.g. homogeneous random or Erdős-Rényi-type random networks.

The generation of such networks is straightforward, and weights can be assigned during link creation in the un-weighted network. For example, upon using the configuration model for generating un-weighted networks, each new link will have a weight assigned to it based on the chosen weight distribution. This means that in a homogeneous random network with each node having k links, the distribution of link weights of different types will be multi-nomial, and it is given by

$$P(n_{w_1}, n_{w_2}, \ldots, n_{w_M}) = \frac{k!}{n_{w_1}! n_{w_2}! \ldots n_{w_M}!} p_1^{n_1} p_2^{n_2} \ldots p_M^{n_M}, \tag{2.1}$$

where, $n_{w_1} + n_{w_2} + \cdots + n_{w_M} = k$ and $P(n_{w_1}, n_{w_2}, \ldots, n_{w_M})$ stands for the probability of a node having n_{w_1}, n_{w_2}, ..., n_{w_M} links with weights w_1, w_2, ..., w_M, respectively. While the above expression is applicable in the most general set-up, it is worth considering the case of weights of only two types, where the distribution of link weights for a homogenous random network becomes binomial

$$P(n_{w_1}, n_{w_2} = k - n_{w_1}) = \binom{k}{n_{w_1}} p_1^{n_1} (1 - p_1)^{k - n_1}, \tag{2.2}$$

where $p_1 + p_2 = 1$ and $n_{w_1} + n_{w_2} = k$. The average link weight in the model above can be easily found as

$$w_{av}^{random} = \sum_{i=1}^{M} p_i w_i,$$

which for the case of weights of two types w_1 and w_2 reduces to

$$w_{av}^{(2r)} = p_1 w_1 + p_2 w_2 = p_1 w_1 + (1 - p_1) w_2.$$

Fixed Deterministic Weight Distribution

As a second example, we consider a network, in which each node has k_i links with weight w_i $(i = 1, 2, \ldots, M)$, where $k_1 + k_2 + \cdots + k_M = k$. The different weights here could be interpreted as being associated with different types of social interaction: e.g. home, workplace, and leisure contacts, or physical and non-physical interactions. In this model, all individuals are identical in terms of their connections, not only having the same number of links (as in the model above), but also having the same set of weights. The average weight in such a model is given by

$$w_{av}^{fixed} = \sum_{i=1}^{M} p_i w_i, \qquad p_i = \frac{k_i}{k},$$

where p_i is the fraction of links of type i for each node. In the case of links of two types with weights w_1 and w_2, the average weight becomes

$$w_{av}^{(2f)} = p_1 w_1 + p_2 w_2 = \frac{k_1}{k} w_1 + \frac{k_2}{k} w_2 = \frac{k_1}{k} w_1 + \frac{k - k_1}{k} w_2.$$

Simulation of Epidemic Dynamics

In this study, the simple SIS and SIR epidemic models are considered. The epidemic dynamics are specified in terms of infection and recovery events. The rate of transmission across an un-weighted edge between an infected and susceptible individual is denoted by τ. This will then be adjusted by the weight of the link which is assumed to be directly proportional to the strength of the transmission along that link. Infected individuals recover independently of each other at rate γ. The simulation is implemented using the Gillespie algorithm [43] with inter-event times distributed exponentially with a rate given by the total rate of change in the network, with the single event to be implemented at each step being chosen at random and proportionally to its rate. All simulations start with most nodes being susceptible and with a few infected nodes chosen at random.

2.3.2 Pairwise Equations and Closure Relations

In this section, we extend the classic pairwise model for un-weighted networks [58, 94] to the case of weighted graphs with M different link-weight types. Pairwise models successfully interpolate between classic compartmental ODE models and full individual-based network simulations with the added advantage of high transparency and a good degree of analytical tractability. These qualities make them an ideal tool for studying dynamical processes on networks [32, 47, 54, 58], and they can be used on their own and/or in parallel with simulation. The original versions of the pairwise models have been successfully extended to networks with heterogenous degree distribution [33], asymmetric networks [105] and situations where transmission happens across different/combined routes [32, 47] as well as when taking into consideration network motifs of higher order than pairs and triangles [52]. The extension that we propose is based on the previously established precise counting procedure at the level of individuals, pairs, and triples, as well as on a careful and systematic account of all possible transitions needed to derive the full set of evolution equations for singles and pairs. These obviously involve the precise dependency of lower order moments on higher order ones, e.g. the rate of change of the expected number of susceptible nodes is proportional to the expected number of links between a susceptible and infected node. We extend the previously well-established notation [58] to account for the added level of complexity due to different link weights. In line with this, the number of singles remains unchanged, with $[A]$ denoting the number of nodes across the whole network in state A. Pairs of type $A-B$, $[AB]$, are now broken down depending on link weights, i.e. $[AB]_i$ represents the number of links of type $A-B$ with the link having weight w_i, where as before $i = 1, 2, \ldots, M$ and $A, B \in \{S, I, R\}$ if an SIR dynamics is used. As before, links are doubly counted (e.g. in both directions), and thus the following relations hold: $[AB]_m = [BA]_m$ and $[AA]_m$ is equal to twice the number of uniquely counted links of weight w_m with nodes at both ends in state A. From this extension, it follows that $\sum_{i=1}^{M}[AB]_i = [AB]$. The same convention holds at the level of triples where $[ABC]_{mn}$ stands for the expected number of triples where a node in state B connects nodes in state A and C via links of weight w_m and w_n, respectively. The weight of the link impacts on the rate of transmission across the link, and this is achieved by using a link-specific transmission rate equal to τw_i, where $i = 1, 2, \ldots, M$. In line with the above, we construct two pairwise models, one for SIS and one for SIR dynamics.

The pairwise model for the SIS dynamics can be written in the form:

$$[\dot{S}] = \gamma[I] - \tau\textstyle\sum_{n=1}^{M} w_n[SI]_n,$$

$$[\dot{I}] = \tau\textstyle\sum_{n=1}^{M} w_n[SI]_n - \gamma[I],$$

$$[\dot{SI}]_m = \gamma([II]_m - [SI]_m) + \tau\textstyle\sum_{n=1}^{M} w_n([SSI]_{mn} - [ISI]_{nm}) - \tau w_m[SI]_m, \qquad (2.3)$$

$$[\dot{II}]_m = -2\gamma[II]_m + 2\tau\textstyle\sum_{n=1}^{M} w_n[ISI]_{nm} + 2\tau w_m[SI]_m,$$

$$[\dot{SS}]_m = 2\gamma[SI]_m - 2\tau\textstyle\sum_{n=1}^{M} w_n[SSI]_{mn},$$

where $m = 1, 2, 3, ..., M$ and infected individuals recover at rate γ. When recovered individuals have life-long immunity, we have the following system of equations describing the dynamics of a pairwise SIR model:

$$[\dot{S}] = -\tau\textstyle\sum_{n=1}^{M} w_n[SI]_n,$$

$$[\dot{I}] = \tau\textstyle\sum_{n=1}^{M} w_n[SI]_n - \gamma[I],$$

$$[\dot{R}] = \gamma[I],$$

$$[\dot{SS}]_m = -2\tau\textstyle\sum_{n=1}^{M} w_n[SSI]_{mn},$$

$$[\dot{SI}]_m = \tau\textstyle\sum_{n=1}^{M} w_n([SSI]_{mn} - [ISI]_{nm}) - \tau w_m[SI]_m - \gamma[SI]_m, \qquad (2.4)$$

$$[\dot{SR}]_m = -\tau\textstyle\sum_{n=1}^{M} w_n[ISR]_{nm} + \gamma[SI]_m,$$

$$[\dot{II}]_m = 2\tau\textstyle\sum_{n=1}^{M} w_n[ISI]_{nm} + 2\tau w_m[SI]_m - 2\gamma[II]_m,$$

$$[\dot{IR}]_m = \tau\textstyle\sum_{n=1}^{M} w_n[ISR]_{nm} + \gamma([II]_m - [IR]_m),$$

$$[\dot{RR}]_m = \gamma[IR]_m,$$

where again $m = 1, 2, 3, ..., M$ with the same notation as above. As a check and reference to previous pairwise models, in Appendix A we show how systems (2.3) and (2.4) reduce to the standard un-weighted pairwise SIS and SIR model [58] when all weights are equal to each other, $w_1 = w_2 = \cdots = w_M = W$.

The above systems (i.e. Eqs. (2.3) and (2.4)) are not closed, as equations for the pairs require knowledge of triples, and thus, equations for triples are needed. This dependency on higher-order moments can be curtailed by closing the equations via approximating triples in terms of singles and pairs [58]. For both systems, the agreement with simulation will heavily depend on the precise distribution of weights across the links, the network topology, and the type of closures that will be used to capture essential features of network structure and the weight distribution. A natural extension of the classic closure is given by

$$[ABC]_{mn} = \frac{k-1}{k}\frac{[AB]_m[BC]_n}{[B]}, \tag{2.5}$$

where k is the number of links per node for a homogeneous network, or the average nodal degree for networks with other than homogenous degree distributions. However, even for the simplest case of homogeneous random networks with two weights (i.e. w_1 and w_2), the average degree is split according to weight. Namely, the average number of links of weight w_1 across the whole network is $k_1 = p_1 k \leq k$, and similarly, the average number of links of weight w_2 is $k_2 = (1-p_1)k \leq k$, where $k = k_1 + k_2$. Attempting to better capture the additional network structure generated by the weights, the closure relation above can be recast to give the following, potentially more accurate, closures

$$\begin{aligned}
[ABC]_{11} &= [AB]_1(k_1-1)\frac{[BC]_1}{k_1[B]} = \frac{k_1-1}{k_1}\frac{[AB]_1[BC]_1}{[B]}, \\
[ABC]_{12} &= [AB]_1 k_2 \frac{[BC]_2}{k_2[B]} = \frac{[AB]_1[BC]_2}{[B]}, \\
[ABC]_{21} &= [AB]_2 k_1 \frac{[BC]_1}{k_1[B]} = \frac{[AB]_2[BC]_1}{[B]}, \\
[ABC]_{22} &= [AB]_2(k_2-1)\frac{[BC]_2}{k_2[B]} = \frac{k_2-1}{k_2}\frac{[AB]_2[BC]_2}{[B]},
\end{aligned} \tag{2.6}$$

where, as in Eq. (2.5), the form of the closure can be derived by considering the central individual in the triple, B. The first pair of the triple ($[AB]_i$) effectively "uses up" one of B's links of weight w_i. For triples of the form $[ABC]_{11}$, the presence of the pair $[AB]_1$ means that B has $(k_1 - 1)$ remaining links of weight w_1 that could potentially connect to C. For triples of the form $[ABC]_{12}$, however, B has k_2 weight w_2 links that could potentially connect to C. Furthermore, expressions such as $\frac{[BC]_i}{k_i[B]}$ denote the fraction of B's edges of weight w_i that connect to an individual of type C. The specific choice of

closure will depend on the structure of the network and, especially, on how the weights are distributed. For example, for the case of the homogeneous random networks with links allocated randomly, both closures offer a viable option. For the case of a network where each node has a fixed pre-allocated number of links with different weights, e.g. k_1 and k_2 links with weights w_1 and w_2, respectively, the second closure (2.6) offers the more natural/intuitive avenue toward closing the system and obtaining good agreement with network simulation.

2.4 Results

In this section, we present analytical and numerical results for weighted networks and pairwise representations of SIS and SIR models in the case of two different link-weight types (i.e. w_1 and w_2).

2.4.1 Threshold Dynamics for the SIR Model - the Network Perspective

The basic reproductive ratio, R_0 (the average number of secondary cases produced by a typical index case in an otherwise susceptible population), is one of the most fundamental quantities in epidemiology [3, 27]. Besides informing us on whether a particular disease will spread in a population, as well as quantifying the severity of an epidemic outbreak, it can be also used to calculate a number of other important quantities that have good intuitive interpretation. In what follows, we will compute R_0 and R_0-like quantities and will discuss their relation to each other, and also issues around these being model-dependent. First, we compute R_0 from an individual-based or network perspective by employing the next generation matrix approach as used in the context of models with multiple transmission routes, such as household models [5].

Random Weight Distribution: First, we derive an expression for R_0 when the underlying network is homogeneous, and the weights of the links are assigned at random according to a prescribed weight distribution. In the spirit of the proposed approach, the next generation matrix can be easily computed to yield

$$NGM = (a_{ij})_{i,j=1,2} = \begin{vmatrix} (k-1)p_1r_1 & (k-1)p_1r_1 \\ (k-1)p_2r_2 & (k-1)p_2r_2 \end{vmatrix},$$

where

$$r_1 = \frac{\tau w_1}{\tau w_1 + \gamma}, \quad r_2 = \frac{\tau w_2}{\tau w_2 + \gamma}$$

represent the probability of transmission from an infected to a susceptible across a link of weight w_1 and w_2, respectively. Here, the entry a_{ij} stands for the average number of infections produced via links of type i (i.e. with weight w_i) by a typical infectious node who itself has been infected across a link of type j (i.e. with weight w_j). Using the fact that $p_2 = 1 - p_1$, the basic reproductive ratio can be found from the leading eigenvalue of the NGM matrix as follows:

$$R_0^1 = (k-1)(p_1 r_1 + (1-p_1) r_2). \tag{2.7}$$

In fact, the expression for R_0 can be generalised to more than two weights to give $R_0 = (k-1)\sum_{i=1}^{M} p_i r_i$, where w_m has frequency given by p_m with the constraint that $\sum_{i=1}^{M} p_i = 1$. It is straightforward to show that upon assuming uniform weight distribution $w_i = W$ for $i = 1, 2, \ldots, M$, the basic reproduction number on a homogeneous graph reduces to $R_0 = (k-1)r$ as expected, where $r = \tau W/(\tau W + \gamma)$.

Deterministic Weight Distribution: The case when the number of links with given weights for each node is fixed can be captured with the same approach, and the next generation matrix can be constructed as follows:

$$NGM = \begin{vmatrix} (k_1 - 1)r_1 & k_1 r_1 \\ k_2 r_2 & (k_2 - 1)r_2 \end{vmatrix}.$$

As before, the leading eigenvalue of the NGM matrix yields the basic reproductive ratio:

$$R_0^2 = \frac{(k_1-1)r_1 + (k_2-1)r_2 + \sqrt{[(k_1-1)r_1 - (k_2-1)r_2]^2 + 4k_1 k_2 r_1 r_2}}{2}. \tag{2.8}$$

It is worth noting that the calculations above are a direct result of a branching process approximation of the pure transmission process which differentiates between individuals depending on whether they were infected via a link of weight w_1 or w_2, with an obvious generalisation to more than two weights. This separation used in the branching process leads to the offspring or next generation matrix of the branching process [5]. Using the two expressions for the basic reproductive ratio, it is possible to prove the following result.

Theorem 1. *Given the setup for the fixed weight distribution and using* $p_1 = k_1/k$, $p_2 = k_2/k$ *and* $k_1 + k_2 = k$, *if* $1 \leq k_1 \leq k-1$ *(which implies that* $1 \leq k_2 \leq k-1$*), then* $R_0^2 \leq R_0^1$.

The proof of this result is sketched out in Appendix B. This Theorem effectively states that provided each node has at least one link of type 1 and one link of type 2, then independently of disease parameters, it follows that the basic reproductive ratio as computed from Eq. (2.7) always exceeds or is equal to an equivalent R_0 computed from Eq. (2.8).

It is worth noting that both R_0 values reduce to

$$R_0^1 = R_0^2 = R_0 = (k-1)r = \frac{(k-1)\tau W}{\tau W + \gamma}, \tag{2.9}$$

if one assumes that weights are equal, i.e. $w_1 = w_2 = W$. As one would expect, the first good indicator of the impact of weights on the epidemic dynamics will be the average weight. Hence, it is worth considering the problem of maximising the values R_0 under assumption of a fixed average weight:

$$p_1 w_1 + p_2 w_2 = W. \tag{2.10}$$

Under this constraint, the following statement holds.

Theorem 2. *For weights constrained by* $p_1 w_1 + p_2 w_2 = W$ *(or* $(k_1/k)w_1 + (k_2/k)w_2 = W$ *for a fixed weights distribution),* R_0^1 *and* R_0^2 *attain their maxima when* $w_1 = w_2 = W$, *and the maximum values for both is* $R_0 = (k-1)r = \dfrac{(k-1)\tau W}{\tau W + \gamma}$.

The proof of this result is presented in Appendix C.

The above results suggest that for the same average link weight and when the one-to-one correspondence between p_1 and k_1/k, and p_2 and k_2/k holds, the basic reproductive ratio is higher on networks with random weight distribution than on networks with a fixed weight distribution. This, however, does not preclude the possibility of having a network with random weight distribution with smaller average weight exhibiting an R_0 value that is bigger than the R_0 value corresponding to a network where weights are fixed and the average weight is higher. The direct implication is that it is not sufficient to know just the average link weight in order to draw conclusions about possible epidemic outbreaks on weighted networks; rather one has to know the precise weight distribution that provides a given average weight.

Figure 2.1 shows how the basic reproductive ratio changes with the transmission rate τ for different weight distributions. When links on a homogeneous network are distributed at random (upper panel), the increase in the magnitude of one specific link weight (e.g. w_1) accompanied by a decrease in its frequency leads to smaller R_0 values. This is to be expected since the contribution of the different link types in this case is kept constant ($p_1 w_1 = p_2 w_2 = 0.5$) and this implies that the overall weight of the network links accumulates in a small number of highly weighted links with most links displaying small weights and thus making transmission less likely. The statement above is more rigorously underpinned by the results of Theorem 1 and 2, which clearly show that equal or more homogeneous weights lead to higher values of the basic reproductive ratio. For the case of fixed weight distribution (lower panel), the changes in the value of R_0 are investigated in terms of varying the weights, so that the overall weight in the network remains constant. This is constrained by fixing values of p_1 and p_2 and, in this case, the highest values are obtained for higher values of w_1. The flexibility here is reduced due to p_1 and p_2 being fixed, and a different link breakdown may lead to different outcomes. The top continuous line in Fig. 2.1 (upper panel) corresponds to the maximum R_0 value achievable for both models if the $p_1 w_1 + p_2 w_2 = 1$ constraint is fulfilled.

2.4.2 R_0-like Threshold for the SIR Model - a Pairwise Model Perspective

To compute the value of R_0-like quantity from the pairwise model, we use the approach suggested by Keeling [58], which utilises the local spatial/network structure and correctly accounts for correlations between susceptible and infectious nodes early on in the epidemics. This can be achieved by looking at the early behaviour of $[SI]_1/[I] = \lambda_1$ and $[SI]_2/[I] = \lambda_2$ when considering links of only two different weights. In line with Eames [32], we start from the evolution equation of $[I]$

$$[\dot{I}] = (\tau w_1 [SI]_1/[I] + \tau w_2 [SI]_2/[I] - \gamma)[I],$$

where from the growth rate $\tau w_1 \lambda_1 + \tau w_2 \lambda_2 - \gamma$ it is easy to define the threshold quantity R as follows:

$$R = \frac{\tau w_1 \lambda_1 + \tau w_2 \lambda_2}{\gamma}. \tag{2.11}$$

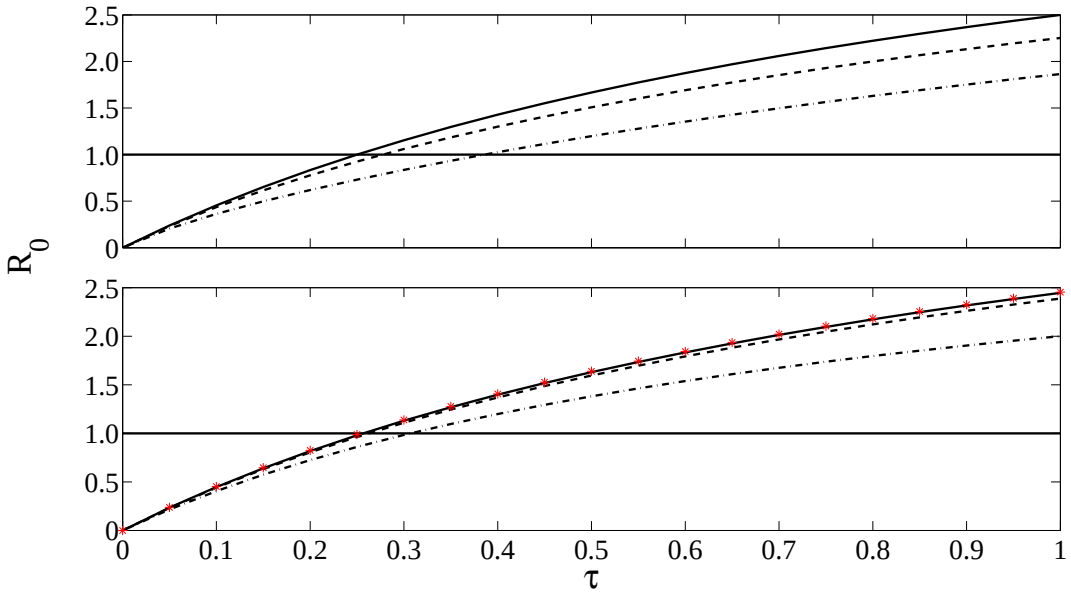


Figure 2.1: Basic reproductive ratio R_0 for random (upper) and deterministic (lower) weight distributions with different weight and weight frequency combinations, but with $p_1w_1 + p_2w_2 = 1$. *Upper panel:* the case of homogenous networks with weights assigned at random considers the situation where the contribution of the two different weight types is equal ($p_1w_1 = p_2w_2 = 0.5$) but with weight w_1 increasing and its frequency decreasing (top to bottom with $(p_1, w_1) = \{(0.5, 1), (0.2, 2.5), (0.05, 10)\}$). Increasing the magnitude of weights, but reducing their frequency leads to smaller R_0 values. *Lower panel:* the case of homogeneous networks with fixed number of links of type w_1 and w_2 illustrates the situation where w_1 increases while $p_1 = k_1/k = 1/3$ and $p_2 = (k - k_1)/k = 2/3$ remain fixed (bottom to top with $w_1 = \{0.1, 0.5, 1.4\}$). Here the opposite tendency is observed with increasing weights leading to higher R_0 values. Finally, for the randomly distributed weights case, setting $p_1 = 1/3, w_1 = 1.4$ and observing $p_1w_1 + p_2w_2 = 1$, we obtain R_0 ($\star$) values which compare almost directly to the fixed-weights case (top continuous line). Other parameters are set to $k = 6$, $k_1 = 2$ and $\gamma = 1$.

For the classic closure (2.5), one can compute the early quasi-equilibria for λ_1 and λ_2 directly from the pairwise equations as follows:

$$\lambda_1 = \frac{\gamma(k-1)p_1 R}{\tau w_1 + \gamma R} \text{ and } \lambda_2 = \frac{\gamma(k-1)(1-p_1)R}{\tau w_2 + \gamma R}.$$

Substituting these into Eq. (2.11) and solving for R yields

$$R = \frac{R_1 + R_2 + \sqrt{(R_1+R_2)^2 + 4R_1R_2Q}}{2}, \tag{2.12}$$

where

$$R_1 = \frac{\tau w_1[(k-1)p_1 - 1]}{\gamma}, \quad R_2 = \frac{\tau w_2[(k-1)p_2 - 1]}{\gamma},$$

$$Q = \frac{k-2}{[(k-1)p_1 - 1][(k-1)p_2 - 1]},$$

with details of all calculations presented in Appendix D. We note that $R > 1$ will result in an epidemic, while $R < 1$ will lead to the extinction of the disease. It is straightforward to show that for equal weights, say W, the expression above reduces to $R = \tau W(k-2)/\gamma$ which is in line with R_0 value in [58] for un-clustered, homogeneous networks. Under the assumption of a fixed total weight W, one can show that similarly to the network-based basic reproductive ratio, R achieves its maximum when $w_1 = w_2 = W$.

In a similar way, for the modified closure (2.6), we can use the same methodology to derive the threshold quantity as

$$R = \frac{R_1 + R_2 + \sqrt{(R_1+R_2)^2 + 4R_1R_2(Q-1)}}{2}, \tag{2.13}$$

where

$$R_1 = \frac{\tau w_1(k_1-2)}{\gamma}, \quad R_2 = \frac{\tau w_2(k_2-2)}{\gamma}, \quad Q = \frac{k_1k_2}{(k_1-2)(k_2-2)}.$$

For this closure once again, $R > 1$ results in an epidemic, while for $R < 1$, the disease dies out. Details of these calculations are shown in Appendix D. It is noteworthy that one can derive expressions (2.12) and (2.13) by considering the leading eigenvalue based on the linear stability analysis of the disease-free steady state of system (2.4) with the corresponding pairwise closures given in Eqs. (2.5) and (2.6).

Finally, we note that this seemingly R_0-lookalike, $R = \tau W(k-2)/\gamma$ for the equal weights case $w_1 = w_2 = W$ is a multiple of $(k-2)$ as opposed to $(k-1)$ as is the case for the R_0 derived based on the individual-based perspective, where, for equal weights,

$R_0^1 = R_0^2 = \tau W(k-1)/(\tau W + \gamma)$. This highlights the importance, in models that are based on an underlying network of population interactions, of the way in which an R_0-like quantity is defined. In simple mass-action-type models the same value is derived irrespective of whether R_0 is thought of as the number of new cases from generation-to-generation (the NGM method), or as the growth rate of the epidemic scaled by the infectious period. In a network model, the two approaches have the same threshold behaviour, but the clusters of infection that appear within the network mean that they produce different values away from the threshold. It is important therefore to be clear about what we mean by "R_0" in a pair-approximation model. It is also important when using empirically-derived R_0 values to inform pairwise models to be clear about how these values were estimated from epidemiological data, and to consider which is the most appropriate way to incorporate the information into the model.

2.4.3 The Performance of Pairwise Models and the Impact of Weight Distributions on the Dynamics of Epidemics

To evaluate the accuracy of the pairwise approximation models, we will now compare numerical solutions of models (2.3) and (2.4) (with closures given by Eq. (2.5) and Eq. (2.6) for random and deterministic weight distributions, respectively) to results obtained from the corresponding network simulation. The discussion around the comparison of the two models is interlinked with the discussion of the impact of different weight distributions/patterns on the overall epidemic dynamics. We begin our numerical investigation by considering weight distributions with moderate heterogeneity. This is illustrated in Fig. 2.2, where excellent agreement between simulation and pairwise models is obtained. The agreement remains valid for both SIS and SIR dynamics, and networks with higher average link weight lead to higher prevalence levels at equilibrium for SIS and higher infectiousness peaks for SIR.

Next, we explore the impact of weight distribution under the condition that the average weight remains constant (i.e. $p_1 w_1 + p_2 w_2 = 1$, where without loss of generality the average weight has been chosen to be equal to 1). First, we keep the proportion of edges of type one (i.e. with weight w_1) fixed and change the weight itself by gradually increasing its magnitude. Due to the constraint on the average weight and the condition $p_2 = 1 - p_1$, the other descriptors of the weight distribution follow. Figure 2.3 shows that concentrating a large portion of the total weight on a few links leads to smaller epidemics, since the majority of links are low-weight and thus have a small potential to

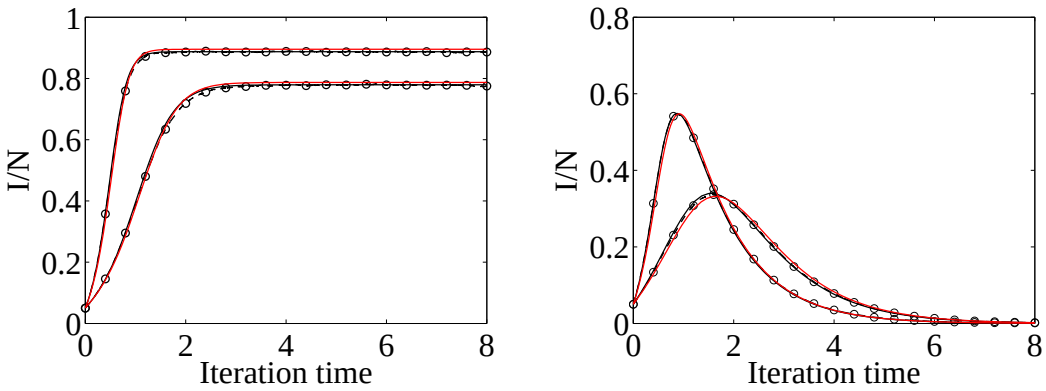


Figure 2.2: The infection prevalence (I/N) from the pairwise and simulation models for homogeneous random networks with random weight distribution (ODE: solid line, simulation: dashed line and (o)). All nodes have degree $k = 5$ with $N = 1000$, $I_0 = 0.05N$, $\gamma = 1$ and $\tau = 1$. From top to bottom, the parameter values are: $w_1 = 5, p_1 = 0.2, w_2 = 1.25, p_2 = 0.8$ (top), and $w_1 = 0.5, p_1 = 0.5, w_2 = 1.5, p_2 = 0.5$ (bottom). The left and right panels represent the SIS and SIR dynamics, respectively.

transmit the disease. This effect is exacerbated for the highest value of w_1; in this case, 95% of the links are of weight $w_2 = (1 - p_1 w_1)/(1 - p_1) = 0.5/0.95$ leading to epidemics of smallest impact (Fig. 2.3(a)) and smallest size of outbreak (Fig. 2.3(b)).

While the previous setup kept the frequency of links constant while changing the weights, one can also investigate the impact of keeping at least one of the weights constant (e.g. the larger one) and changing its frequency. To ensure a meaningful comparison, here we also require that the average link weight over the whole network is kept constant. When such highly weighted links are rare, the system approaches the non-weighted network limit where the transmission rate is simply scaled by w_2 (the most abundant link type). As Fig. 2.4 shows, in this case, the agreement is excellent, and as the frequency of the highly weighted edges/links increases, disease transmission is less severe.

Regarding the comparison of the pairwise and simulation models, we note that while the agreement is generally good for a large part of the disease and weight parameter space, the more extreme scenarios of weight distribution result in poorer agreement. This is illustrated in both Figs. 2.3 and 2.4 (see bottom curves), with the worst agreement for the SIS dynamics. The insets in Fig. 2.3 show that increasing the average connectivity improves the agreement. However, the cause of disagreement is due to a more subtle effect driven also by the weight distribution. For example, in Fig. 2.4, the

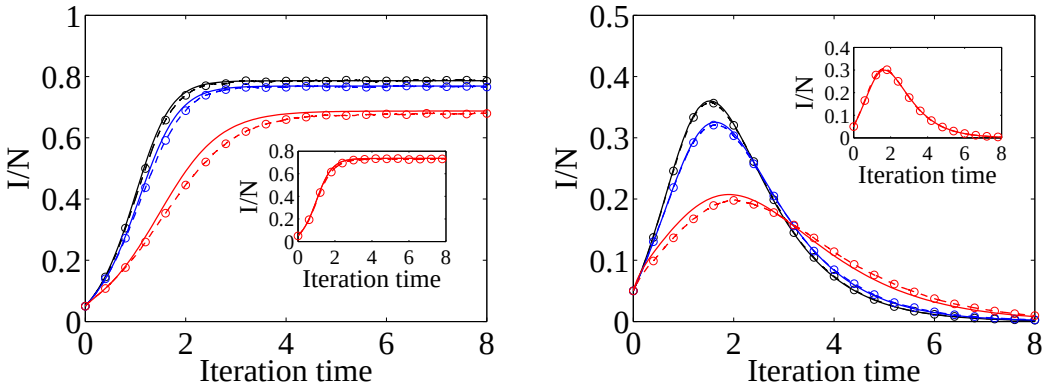


Figure 2.3: The infection prevalence (I/N) from the pairwise and simulation models for homogenous networks with random weight distribution (ODE: solid line, simulation: dashed line and (o)). All numerical tests use $N = 1000$, $I_0 = 0.05N$, $k = 5$, $\gamma = 1, \tau = 1$ and $p_1 = 0.05$ $(p_2 = 1 - p_1 = 0.95)$. From top to bottom, $w_1 = 2.5, 5, 10$, $w_2 = 0.875/0.95, 0.75/0.95, 0.5/0.95$. The weight distributions are chosen such that the average link weight, $p_1 w_1 + p_2 w_2 = 1$, remains constant. Insets of (a) and (b): the same parameter values as for the lowest prevalence plots but, with $k = 10$ and $\tau = 0.5$. The left and right panel represent the SIS and SIR dynamics, respectively.

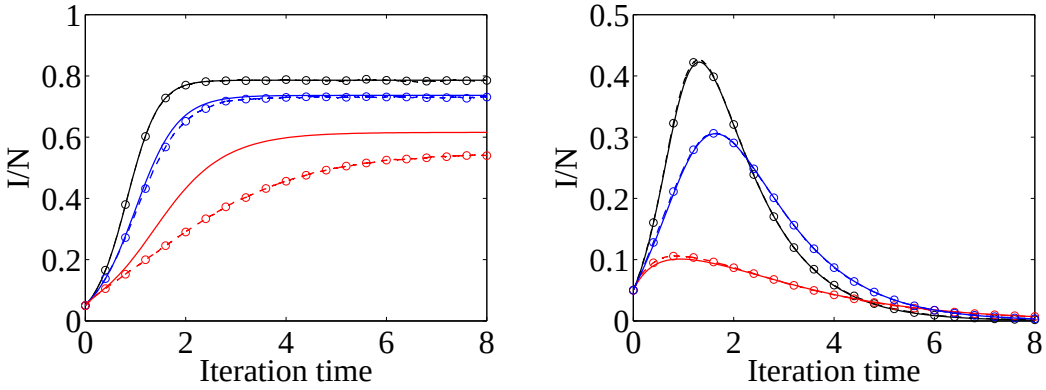


Figure 2.4: The infection prevalence (I/N) from the pairwise and simulation model for homogenous networks with random weight distribution (ODE: solid line, simulation: dashed line and (o)). All numerical tests use $N = 1000$, $I_0 = 0.05N$, $k = 10$, $\gamma = 1, \tau = 0.5$ and $w_1 = 10$. From top to bottom, $P(w_1) = 0.01, 0.05, 0.09$, $w_2 = 0.9/0.99, 0.5/0.95, 0.1/0.91$. Here, also $p_2 = 1 - p_1$ and $p_1 w_1 + p_2 w_2 = 1$. The left and right panels represent the SIS and SIR dynamics, respectively.

average degree in the network is 10, higher then used previously and equal to that in the insets from Fig. 2.3, but despite this, the agreement is still poor.

The two different weighted network models are compared in Fig. 2.5. This is done by using the same link weights and setting $p_1 = k_1/k$ and $p_2 = k_2/k$. Epidemics on networks with random weight distributions grow faster and, given the same time scales of the epidemic, this is in line with results derived in Theorem 1 and 2 and findings concerning the growth rates. The difference is less marked for larger values of τ where a significant proportion of the nodes becomes infected.

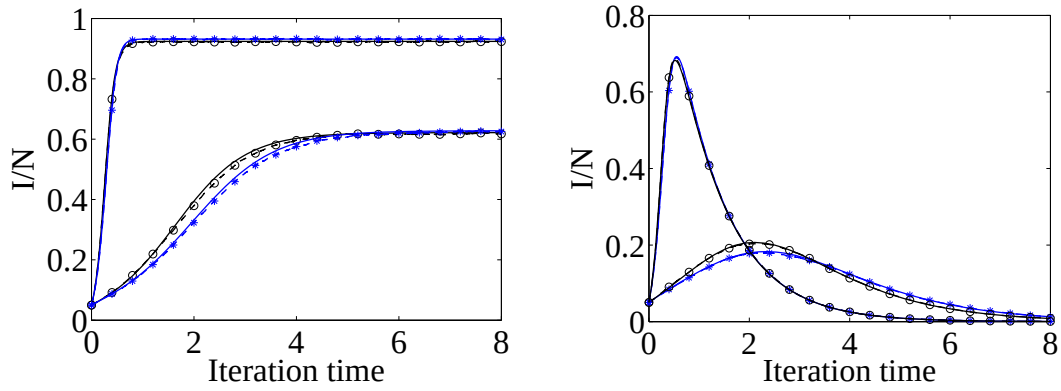


Figure 2.5: The infection prevalence (I/N) based on random (model 1) and fixed (model 2) weight distribution (ODE: black (1) and blue (2) solid line, simulation results: same as ODE but dashed lines, and ($\circ$) and ($*$)). All numerical tests use $N = 1000$, $I_0 = 0.05N$, $k = 10$, $k_1 = 2$, $k_2 = 8$, $p_1 = k_1/k$, $p_2 = k_2/k$, $w_1 = 10$, $w_2 = 1.25$ and $\gamma = 1$. The rate of infection $\tau = 0.5$ (top) and $\tau = 0.1$ (bottom). The left and right panels represent the SIS and SIR dynamics, respectively.

In Fig. 2.6, the link weight composition is altered by decreasing the proportion of highly-weighted links. As expected, the reduced average link weight across the network leads to smaller epidemics while keeping the excellent agreement between simulation and pairwise model results.

2.5 Discussion

The present study has explored the impact of weight heterogeneity and highlighted that the added heterogeneity of link weights does not manifest itself in the same way as most other heterogeneities in epidemic models on networks. Usually, heterogeneities lead to an increase in R_0 but potentially to a fall in the final epidemic size [65]. However,

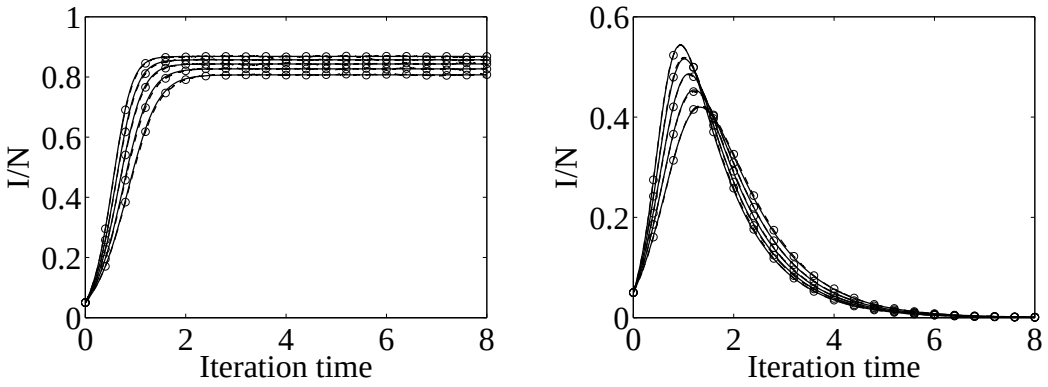


Figure 2.6: The infection prevalence (I/N) for a fixed weight distribution (ODE: solid lines, simulation results: dashed lines and (o)). All numerical tests use $N = 1000$, $I_0 = 0.05N$, $k = 6$, $\gamma = 1, \tau = 1$ and $w_1 = 1.4, w_2 = 0.8$. From top to bottom: $k_1 = 5, 4, 3, 2, 1$ and $k_2 = k - k_1$. The left and right panels represent the SIS and SIR dynamics, respectively.

for weighted networks the concentration of infectiousness on fewer target links, and thus target individuals, leads to a fall in R_0 for both homogeneous random and fixed weight distribution models. Increased heterogeneity in weights accentuates the locality of contact and is taking the model further from the mass-action type models. Infection is concentrated along a smaller number of links, which results in wasted infectivity and lower R_0. This is in line with similar results [16, 17, 113], where different modelling approaches have been used to capture epidemics on weighted networks.

The models proposed in this paper are simple mechanistic models with basic weight distributions, but despite their simplicity they provide a good basis for analysing disease dynamics on weighted networks in a rigorous and systematic way. The modified pairwise models have performed well, and provide a good approximation to direct simulation. As expected, the agreement with simulations typically breaks down at or close to the threshold but, away from it, pairwise models provide a good counterpart or alternative to simulation. Disagreement only appears for extreme weight distributions, and we hypothesise that this is mainly due to the network becoming more modular with islands of nodes connected by links of low weight being bridged together by highly weighted links. A good analogy to this is provided by considering the case of a pairwise model on un-weighted networks specified in terms of two network metrics, node number N and average number of links k. The validity of the pairwise model relies on the network being connected up at random, or according to the configuration model.

This can be easily broken by creating two sub-networks of equal size both exhibiting the same average connectivity. Simulations on such type of networks will not agree with the pairwise model, highlighting that the network generating algorithm can push the network out of the set of 'acceptable' networks. We expect that this or a similar argument can more precisely explain why the agreement breaks down for significant link-weight heterogeneity.

The usefulness of pairwise models is illustrated in Fig. 2.7, where the I/N values are plotted for a range of τ values and for different weight distributions. Here, the equilibrium value has been computed by finding the steady state directly from the ODEs (2.3) by finding numerically the steady state solution of a set on non-linear equations (i.e. $[\dot{A}] = 0$ and $[\dot{AB}] = 0$). To test the validity, the long term solution of the ODE is plotted along with results based on simulation. The agreement away from the threshold is excellent and illustrates clearly the impact of different weight distributions on the magnitude of the endemic threshold.

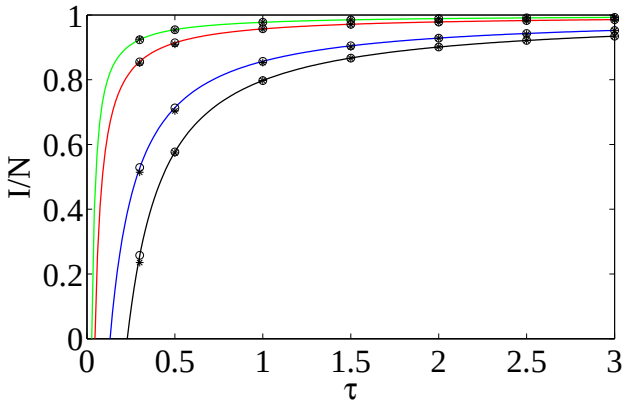


Figure 2.7: Endemic steady state from the SIS model on networks with random weight distribution. The continuous lines correspond to the steady state computed numerically by setting all evolution equations in the pairwise system to zero. These are complemented by finding the endemic steady state through direct integration of the ODE system for a long-enough time ($\circ$), as well as direct simulation ($*$). The first marker corresponds to $\tau = 0.3$ followed by $\tau = 0.5, 1.0, \ldots, 3.0$. All results are based on: $k = 5$, $\gamma = 1$ and $w_1 = 10, w_2 = 1$. From top to bottom : $p_1 = 0.9, 0.5, 0.1, 0.01$ and $p_2 = 1 - p_1$.

The models proposed here can be extended in a number of different ways. One potential avenue for further research is the analysis of correlations between link weight and node degree. This direction has been explored in the context of classic compartmental

mean-field models based on node degree [55, 88]. Given that pairwise models extend to heterogeneous networks, such avenues can be further explored to include different types of correlations or other network-dependent weight distributions. While this is a viable direction, it is expected that the extra complexity will make the pairwise models more difficult to analyse and disagreement between pairwise and simulation models more likely. Another theoretically interesting and practically important aspect is the consideration of different types of time delays, representing latency or temporary immunity [13], and the analysis of their effects on the dynamics of epidemics on weighted networks. The methodology presented in this paper can be of wider relevance to phenomena that take place simultaneously on more that one type of network. Examples of such systems include the co-circulation of two different diseases in the same population [12], the spread of the same disease but via different routes [65], or the spread of epidemics concurrently with information about the disease [47, 64]. These areas offer other important avenues for further extensions.

Acknowledgements

P. Rattana acknowledges funding for her PhD studies from the Ministry of Science and Technology, Thailand. K.T.D Eames is funded by a Career Development Fellowship award from the National Institute for Health Research. I Z. Kiss acknowledges useful discussions with Professor Frank Ball on aspects of the epidemic threshold calculation. The authors wish to thank the two anonymous reviewers for their useful comments and suggestions which have contributed to improving the structure and clarify of the paper.

2.6 Appendices

2.6.1 Appendix A - Reducing the weighted pairwise models to the un-weighted equivalents

We start from the system

$$[\dot{S}] = \gamma[I] - \tau\textstyle\sum_{n=1}^{M} w_n[SI]_n,$$

$$[\dot{I}] = \tau\textstyle\sum_{n=1}^{M} w_n[SI]_n - \gamma[I],$$

$$[\dot{SI}]_m = \gamma([II]_m - [SI]_m) + \tau\textstyle\sum_{n=1}^{M} w_n([SSI]_{mn} - [ISI]_{nm}) - \tau w_m[SI]_m,$$

$$[\dot{II}]_m = -2\gamma[II]_m + 2\tau\textstyle\sum_{n=1}^{M} w_n[ISI]_{nm} + 2\tau w_m[SI]_m,$$

$$[\dot{SS}]_m = 2\gamma[SI]_m - 2\tau\textstyle\sum_{n=1}^{M} w_n[SSI]_{mn},$$

where $m = 1, 2, \ldots, M$. To close this system of equations at the level of pairs, we use the approximations

$$[ABC]_{mn} = \frac{k-1}{k}\frac{[AB]_m[BC]_n}{[B]}.$$

To reduce these equations to the standard pairwise model for un-weighted networks we use the fact that $\sum_{m=1}^{M}[AB]_m = [AB]$ for $A, B \in \{S, I\}$ and aim to derive the evolution equation for $[AB]$. Assuming that all weights are equal to some W, the following relations hold:

$$\begin{aligned}
[\dot{SI}] &= \sum_{m=1}^{M} [\dot{SI}]_m \\
&= \sum_{m=1}^{M} \left(\gamma([II]_m - [SI]_m) + \tau\sum_{n=1}^{M} w_n([SSI]_{mn} - [ISI]_{nm}) - \tau w_m[SI]_m \right) \\
&= \gamma([II] - [SI]) - \tau W[SI] + \tau W \sum_{m=1}^{M}\sum_{n=1}^{M} ([SSI]_{mn} - [ISI]_{nm}),
\end{aligned}$$

where the summations of the triples can be resolved as follows:

$$\sum_{m=1}^{M}\sum_{n=1}^{M}[SSI]_{mn} = \frac{k-1}{k}\sum_{m=1}^{M}[SS]_m\sum_{n=1}^{M}\frac{[SI]_n}{[S]}$$

$$= \frac{k-1}{k}\frac{[SS][SI]}{[S]} = [SSI].$$

Using the same argument for all other triples, the pairwise model for weighted networks with all weights being equal (without loss of generality $W = 1$) reduces to the classic pairwise model, that is

$$[\dot{S}] = \gamma[I] - \tau[SI],$$

$$[\dot{I}] = \tau[SI] - \gamma[I],$$

$$\sum_{m=1}^{M}[\dot{SI}] = [\dot{SI}] = \gamma([II] - [SI]) + \tau[SSI] - [ISI] - [SI],$$

$$\sum_{m=1}^{M}[\dot{II}] = [\dot{II}] = -2\gamma[II] + 2\tau([ISI] + [SI]),$$

$$\sum_{m=1}^{M}[\dot{SS}] = [\dot{SS}] = 2\gamma[SI] - 2\tau[SSI].$$

A similar argument holds for the pairwise model on weighted networks with SIR dynamics.

2.6.2 Appendix B - Proof of Theorem 1

We illustrate the main steps needed to complete the proof of Theorem 1. This revolves around starting from the inequality itself and showing via a series of algebraic manipulations that it is equivalent to a simpler inequality that holds trivially. Upon using that $p_1 k = k_1$, $p_2 k = k_2$ and $p_2 + p_1 = 1$, the original inequality can be rearranged to give

$$\sqrt{[(k_1-1)r_1 - (k_2-1)r_2]^2 + 4k_1k_2r_1r_2} \leq (k_1-1)r_1 + (k_2-1)r_2 + 2r_1p_2 + 2r_2p_1.$$

Based on the assumptions of the theorem, the right-hand side is positive, and thus this inequality is equivalent to the one where both the left- and right-hand sides are squared. Combined with the fact that $p_2 = 1 - p_1$, after a series of simplifications and factorizations this inequality can be recast as

$$4p_1(1-p_1)(r_1^2 + r_2^2) + 8kp_1(1-p_1)r_1r_2 \leq 4kp_1(1-p_1)(r_1^2 + r_2^2) + 8p_1(1-p_1)r_1r_2,$$

which can be further simplified to

$$4p_1(1-p_1)(r_1 - r_2)^2(k-1) \geq 0,$$

which holds trivially and thus completes the proof. We note that in the strictest mathematical sense the condition of the theorem should be $(k_1 - 1)r_1 + (k_2 - 1)r_2 + 2r_1p_2 + 2r_2p_1 \geq 0$. This holds if the current assumptions are observed since these are stronger but follow from a practical reasoning whereby for the network with fixed weight distribution, a node should have at least one link with every possible weight type.

2.6.3 Appendix C - Proof of Theorem 2

First, we show that R_0^1 is maximised when $w_1 = w_2 = W$. R_0^1 can be rewritten to give

$$R_0^1 \;=\; (k-1)\left(p_1\frac{\tau w_1}{\tau w_1+\gamma}+(1-p_1)\frac{\tau w_2}{\tau w_2+\gamma}\right).$$

Maximising this given the constraint $w_1p_1 + w_2(1 - p_1) = W$ can be achieved by considering R_0^1 as a function of the two weights and incorporating the constraint into it via the Lagrange multiplier method. Hence, we define a new function $f(w_1, w_2, \lambda)$ as follows:

$$f(w_1,w_2,\lambda) = (k-1)\left(p_1\frac{\tau w_1}{\tau w_1+\gamma}+(1-p_1)\frac{\tau w_2}{\tau w_2+\gamma}\right)$$

$$+\lambda(w_1p_1 + w_2(1 - p_1) - W).$$

Finding the extrema of this function leads to a system of three equations

$$\frac{\partial f}{\partial w_1} = \frac{(k-1)p_1\tau\gamma}{(\tau w_1+\gamma)^2} + \lambda p_1 = 0,$$

$$\frac{\partial f}{\partial w_2} = \frac{(k-1)(1-p_1)\tau\gamma}{(\tau w_2+\gamma)^2} + \lambda(1-p_1) = 0,$$

$$\frac{\partial f}{\partial \lambda} = w_1p_1 + w_2(1-p_1) - W = 0.$$

Expressing λ from the first two equations and equating these two expressions yields

$$\frac{(k-1)\tau\gamma}{(\tau w_1+\gamma)^2} = \frac{(k-1)\tau\gamma}{(\tau w_2+\gamma)^2}.$$

Therefore,

$$w_1 = w_2 = W,$$

and it is straightforward to confirm that this is a maximum.

Performing the same analysis for R_0^2 is possible but it is more tedious. Instead, we propose a more elegant argument to show that R_0^2 under the constraint of constant total link weight achieves its maximum when $w_1 = w_2 = W$. The argument starts by considering R_0^2 when $w_1 = w_2 = W$. In this case, and using that $r_2 = r_1 = r = \tau W/(\tau W + \gamma)$ we can write:

$$R_0^{2*} = \frac{(k_1-1)r_1 + (k_2-1)r_2 + \sqrt{[(k_1-1)r_1 - (k_2-1)r_2]^2 + 4k_1k_2r_1r_2}}{2}$$

$$= \frac{r(k_1+k_2-2) + \sqrt{r^2[(k_1-1)-(k_2-1)]^2 + 4r^2k_1k_2}}{2}$$

$$= \frac{r(k_1+k_2-2) + r\sqrt{(k_1+k_2)^2}}{2}$$

$$= \frac{r(2k_1+2k_2-2)}{2} = r(k_1+k_2-1) = (k-1)r.$$

However, it is known from Theorem 1 that $R_0^2 \leq R_0^1$, and we have previously shown that R_0^1 under the present constraint achieves its maximum when $w_1 = w_2 = W$, and its maximum is equal to $(k-1)r$. All the above can be written as

$$R_0^2 \leq R_0^1 \leq (k-1)r.$$

Now taking into consideration that $R_0^{2*} = (k-1)r$, the inequality above can be written as

$$R_0^2 \leq R_0^1 \leq (k-1)r = R_0^{2*},$$

and this concludes the proof.

2.6.4 Appendix D - The R_0-like threshold R

Let us start from the evolution equation for $[I](t)$,

$$\begin{aligned}[\dot{I}] &= \tau(w_1[SI]_1 + w_2[SI]_2) - \gamma[I] \\ &= \left[\tau w_1\left(\frac{[SI]_1}{[I]}\right) + \tau w_2\left(\frac{[SI]_2}{[I]}\right) - \gamma\right][I] \\ &= (\tau w_1\lambda_1 + \tau w_2\lambda_2 - \gamma)[I],\end{aligned}$$

where $\lambda_1 = \frac{[SI]_1}{[I]}$ and $\lambda_2 = \frac{[SI]_2}{[I]}$, and let R be defined as

$$R = \frac{\tau w_1\lambda_1 + \tau w_2\lambda_2}{\gamma}. \tag{2.14}$$

Following the method outlined by Keeling [58] and Eames [32], we calculate the early quasi-equilibrium values of $\lambda_{1,2}$ as follows:

$$\dot{\lambda_1} = 0 \Leftrightarrow [\dot{SI}]_1[I] = [\dot{I}][SI]_1,$$
$$\dot{\lambda_2} = 0 \Leftrightarrow [\dot{SI}]_2[I] = [\dot{I}][SI]_2.$$

Upon using the pairwise equations and the closure, consider $[\dot{SI}]_1[I] = [\dot{I}][SI]_1$:

$$\begin{aligned}[\dot{SI}]_1[I] &= (\tau w_1[SSI]_{11} + \tau w_2[SSI]_{12} - \tau w_1[ISI]_{11} \\ &\quad -\tau w_2[ISI]_{21} - \tau w_1[SI]_1 - \gamma[SI]_1)[I] \\ &= (\tau w_1[SI]_1 + \tau w_2[SI]_2 - \gamma[I])[SI]_1. \end{aligned} \tag{2.15}$$

Using the classical closure

$$[ABC]_{12} = \frac{k-1}{k}\frac{[AB]_1[BC]_2}{[B]},$$

$$[ABC]_{21} = \frac{k-1}{k}\frac{[AB]_2[BC]_1}{[B]},$$

and making the substitution : $[SI]_1 = \lambda_1[I]$, $[SI]_2 = \lambda_2[I]$, $[I] \ll 1$, $[S] \approx N$, $[SS]_1 \approx kNp_1$, $[SS]_2 \approx kN(1-p_1)$ together with $\gamma R = \tau w_1\lambda_1 + \tau w_2\lambda_2$, we have

$$(\tau w_1\lambda_1 + \tau w_2\lambda_2)kp_1 - (\tau w_1\lambda_1 + \tau w_2\lambda_2)p_1 - (\tau w_1\lambda_1 + \tau w_2\lambda_2)\lambda_1 - \tau w_1\lambda_1 = 0,$$

which can be solved for λ_1 to give

$$\lambda_1 = \frac{\gamma(k-1)p_1R}{\tau w_1 + \gamma R}.$$

Similarly, λ_2 can be found as

$$\lambda_2 = \frac{\gamma(k-1)(1-p_1)R}{\tau w_2 + \gamma R}.$$

Substituting the expressions for $\lambda_{1,2}$ into the original equation for R yields

$$R = \frac{A + B + \sqrt{(A+B)^2 + 4\tau^2 w_1 w_2 (k-2)}}{2\gamma},$$

where $A = \tau w_1[(k-1)p_1 - 1]$ and $B = \tau w_2[(k-1)p_2 - 1]$. If we define

$$R_1 = \frac{\tau w_1[(k-1)p_1 - 1]}{\gamma}, \quad \text{and} \quad R_2 = \frac{\tau w_2[(k-1)p_2 - 1]}{\gamma},$$

the expression simplifies to

$$R = \frac{R_1 + R_2 + \sqrt{(R_1 + R_2)^2 + 4R_1R_2Q}}{2},$$

where $Q = \dfrac{(k-2)}{[(k-1)p_1 - 1][(k-1)p_2 - 1]}$.

Substituting the modified closure

$$[ABC]_{11} = \frac{k_1 - 1}{k_1}\frac{[AB]_1[BC]_1}{[B]},$$

$$[ABC]_{12} = \frac{[AB]_1[BC]_2}{[B]},$$

$$[ABC]_{21} = \frac{[AB]_2[BC]_1}{[B]},$$

$$[ABC]_{22} = \frac{k_2 - 1}{k_2}\frac{[AB]_2[BC]_2}{[B]},$$

into Eq. (2.15) and making further substitution : $[SI]_1 = \lambda_1[I]$, $[SI]_2 = \lambda_2[I]$, $[I] \ll 1$, $[S] \approx N$, $[SS]_1 \approx k_1N$, $[SS]_2 \approx k_2N$, we have

$$(\tau w_1\lambda_1 + \tau w_2\lambda_2)k_1 - (\tau w_1\lambda_1 + \tau w_2\lambda_2)\lambda_1 - 2\tau w_1\lambda_1 = 0 \Longrightarrow \lambda_1 = \frac{\gamma k_1 R}{2\tau w_1 + \gamma R}.$$

Similarly, the equation $[\dot{S}I]_2[I] = [\dot{I}][SI]_2$ yields

$$\lambda_2 = \frac{\gamma k_2 R}{2\tau w_2 + \gamma R}.$$

Substituting these expressions for $\lambda_{1,2}$ into Eq. (2.14), we have

$$R = \frac{\tau(w_1k_1 + w_2k_2) - 2\tau(w_1 + w_2)}{2\gamma}$$

$$+\frac{\sqrt{[2\tau(w_1 + w_2) - \tau(w_1k_1 + w_2k_2)]^2 + 8\tau^2w_1w_2(k_1 + k_2 - 2)}}{2\gamma}.$$

If we define

$$R_1 = \frac{\tau w_1(k_1 - 2)}{\gamma}, \quad R_2 = \frac{\tau w_2(k_2 - 2)}{\gamma},$$

the above expression for R simplifies to

$$R = \frac{R_1 + R_2 + \sqrt{(R_1 + R_2)^2 + 4R_1R_2(Q-1)}}{2}$$

where $Q = \dfrac{k_1 k_2}{(k_1 - 2)(k_2 - 2)}$.

Chapter 3

Paper 2: Pairwise and Edge-based Models of Epidemic Dynamics on Correlated Weighted Networks

P. Rattana [1], J.C. Miller [2] and I.Z. Kiss [1]

[1] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK

[2] School of Mathematical Sciences, School of Biological Sciences, and the Monash Academy for Cross & Interdisciplinary Mathematics, Monash University, VIC 3800, Australia

3.1 Abstract

In this paper we explore the potential of the pairwise-type modelling approach to be extended to weighted networks where nodal degree and weights are not independent. As a baseline or null model for weighted networks, we consider undirected, heterogenous networks where edge weights are randomly distributed. We show that the pairwise model successfully captures the extra complexity of the network, but does this at the cost of limited analytical tractability due the high number of equations. To circumvent this problem, we employ the edge-based modelling approach to derive models corresponding to two different cases, namely for degree-dependent and randomly distributed weights. These models are more amenable to compute important epidemic descriptors, such as early growth rate and final epidemic size, and produce similarly excellent agreement with simulation. Using a branching process approach we compute the basic reproductive ratio for both models and discuss the implication of random and correlated weight distributions on this as well as on the time evolution and final outcome of epidemics. Finally, we illustrate that the two seemingly different modelling approaches, pairwise and edge-based, operate on similar assumptions and it is possible to formally link the two.

3.2 Introduction

The study of epidemic spread through contact networks has significantly improved our understanding of how the structure of interactions influences the spread of an infectious disease. One of the most recognised facts is that individuals with more contacts tend to become infected sooner and then spread the disease more quickly than others. Thus, for a given average degree, epidemics tend to spread faster if the population has a more heterogeneous degree distribution.

A number of models have been introduced to study the spread of an SIR (susceptible-infectious-recovered) infectious disease through a class of random networks known as configuration model networks [81]. The earliest models [86] were restricted to final size calculations, predicting how the total number infected depends on the transmission probability. More recently, models have been introduced which attempt to predict the dynamics of an epidemic, with varying levels of success and degrees of complexity. There are now several models available which can predict with high accuracy the population-scale dynamics of an SIR epidemic spreading through a configuration

model network [33, 58, 70, 72, 78].

However, these analyses assume that all interactions have the same strength. In fact some connections are expected to transmit infection quicker than others as a result of the closeness of interaction of the individuals. By itself, a heterogeneous distribution of contact weights would affect the dynamics of an epidemic. However, we further expect that an individual's contact-weights are likely to have some dependence on the degree of the nodes that the edges/links connect. Previous studies have considered and analysed different scenarios of weighted networks based on theoretical/synthetic network models [16, 25, 75, 95], as well as empirical networks reconstructed from real data (e.g. social mixing data [34] and cattle movements between farms [41]). These studies have typically focused on specific models that either gave information about (a) threshold quantities and final epidemic size, (b) mean-field type models for describing the time evolution of infection or (c) simulation. Here, we will aim to cover as many of these aspects as possible in one single body of work.

In this paper we develop and analyse models which allow us to incorporate edge-weights into the epidemic dynamics and we explore this via pairwise and edge-based compartmental models, as well as simulation. In particular, we focus on weighted networks where link or edge weights and node degree are not independent, see for example [55, 88]. The aim of this study is twofold. First, we explore the flexibility of the pairwise and edge-based compartmental modelling frameworks to account for this added level of complexity, and second, to gain better understanding on the precise impact of different weight distributions and of correlations between link-weight and degree on epidemic threshold, growth rate and epidemic dynamics. The paper is organised as follows. Section 2 is dedicated to model derivation starting with network construction and edge-weight distribution, including some null models, such as where link-weights are randomly distributed and where all link weights are equal to some predetermined average. In this same section, we derive and present the pairwise and edge-based models for random and degree-dependent weights cases. Section 3 is dedicated to results, and it is divided into analytic, numeric and model comparison parts. Finally, in section 4, we provide further aspects for discussion and future directions.

3.3 Model Derivation

The models are built in a bottom up approach. We first describe the construction of the networks we study and how their edge-weights are assigned. We then describe the disease dynamics and simulation model. We conclude this section with the formulation and derivation of the pairwise and edge-based compartmental models for two distinct classes of weighted networks.

3.3.1 Network construction and simulation

Our focus here is the construction of our model networks and the simulation of an epidemic through those networks. Our model networks use the configuration model framework [81] with each edge assigned one of M possible weights. The two network types we consider differ in how those weights are assigned to edges. We make standard assumptions about the disease spread, but we let the rate of transmission along an edge depend on its weight.

Networks with randomly-distributed edge weights

In this case a two step approach is used to generate networks with randomly-distributed edge weights. First, a network of N nodes with prescribed degree distribution $P(k)$ is generated according to the configuration model. This procedure leads to an undirected un-weighted network where edge weights can be now assigned at random according to a specified weight distribution $Q(w)$. If $Q(w)$ is defined across weights w_i, where $Q(w_i) = q_i$ and $i = 1, 2, \ldots, M$, then in a homogenous random network (i.e. all nodes have degree k), the distribution of edge-weights of various types is multi-nomial with parameters k - number of trials and q_i - the probability of a link being of weight w_i with $i = 1, 2, \ldots, M$. The average weight in the network is given by $\langle w \rangle_{random} = \sum_{i=1}^{M} q_i w_i$.

While this is a good baseline model it is unlikely that this scenario would be a true representation of social interactions. For example, different weights could be interpreted as representing different social interactions (e.g. household, workplace and casual) and this could suggest a model where each individual has a certain number of links of different weights. Ignoring degree heterogeneity and considering individuals to be equal can result in a weighted network with fixed edge-weights, e.g. each node has k links with k_1 being of household type and with $k_2 = k - k_1$ being of workplace type and thus of different weights, say w_1 and w_2 [95].

Degree-dependent weighted networks

While many different edge-weight allocation scenarios are possible, we opt to investigate the case where edge weights and node degrees are not independent. This is in contrast with the random edge weights case, where the network topology and the edge weight distribution and allocation are totally uncoupled. In particular, we wish to investigate an intuitively plausible idea which suggests that the weight or 'strength' of a link is negatively correlated to node-degree since individuals with many contacts are likely to afford a limited time commitment per link, and thus less of an opportunity for the disease to transmit [55, 88]. In line with these studies, we propose a weighted network model where the link-weight between two nodes of degree i and j, respectively, is given by $w(i,j) = w_{ij}$ with some functional form such that link-weight decreases as the degrees of nodes that it connects increase. Generating such a network is straightforward and it requires that first a configuration network with given degree distribution (i.e. $P(k)$) is created. This is followed by allocating weights to all links based on the degrees of the end nodes and according some pre-specified function $w(i,j)$, where $i,j = k_{min},\ldots,k_{max}$ with k_{min} and k_{max} being the minimal and maximal nodal degree in the network. In this case, the distribution of weights is such that $Q(W = w(i,j)) = NP(i)i\frac{jP(j)}{\langle k\rangle}/(\langle k\rangle N/2) = \frac{2ijP(i)P(j)}{\langle k\rangle^2}$ and $Q(W = w(i,i)) = NP(i)i\frac{iP(i)}{\langle k\rangle}/(\langle k\rangle N) = \frac{i^2P^2(i)}{\langle k\rangle^2}$, where W is the random variable corresponding to link-weights. Furthermore, discarding information about the degree of the nodes for a link and simply assigning a random variable W according to the distribution Q provides another way to allocate weights of different type. This setup makes it possible to construct at least two possible null-model-type weighted networks:

(i.) the first is a network that has the same topology and weight distribution but, with weights allocated at random (i.e. ignoring degree-weight correlations) as prescribed by the random variable W and its distribution Q, and

(ii.) the second is simply a weighted network where all link-weights are equal to the average weight computed as

$$\langle w\rangle_{dd} = \frac{\sum_{i=k_{min}}^{k_{max}} NP(i)i\frac{iP(i)}{\langle k\rangle}w_{ii} + 2\sum_{i=k_{min}}^{k_{max}-1}\sum_{j=i+1}^{k_{max}} NP(i)i\frac{jP(j)}{\langle k\rangle}w_{ij}}{\langle k\rangle N},$$

where $NP(i)i\frac{jP(j)}{\langle k\rangle} = \frac{ijNP(i)P(j)}{\langle k\rangle}$ stands for the actual expected number of links between nodes of degree i and j, and $\langle k\rangle = \sum kP(k)$ is the average nodal degree.

These two null models will be used as baseline models for comparison when looking to determine the effect of degree-dependent weights on epidemic dynamics and other

important indicators, such as R_0 and final epidemic size.

Epidemic model and simulation

In this paper, the simple SIR (susceptible-infective-recovered) epidemic model is considered. Disease transmission is specified in terms of infection and recovery events. The rate of transmission over an edge of weight 1 is denoted by τ and this is adjusted by the edge weight by assuming that transmission is directly proportional to it, i.e. rate of transmission across an edge of weight w is τw. Infected individuals recover independently of each other at rate γ. The simulation is implemented using the Gillespie algorithm [43] with exponentially distributed (rate given by the total rate of change in the system) inter-event times, with the single event to be implemented at each step being chosen at random and proportionally to its rate. All simulations start with a few infected nodes chosen at random with the remaining nodes being susceptible.

3.3.2 Approximate ODE models

Markovian processes on networks, being disease, rumour, information, innovation transmission or firing neurones result in an exact mathematical description in terms of Kolmogorov/master equations. Their high dimensionality, even for small networks, renders them difficult to use and often these can only be used to ascertain results of a theoretical nature but may offer less insight for specific applications. Notably, for highly symmetric or regular networks, the exact equations can be used directly and this is an area that has been well exploited and has been used to provide and illustrate linkages between stochastic and approximate ODE models. However, for more general networks, the drawback of the exact model remains. This has led to the development of a number of approaches and models that do an excellent job in approximating results from explicit simulations on networks which correspond to what would be regarded as the exact model. Examples include: (a) pairwise models [33, 58, 94, 105], (b) edge-based compartmental models and in general approaches that require the use of probability generating functions [78], (c) effective degree models [70, 72], and other variations or combinations based around these. In this paper, we will concentrate on pairwise and edge-based compartmental models and will assess their flexibility and performance in accounting and approximating epidemic dynamics unfolding on two main classes of weighted networks.

Pairwise models

The model extension that we propose is partly covered in Rattana et al. [95]. However, here we extend this from homogenous to heterogeneous networks with random weights as well as to the case where edge weights and node degree are not independent. Before writing down the two models, we refresh the notation and counting procedures. In line with the notation used for pairwise models, the number of singles remains unchanged, with $[A_k]$ denoting the number of nodes across the whole network which have degree k and are in state A. Pairs of type $A_k - B_{k'}$, $[A_k B_{k'}]$, are now further divided depending on edge weights, i.e. $[A_k B_{k'}]_i$ represents the number of links of type $A_k - B_{k'}$ with the edge having weight w_i, where as before $i = 1, 2, \ldots, M$ and $A, B \in \{S, I, R\}$. Edges are doubly counted (e.g. in both directions) and thus the following relations hold: $[A_k B_{k'}]_m = [B_{k'} A_k]_m$ and $[A_k A_k]_m$ is equal to twice the number of uniquely counted links of weight w_m with nodes at both ends in state A and having degree k. From this extension it follows that $\sum_{i=1}^{M} [A_k B_{k'}]_i = [A_k B_{k'}]$. The same convention holds at the level of triples where $[A_k B_{k'} C_q]_{mn}$ stands for the expected number of triples where a node in state B and of degree k' connects a node in state A and of degree k and a node in state C and of degree q via links of weight w_m and w_n, respectively. The weight of the edge impacts on the rate of transmission across that edge, and this is achieved by using a link-specific transmission rate equal to τw_i, where $i = 1, 2, \ldots, M$. In line with the above, we construct two pairwise models, one for randomly distributed weights across edges and one for the case where edge weights and node degrees are correlated.

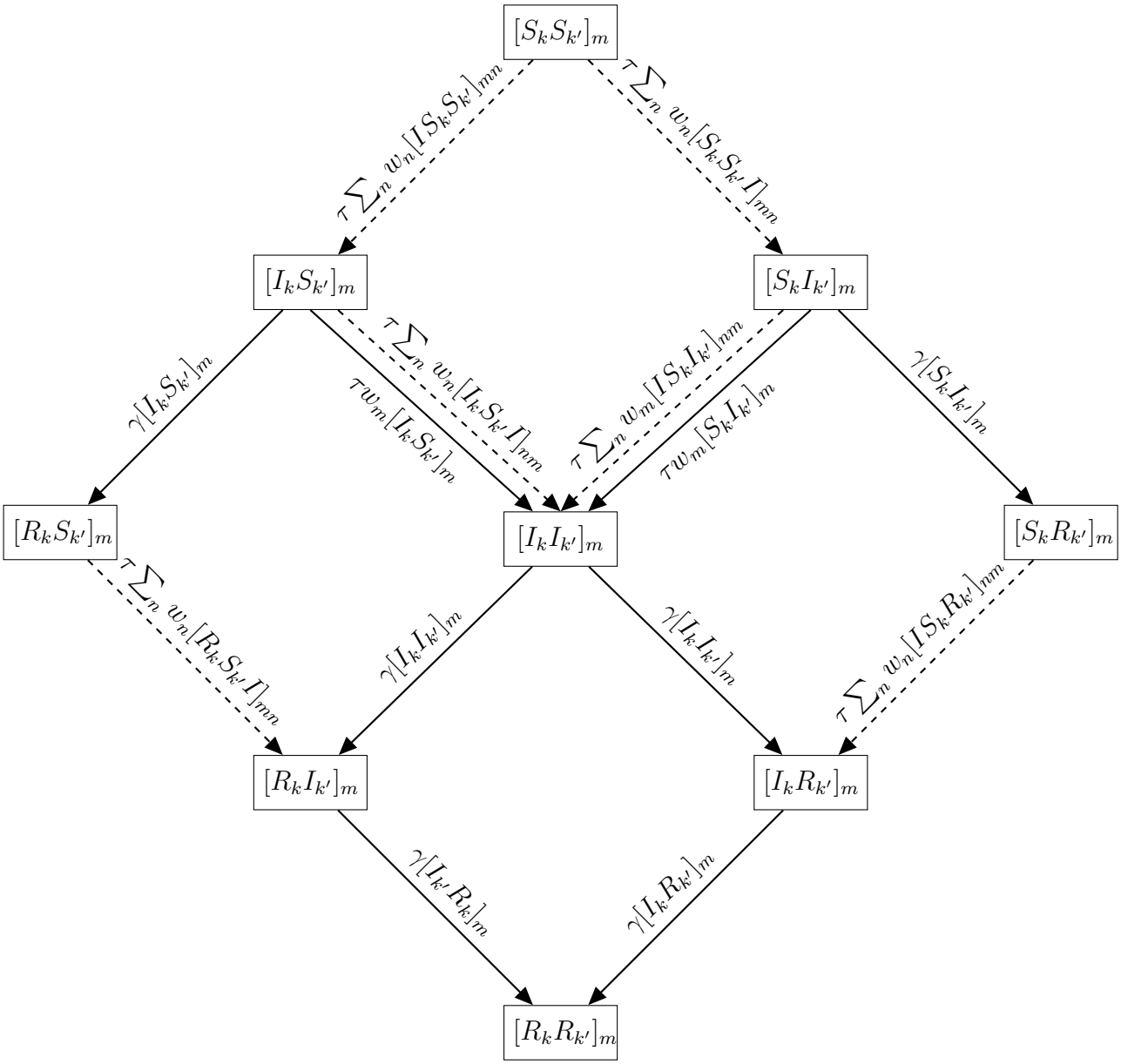


Figure 3.1: Flow diagram showing the evolution of pairs in the random weight case. The only pairs which have the potential to *eventually* transmit are the $[SS]$, $[SI]$ and $[IS]$ pairs, and hence, these need to be tracked. Solid and dashed arrows denote transmission within and from outside the pairs, respectively. We are able to find a closed system of equations which does not require calculating the other terms.

Evolution equations for SIR dynamics on heterogenous networks with random weights

$$
\begin{aligned}
[\dot{S_k}] &= -\tau \textstyle\sum_{n=1}^{M} w_n [S_k I]_n, \\
[\dot{I}] &= \tau \textstyle\sum_k \sum_{n=1}^{M} w_n [S_k I]_n - \gamma [I], \\
[\dot{S_k S_{k'}}]_m &= -\tau \textstyle\sum_{n=1}^{M} w_n \left([S_k S_{k'} I]_{mn} + [S_{k'} S_k I]_{mn} \right), \\
[\dot{S_k I}]_m &= \tau \left(\textstyle\sum_{k'} \sum_{n=1}^{M} w_n [S_k S_{k'} I]_{mn} - \sum_{n=1}^{M} w_n [I S_k I]_{nm} - w_m [S_k I]_m \right) - \gamma [S_k I]_m,
\end{aligned}
\tag{3.1}
$$

where $k, k' \in \{k_{min}, k_{min}+1, \ldots, k_{max}\}$ and $m = 1, 2, 3, ..., M$. Here, k_{min} and k_{max} stands for the smallest and largest nodal degree in the network. We further note that the system above stems from a reduction applied to a fuller version, see flow diagram in Fig. 3.1, where evolution equations for all $[I_k]$ classes are given (i.e. $[\dot{I_k}] = \tau \sum_{l=k_{min}}^{k_{max}} \sum_{n=1}^{M} w_n [S_k I_l]_n - \gamma [I_k]$). Summing this for $k = k_{min}, k_{min}+1, \ldots, k_{max}$ gives the evolution equations for $[I]$, as shown above. A similar notational procedure has been applied at the level of triples where in general $[A_k B_{k'} I]_{mn} = \sum_{q=k_{min}}^{k_{max}} [A_k B_{k'} I_q]_{mn}$.

The above system of Eq. (3.1) is not closed. Singles depend on pairs, and pairs depend on triples. Thus equations for triples are needed. This dependency on higher-order moments can be broken via approximating triples in terms of singles and pairs [58]. The agreement of the results from the closed system with simulation depends on how well the closure captures essential features of network structure and the edge weight distribution. Following Eames [33], the following closure is applied,

$$
[A_m B_n I] = \frac{n-1}{n} \frac{[A_m B_n][B_n I]}{[B_n]} \text{ or } [A_m B_n C_p] = \frac{n-1}{n} \frac{[A_m B_n][B_n C_p]}{[B_n]}. \tag{3.2}
$$

It is worth noting that the equations only rely on triples for which the central individual is susceptible. Thus individuals at the "ends" of a triple cannot affect one another's status through the central node until after they no longer affect the equations at the pair level.

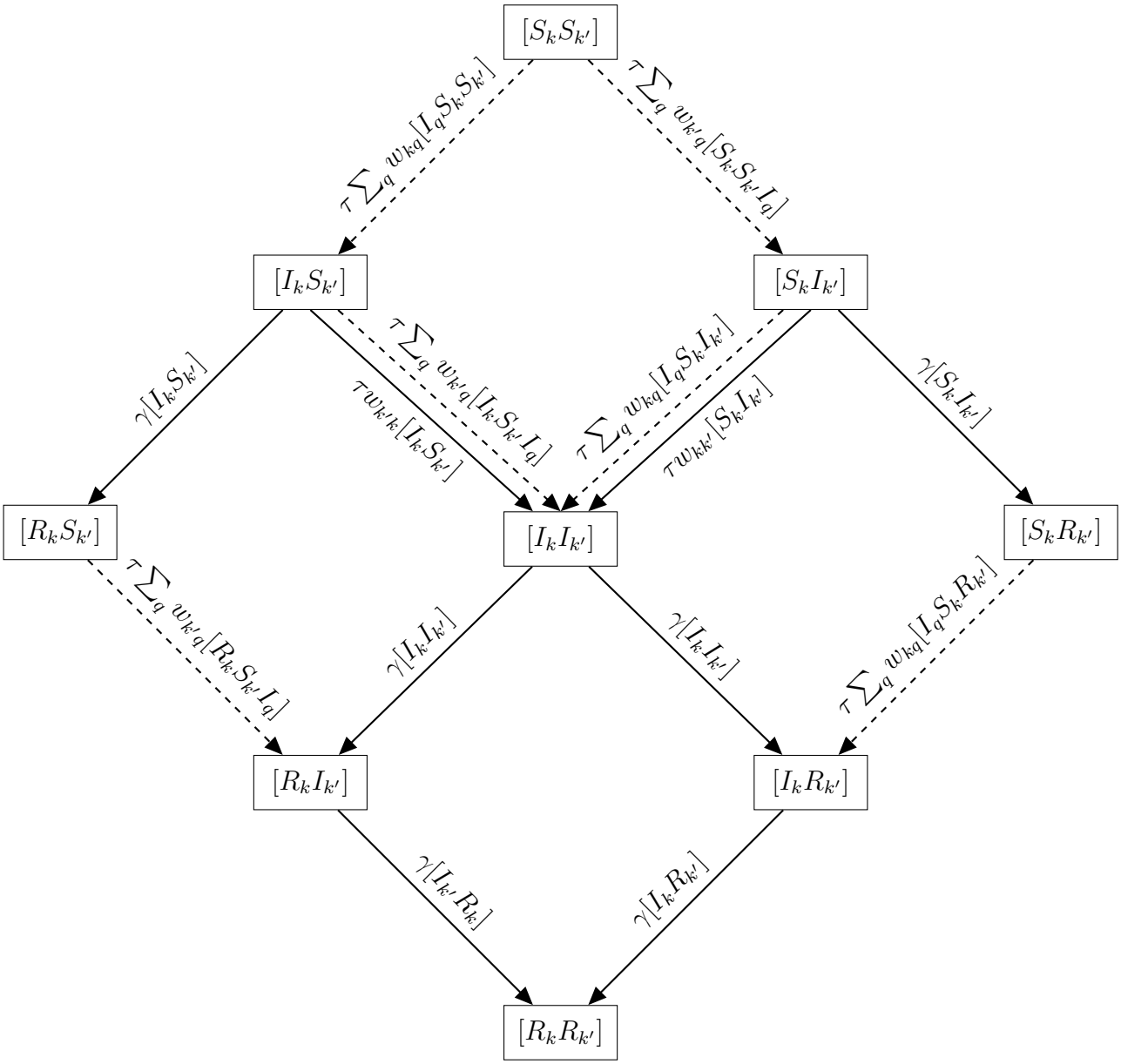


Figure 3.2: Flow diagram showing the evolution of pairs in the degree-dependent weight case. The only pairs which have the potential to *eventually* transmit are the $[SS]$, $[SI]$ and $[IS]$ pairs, and hence, these need to be tracked. Solid and dashed arrows denote transmission within and from outside the pairs, respectively. Again we are able to find a closed system of equations which only requires the $[SS]$, $[SI]$, and $[IS]$ terms.

Evolution equations for SIR dynamics on networks with degree-dependent weights

The focus now shifts to the case where we wish to incorporate some general correlation between edge weights and nodal degree. This is done by assuming that transmission between a susceptible node of degree k and an infected node of degree q happens at rate τw_{kq}, where $w_{kq} = w(k, q)$ can accommodate various dependencies of edge weight on nodal degree. The pairwise equations follow in the same way as before and are given by

$$\begin{aligned}
[\dot{S_k}] &= -\tau \textstyle\sum_q w_{kq}[S_k I_q], \\
[\dot{I_k}] &= \tau \textstyle\sum_q w_{kq}[S_k I_q] - \gamma[I_k], \\
[\dot{S_k S_{k'}}] &= -\tau \textstyle\sum_q (w_{k'q}[S_k S_{k'} I_q] + w_{kq}[S_{k'} S_k I_q]), \\
[\dot{S_k I_{k'}}] &= \tau \textstyle\sum_q (w_{k'q}[S_k S_{k'} I_q] - w_{kq}[I_{k'} S_k I_q]) - \tau w_{kk'}[S_k I_{k'}] - \gamma[S_k I_{k'}],
\end{aligned} \tag{3.3}$$

where as before $k, k', q \in \{k_{min}, k_{min}+1, \ldots, k_{max}\}$ and with w_{xy} yet unspecified. A corresponding flow diagram is given in Fig. 3.2. This system is closed in the same way as before using Eq. (3.2).

Edge-based compartmental models for weighted networks

We follow the derivation of Edge-based compartmental models (EBCM) of [76, 78, 79]. We assume that the population is connected according to the configuration model. We assume that the population-scale measures of infection (number infected, etc) are behaving deterministically. A consequence of this assumption is that if we choose a random individual u, the random event of whether u is or is not infected cannot have any impact on the population scale. So if we alter a single individual u so that u can become infected but cannot transmit to its partners, this can have no population-scale impact.

We define a *test individual* as follows: u is a test individual if u is randomly selected from the population and prevented from transmitting to its neighbours. Because the dynamics are deterministic and u is selected randomly, the probability u is in a given state equals the proportion of the population in that state. So to calculate the proportion infected, we can simply calculate the probability u is infected. This depends

on the probabilities that the partners of u are infected. Because we have prevented u from causing any infections, the status of each partner of u is independent of any other partner, which will simplify our calculations without altering the time of first infection of u. This is closely related to the observation for the pairwise equations that the triples only appear in the pair equations if the central individual is still susceptible.

EBCM Evolution equations for SIR dynamics on heterogeneous networks with random weights

As before, let us assume that there is a weight distribution $Q(w)$ assigned to the edges. We assume that the transmission rate for an edge with a given w is simply τw for some parameter τ. We further assume that; (a) infected individuals recover at rate γ, which is independent of how they were infected and that (b) at the initial time $t = t_0$, the probability an individual of degree k is susceptible is $S(k, t_0)$.

Let us now consider a test individual u, and let v be a random neighbour of u. Let θ be the probability that v has not transmitted to u given that at time t_0 v had not yet transmitted to u. Then trivially, $\theta = \sum_w Q(w)\theta_w$ where θ_w is the probability a neighbour along a weight w edge has not transmitted to u given that it had not yet transmitted at time t_0. Note that $\theta(t_0) = 1$. These probabilities are not affected by the degree of u, so the probability u is susceptible is

$$S(t) = \sum_k P(k)S(k, t_0)\theta^k = \psi(\theta).$$

Once we know $S(t)$, we can find the probability that u is infected or recovered simply by noting that $\dot{R} = \gamma I$ and $I = 1 - S - R$.

To complete the system, all the θ_w need to specified. Assuming that the edge connecting v to u has weight w, we define $\phi_{S,w}$ to be the probability that v is still susceptible. We define $\phi_{I,w}$ to be the probability v is infected but has not transmitted to u. We define $\phi_{R,w}$ to be the probability v has recovered and did not transmit to u. Then $\theta_w = \phi_{S,w} + \phi_{I,w} + \phi_{R,w}$ and $1 - \theta_w$ is the probability transmission has occurred (given that it had not occurred prior to t_0). Note however, that $\phi_{S,w}$ is independent of w because the weight of the edge from u to v does not influence the probability v has become infected. So we can treat $\phi_{S,w}$ as simply ϕ_S.

To find $\phi_S(t)$, we assume its initial value $\phi_S(t_0)$ is known. We need to find the probability that v has degree k given that it was chosen as a neighbour of u and was susceptible at time t_0. To do this, we count all edges belonging to susceptible nodes of

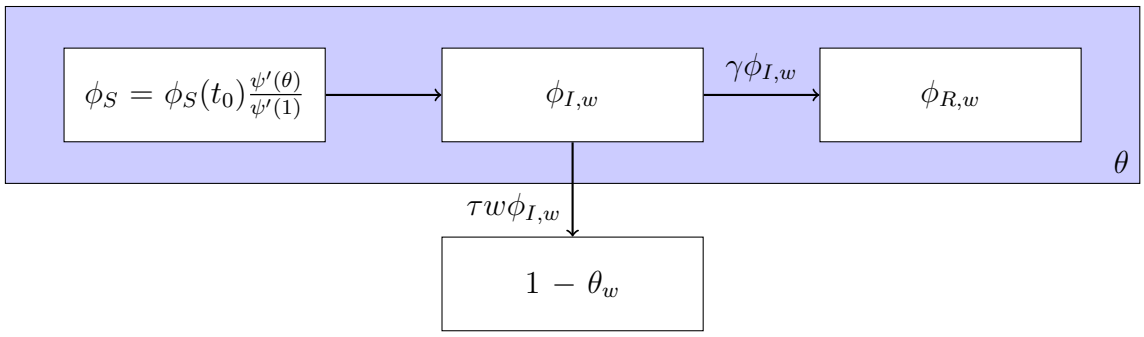


Figure 3.3: Flow diagram for EBCM model of random weight case. The large, shaded box contains all compartments, where transmission has not yet occurred.

degree k at time t_0 and divide by the number of all edges belonging to susceptible nodes at time t_0. This yields $kP(k)S(k,t_0)N/\sum_{k'} k'P(k')S(k',t_0)N = kP(k)S(k,t_0)/\psi'(1)$. The probability that v is still susceptible if it started susceptible and has degree k is θ^{k-1}. So $\phi_S(t) = \phi_S(t_0)\psi'(\theta)/\psi'(1)$. Note that this is independent of w.

We can find $\phi_{R,w}(t)$ in terms of θ_w. We assume that its initial value $\phi_{R,w}(t_0)$ is known. By definition, $\theta_w(t_0) = 1$. An infected neighbor along a weight-w edge transmits at rate τw and recovers at rate γ. Thus it moves from being counted towards $\phi_{I,w}$ to being counted towards $\phi_{R,w}$ at rate γ and to being counted towards $1-\theta_w$ at rate τw. Thus the rate of increase of $\phi_{R,w}$ is $\gamma/\tau w$ times the rate of increase of $1-\theta_w$. Using this argument, we conclude that

$$\phi_{R,w} = \frac{\gamma}{\tau w}(1-\theta_w) + \phi_{R,w}(t_0).$$

The arguments above are summarised in Fig. 3.3.

Then, since $\phi_S + \phi_{I,w} + \phi_{R,w} = \theta_w$ and we know ϕ_S and $\phi_{R,w}$, we can compute $\phi_{I,w}$. Summarising the findings above leads to

$$\begin{aligned}\dot{\theta}_w &= -\tau w\phi_{I,w} \\ &= -\tau w\left(\theta_w - \phi_S(0)\frac{\psi'(\theta)}{\psi'(1)} - \frac{\gamma(1-\theta_w)}{\tau w} - \phi_{R,w}(0)\right).\end{aligned}$$

So we end up with the system

$$\dot{\theta}_w = -\tau w\theta_w + \tau w\phi_S(0)\frac{\psi'(\theta)}{\psi'(1)} + \gamma(1-\theta_w) + \tau w\phi_{R,w}(t_0), \tag{3.4}$$

$$\theta = \sum_w Q(w)\theta_w, \tag{3.5}$$

where as for the pairwise model $w \in \{w_1, w_2, \ldots, w_M\}$. The initial conditions on $\phi_{S,w}(t_0)$ and $\phi_{R,w}(t_0)$ depend on how the epidemic is initialized. We have $\theta_w(t_0) = 1$. Noting that in $\psi'(\theta)$ it is θ, not θ_w, and combining the above with

$$S = \psi(\theta)\,, \qquad I = 1 - S - R\,, \qquad \dot{R} = \gamma I,$$

completes the system.

In general starting by randomly selecting a proportion ρ of individuals yields $S(k, t_0) = \phi_S(t_0) = 1 - \rho$ and $\phi_{R,w}(t_0) = R(t_0) = 0$. If instead the disease starts with a very small number and set t_0 when enough infections are present to be deterministic, then the initial conditions are different, and depend on the state of the population at this initial time [76]. In particular $S(k, t_0)$ may depend on k and not match exactly with $\phi_S(t_0)$.

EBCM evolution equations for SIR dynamics on networks with degree-dependent weights

The focus now shifts to the case when across each edge there is a weight $w_{kk'} = w(k, k')$ which depends on the degrees k and k' of the neighbouring nodes. Transmission happens at rate $\tau w_{kk'}$. We define θ_k to be the probability a neighbour of a degree k test node has not transmitted to it (given that it had not at time t_0). Due to this being k dependent, the expression for $\psi(\theta)$ will be more complicated compared to the random weights case. Instead, the probability the test node is susceptible is

$$S(t) = \sum_k P(k) S(k, t_0) \theta_k^k = \psi(\theta_{k_{min}}, \theta_{k_{min}+1}, \ldots, \theta_{k_{max}}).$$

Assume the neighbor v has degree k'. We define $\theta_{k,k'}$ to be the probability that v has not transmitted given that it has degree k', u has degree k, and v had not transmitted to u by time t_0. Then v is in the same states as before with probabilities $\phi_{S,k,k'}(t)$, $\phi_{I,k,k'}(t)$, and $\phi_{R,k,k'}(t)$. We find $\phi_{S,k,k'}(t) = \phi_{S,k,k'}(t_0)\theta_{k'}^{k'-1}$. We find that $\phi_{R,k,k'} = \gamma(1 - \theta_{k,k'})/\tau w_{kk'} + \phi_{R,k,k'}(t_0)$. The picture underlying this process of thought is given in Fig. 3.4.

The final equations are

$$\begin{aligned}
\dot{\theta}_{k,k'} &= -\tau w_{kk'} \theta_{k,k'} + \tau w_{kk'} \phi_{S,k,k'}(t_0) \theta_{k'}^{k'-1} + \gamma(1 - \theta_{k,k'}) + \tau w_{kk'} \phi_{R,k,k'}(t_0), && (3.6)\\
\theta_k &= \sum_{k'} P_n(k, k') \theta_{k,k'} && (3.7)\\
\dot{R} &= \gamma I\,, \qquad I = 1 - S - R\,, \qquad S = \sum_k P(k) S(k, t_0) \theta_k^k, && (3.8)
\end{aligned}$$

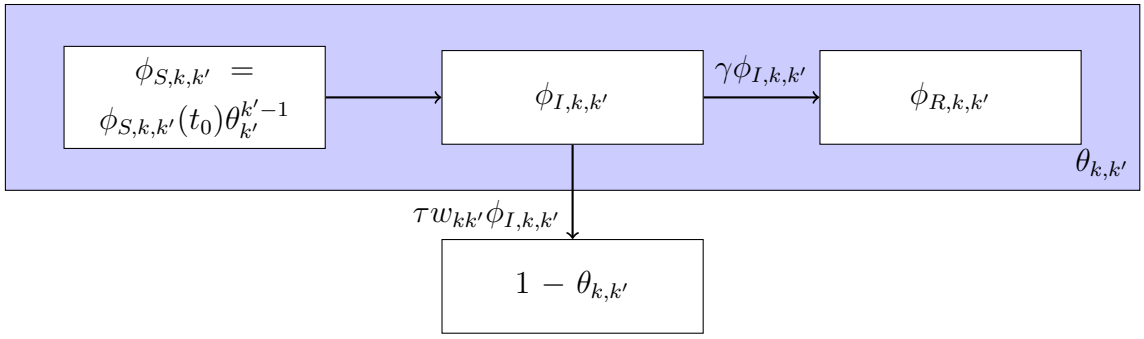


Figure 3.4: Flow diagram for the EBCM model with weights dependent on degree. The large, shaded box contains all compartments, where transmission has not yet occurred.

where $P_n(k, k')$ is the probability the neighbour of u has degree k' given that it hadn't transmitted to u by time t_0.

As before if we start by randomly selecting a proportion ρ of individuals at time t_0, we have $S(k, t_0) = \phi_{S,k,k'}(t_0) = 1 - \rho$, and $\phi_{R,k,k'}(t_0) = R(t_0) = 0$. In this case we get $P_n(k, k') = k'P(k')/\sum_{k''} k''P(k'')$. Hence, if the initial infected proportion is a randomly chosen proportion ρ, then the initial conditions are:

$$\begin{aligned} R(t_0) &= 0, \\ \phi_{R,k,k'}(t_0) &= 0, \\ S(k, t_0) &= 1 - \rho, \\ \phi_{S,k,k'}(t_0) &= 1 - \rho, \\ \theta_{k,k'}(t_0) &= 1, \end{aligned}$$

and

$$P_n(k, k') = \frac{k'P(k')}{\langle k \rangle}.$$

If the disease has been spreading for some time, the considerations above will not hold. In many cases, $P_n(k, k')$ can be calculated rather than taken as an 'initial condition'. If the infection has been spreading for some time before t_0, then the probability a neighbour has transmitted to u before t_0 depends on the degree of the neighbour. Since we define θ to be conditional on transmission to u never happening prior to t_0, this needs to be corrected for, and thus $P_n(k, k')$ will be different.

3.4 Results

In this section we present analytical and numerical results from network simulations, pairwise and edge-based representations of SIR dynamics. To compute the early growth rate and final epidemic size, we first write out the edge-based system for the special case of a heterogeneous network with low (degree l with probability $P(l)$) and high (degree h with probability $P(h)$) degree. This automatically induces three weights $w_1 = w_{ll}$, $w_2 = w_{lh} = w_{hl}$ and $w_3 = w_{hh}$. Moreover, for the degree-dependent weighted network, the distribution of weights is given by: $q_1 = q_{ll} = \frac{l^2P^2(l)}{\langle k\rangle^2}$, $q_2 = q_{lh} = q_{hl} = \frac{2lhP(l)P(h)}{\langle k\rangle^2}$ and $q_3 = q_{hh} = \frac{h^2P^2(h)}{\langle k\rangle^2}$, where $\langle k\rangle = lP(l) + hP(h)$ is the average nodal degree, and q_2 stands for the proportion of uniquely counted links between l and h nodes.

3.4.1 Epidemic threshold and final epidemic size

While pairwise models can be used to compute R_0 [58] and early growth rate [95], this is only practical for special cases where the number of equations remains relatively low. Such calculations are possible for homogenous un-weighted networks [58] and even for homogenous networks with two different edge weight types [32, 95]. In general and as we show, the edge-based compartmental models are more amenable to such analysis due to their smaller dimensionality, see Table 3.1.

Table 3.1: System complexity in terms of the number of differential equations needed to fully describe the epidemic dynamics. As before, M is the number of different weight types and K is the number of different nodal degrees, e.g. $K = k_{max} - k_{min} + 1$ provided that nodes of any degree between minimum and maximum degree exist.

Type of weighted network	Pairwise model	Edge-based model
randomly distributed weights	full system: $2K + \frac{K(K+1)}{2}M + K^2M$ reduced-system : $K + 1 + \frac{K(K+1)}{2}M + KM$	$M + 1$
degree-dependent weights	$2K + \frac{K(K+1)}{2} + K^2$	$K^2 + 1$

Random edge weight distribution for heterogeneous networks

The three weights system leads to working with $\theta_{w_1}, \theta_{w_2}$ and θ_{w_3}, where $Q(w_1) = q_1, Q(w_2) = q_2$ and $Q(w_3) = 1 - q_1 - q_2 = q_3$. Based on Eq. (3.4), the evolution

equations for these are,

$$\dot{\theta}_{w_1} = -\tau w_1\theta_{w_1} + (1-\rho)\tau w_1\frac{\psi'_{\theta_{w_1}}(\theta)}{\psi'_{\theta_{w_1}}(1)} + \gamma(1-\theta_{w_1}), \tag{3.9}$$

$$\dot{\theta}_{w_2} = -\tau w_2\theta_{w_2} + (1-\rho)\tau w_2\frac{\psi'_{\theta_{w_2}}(\theta)}{\psi'_{\theta_{w_2}}(1)} + \gamma(1-\theta_{w_2}), \tag{3.10}$$

$$\dot{\theta}_{w_3} = -\tau w_3\theta_{w_3} + (1-\rho)\tau w_3\frac{\psi'_{\theta_{w_3}}(\theta)}{\psi'_{\theta_{w_3}}(1)} + \gamma(1-\theta_{w_3}). \tag{3.11}$$

For a heterogenous network with N nodes where a node has degree l (e.g. low degree) with probability $P(l)$ or degree h (e.g. high degree) with probability $P(h) = 1 - P(l)$, the proportion of susceptibles at time t (based on Eq. (3.5)) is given by

$$S(t) = (1-\rho)(P(l)\theta^l + P(h)\theta^h) = \psi(\theta),$$

where $\theta = q_1\theta_{w_1} + q_2\theta_{w_2} + q_3\theta_{w_3}$.

Early growth rate

To compute the early growth rate, the assumption of an infinitesimally small initial infection must hold. Hence, to satisfy this requirement, we modify Eqs. (3.9-3.11) by taking $(1-\rho) \to 1$. This gives

$$\begin{aligned}
\dot{\theta}_{w_1} &= -\tau w_1\theta_{w_1} + \tau w_1\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right] + \gamma(1-\theta_{w_1}),\\
\dot{\theta}_{w_2} &= -\tau w_2\theta_{w_2} + \tau w_2\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right] + \gamma(1-\theta_{w_2}),\\
\dot{\theta}_{w_3} &= -\tau w_3\theta_{w_3} + \tau w_3\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right] + \gamma(1-\theta_{w_3}),
\end{aligned}$$

where $P_e(l) = lP(l)/\langle k\rangle$, $P_e(h) = hP(h)/\langle k\rangle$ and $\langle k\rangle = lP(l) + hP(h)$. Here, $P_e(k)$ represents the probability of finding a node of degree k when picking an edge at random and considering either of the nodes at its ends. We set $\theta_{w_1} = 1+\varepsilon_1, \theta_{w_2} = 1+\varepsilon_2$ and $\theta_{w_3} = 1+\varepsilon_3$. We linearise about the equilibrium and have the matrix equation

$$\begin{pmatrix} \dot{\varepsilon_1} \\ \dot{\varepsilon_2} \\ \dot{\varepsilon_3} \end{pmatrix} = \begin{pmatrix} -\tau w_1 + \tau w_1 q_1\zeta - \gamma & \tau w_1 q_2\zeta & \tau w_1 q_3\zeta \\ \tau w_2 q_1\zeta & -\tau w_2 + \tau w_2 q_2\zeta - \gamma & \tau w_2 q_3\zeta \\ \tau w_3 q_1\zeta & \tau w_3 q_2\zeta & -\tau w_3 + \tau w_3 q_3\zeta - \gamma \end{pmatrix} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix},$$

where

$$\zeta = (l-1)P_e(l) + (h-1)P_e(h).$$

Thus, the eigenvalues are the solutions of a 3rd order equation given by $\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0$, where

$$a_1 = u_1 + u_2 + u_3 - v_1 - v_2 - v_3,$$

$$\begin{aligned} a_2 &= u_1u_2 + u_1u_3 + u_2u_3 - u_1(v_2+v_3) - u_2(v_1+v_3) - u_3(v_1+v_2), \\ a_3 &= u_1u_2u_3 - u_1u_2v_3 - u_1u_3v_2 - u_2u_3v_1, \end{aligned}$$

where, u_i-s and v_i-s are given by

$$u_i = \tau w_i + \gamma\,, \quad v_i = \tau w_i q_i \zeta \quad \text{for} \quad i = 1, 2, 3.$$

By considering the case of $\lambda = 0$, the critical point for change of stability, the third order equation yields $a_3 = 0$. This means that at the point at which the eigenvalue changes sign $a_3 = 0$, and this gives a relation between the system parameters which determines the threshold condition.

The basic reproduction number - R_0

The basic reproduction number R_0 can be computed in two different ways. First, by using an individual-level view and average across nodes of different degrees that have become infected from the very initial index case. By doing this, we average the expected number of infections in the second generation. This approach yields,

$$R_0^{rw} = (l-1)P_e(l)(q_1r_1 + q_2r_2 + q_3r_3) + (h-1)P_e(h)(q_1r_1 + q_2r_2 + q_3r_3),$$

where

$$r_i = \frac{\tau w_i}{\tau w_i + \gamma} \quad \text{for} \quad i = 1, 2, 3.$$

A more rigorous and widely applicable approach is to compute R_0 as the leading eigenvalue of the next generation matrix (NGM). In this case, we can consider the epidemic in terms of an embedded multi-type branching process [5, 4], where the $NGM = (m_{ij})_{i,j=1,2,\ldots,N_t}$ (N_t - number of different types) consists of entries giving the expected number of offsprings of type i produced by a single individual of type i. Once, the different types have been defined, then NGM can be constructed, and R_0 will be equivalent to the leading eigenvalue of the NGM. In this case, we have individuals of two different types (individuals of low and high degree) and the NGM is given by,

$$NGM = \begin{bmatrix} (l-1)P_e(l)(q_1r_1 + q_2r_2 + q_3r_3) & (h-1)P_e(l)(q_1r_1 + q_2r_2 + q_3r_3) \\ \\ (l-1)P_e(h)(q_1r_1 + q_2r_2 + q_3r_3) & (h-1)P_e(h)(q_1r_1 + q_2r_2 + q_3r_3) \end{bmatrix},$$

where, for example $(h-1)P_e(l)(q_1r_1 + q_2r_2 + q_3r_3)$ stands for the expected number of individuals of low degree infected by a typical infected individual with high degree.

Hence,

$$R_0^{rw} = \Big((l-1)P_e(l) + (h-1)P_e(h)\Big)(q_1 r_1 + q_2 r_2 + q_3 r_3), \tag{3.12}$$

and this is identical to the previously computed value. A further consistency check of our calculations can be performed. Namely, the relation $R_0^{rw} = 1 \Leftrightarrow \lambda = 0$ should hold. Indeed, using condition $a_3 = 0$ leads to $R_0^{rw} = 1$.

Final epidemic size

To compute the final epidemic size, we need to return to the original equations that account for the initial conditions as given by Eqs. (3.9-3.11). By setting the derivatives to zero, it is possible to find asymptotic values of θ_{w_1}, θ_{w_2} and θ_{w_3}, i.e. $\theta_{w_1}(\infty)$, $\theta_{w_2}(\infty)$ and $\theta_{w_3}(\infty)$. Once these values are know the final epidemic size is given by $R(\infty) = 1 - \psi(\theta_{w_1}(\infty), \theta_{w_2}(\infty), \theta_{w_3}(\infty))$, where $\theta_{w_1}(\infty), \theta_{w_2}(\infty)$ and $\theta_{w_3}(\infty)$ are the solutions of the following system,

$$\theta_{w_1}(\infty) = \frac{\gamma + (1-\rho)\tau w_1 \left[P_e(l)\left(\theta(\infty)\right)^{l-1} + P_e(h)\left(\theta(\infty)\right)^{h-1}\right]}{\tau w_1 + \gamma}, \tag{3.13}$$

$$\theta_{w_2}(\infty) = \frac{\gamma + (1-\rho)\tau w_2 \left[P_e(l)\left(\theta(\infty)\right)^{l-1} + P_e(h)\left(\theta(\infty)\right)^{h-1}\right]}{\tau w_2 + \gamma}, \tag{3.14}$$

$$\theta_{w_3}(\infty) = \frac{\gamma + (1-\rho)\tau w_3 \left[P_e(l)\left(\theta(\infty)\right)^{l-1} + P_e(h)\left(\theta(\infty)\right)^{h-1}\right]}{\tau w_3 + \gamma}, \tag{3.15}$$

where $\theta(\infty) = q_1\theta_{w_1}(\infty) + q_2\theta_{w_2}(\infty) + q_3\theta_{w_3}(\infty)$. By treating the above as a fixed point problem, it can be shown that a numerical recursion will converge quickly to the true solution and we compare these simulation results in the numerical analysis part.

Degree-dependent weights

For the same simplified scenario with a network with bimodal degree distribution and weights that correlate with node-degree, Eqs. (3.6-3.8) yield

$$\dot{\theta}_{ll} = -\tau w_{ll}\theta_{ll} + (1-\rho)\tau w_{ll}\theta_l^{l-1} + \gamma(1-\theta_{ll}), \tag{3.16}$$

$$\dot{\theta}_{lh} = -\tau w_{lh}\theta_{lh} + (1-\rho)\tau w_{lh}\theta_h^{h-1} + \gamma(1-\theta_{lh}), \tag{3.17}$$

$$\dot{\theta}_{hl} = -\tau w_{hl}\theta_{hl} + (1-\rho)\tau w_{hl}\theta_l^{l-1} + \gamma(1-\theta_{hl}), \tag{3.18}$$

$$\dot{\theta}_{hh} = -\tau w_{hh}\theta_{hh} + (1-\rho)\tau w_{hh}\theta_h^{h-1} + \gamma(1-\theta_{hh}). \tag{3.19}$$

According to the model derivation θ_l and θ_h can be found as

$$\begin{aligned}\theta_l &= P_n(l,l)\theta_{ll} + P_n(l,h)\theta_{lh},\\ \theta_h &= P_n(h,l)\theta_{hl} + P_n(h,h)\theta_{hh},\end{aligned}$$

with $P_n(k,k') = k'P(k')/\langle k\rangle$. This complemented by

$$S(t) = (1-\rho)(P(l)\theta_l^l + P(h)\theta_h^h),$$

gives the full system.

Early growth rate

As before, we note that for the correct calculation of the early growth rate, Eqs. (3.16-3.19) must be used with $(1-\rho) \to 1$. By setting $\theta_{ll} = 1+\varepsilon_1$, $\theta_{lh} = 1+\varepsilon_2$, $\theta_{hl} = 1+\varepsilon_3$ and $\theta_{hh} = 1+\varepsilon_4$, and linearising around the disease-free steady state leads to the following Jacobian,

$$J = \begin{bmatrix} -\tau w_1 + v_1 - \gamma & \tau w_1(l-1)P_n(l,h) & 0 & 0 \\ 0 & -\tau w_2 - \gamma & \tau w_2(h-1)P_n(h,l) & \tau w_2(h-1)P_n(h,h) \\ \tau w_2(l-1)P_n(l,l) & \tau w_2(l-1)P_n(l,h) & -\tau w_2 - \gamma & 0 \\ 0 & 0 & \tau w_3(h-1)P_n(h,l) & -\tau w_3 + v_2 - \gamma \end{bmatrix},$$

where

$$v_1 = \tau w_1(l-1)P_n(l,l)\,, \qquad v_2 = \tau w_3(h-1)P_n(h,h).$$

The eigenvalues will be the solution of $\det(J-\lambda I) = 0$, where I is the identity matrix. Thus, the eigenvalues are the solutions of a 4th order equation given by $\lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4 = 0$, where

$$\begin{aligned}a_1 &= u_1(1-R_1) + 2u_2 + u_3(1-R_2),\\ a_2 &= 2u_2\Big(u_1(1-R_1) + u_3(1-R_2)\Big) + u_2^2 + u_1u_3(1-R_1)(1-R_2) - v_3,\\ a_3 &= 2u_1u_2u_3(1-R_1)(1-R_2) + u_2^2\left(u_1(1-R_1) + u_3(1-R_2)\right) - v_3\left(u_1 + u_3(1-R_2)\right) - v_2v_3,\\ a_4 &= u_1u_2^2u_3(1-R_1)(1-R_2) - u_1u_3v_3(1-R_2) - u_1v_2v_3,\end{aligned}$$

where

$$R_1 = (l-1)P_n(l,l)r_1,\;\; R_2 = (h-1)P_n(h,h)r_3,\;\; v_3 = (\tau w_2)^2(l-1)P_n(h,l)(h-1)P_n(l,h),$$

and where u_i-s are given by

$$u_i = \tau w_i + \gamma \quad \text{for} \quad i = 1, 2, 3.$$

By considering the case of $\lambda = 0$, the critical point for change of stability, the fourth order equation yields $a_4 = 0$. This means that at the point at which the eigenvalue changes sign $a_4 = 0$, and this gives a threshold condition. As expected, it can be shown that $a_4 = 0$ is equivalent to $R_0^{dd} = 1$ (below). This confirms that the calculations are consistent.

The basic reproduction number - R_0

In this case, we calculate R_0 only by using the next generation matrix approach, and R_0 is the leading eigenvalue of the next generation matrix. Before writing down the NGM we need to specify the choice of individual types, and then the entries of the $NGM = (m_{ij})_{i,j=1,2,\ldots,N_t}$. For this case, the types will be depend solely on the degree of the nodes, and thus, the NGM is given by,

$$NGM = \begin{bmatrix} (l-1)P_e(l)r_1 & (h-1)P_e(l)r_2 \\ \\ (l-1)P_e(h)r_2 & (h-1)P_e(h)r_3 \end{bmatrix}.$$

For example, the expected number of low degree individuals produced by a single high degree individual h, is given by $(h-1)P_e(l)r_2$. The leading eigenvalue of the above matrix, and thus R_0 is given by

$$R_0^{dd} = \frac{R_1 + R_2 + \sqrt{(R_1 - R_2)^2 + 4F}}{2},$$

where

$$R_1 = (l-1)P_e(l)r_1, \qquad R_2 = (h-1)P_e(h)r_3,$$

and

$$F = (l-1)P_e(l)(h-1)P_e(h)r_2^2.$$

Final epidemic size

Using the same approach as before and taking into account the initial condition in terms of ρ, the final epidemic size is given by $R(\infty) = 1 - \psi(\theta_l(\infty), \theta_h(\infty))$ where $\theta_{ll}(\infty)$, $\theta_{lh}(\infty)$, $\theta_{hl}(\infty)$ and $\theta_{hh}(\infty)$ are the solutions of the following system,

$$\theta_{ll}(\infty) \;=\; \frac{\gamma + (1-\rho)\tau w_{ll}\theta_l^{l-1}(\infty)}{\tau w_{ll} + \gamma}, \tag{3.20}$$

$$\theta_{lh}(\infty) = \frac{\gamma + (1-\rho)\tau w_{lh}\theta_h^{h-1}(\infty)}{\tau w_{lh} + \gamma}, \tag{3.21}$$

$$\theta_{hl}(\infty) = \frac{\gamma + (1-\rho)\tau w_{hl}\theta_l^{l-1}(\infty)}{\tau w_{hl} + \gamma}, \tag{3.22}$$

$$\theta_{hh}(\infty) = \frac{\gamma + (1-\rho)\tau w_{hh}\theta_h^{h-1}(\infty)}{\tau w_{hh} + \gamma}. \tag{3.23}$$

Comparison of R_0 and final epidemic size

Based on the analytic and semi-analytic calculations above, we provide a few examples where R_0 and the final epidemic size (Fig. 3.5) are compared for networks with heterogenous degree and weight distributions. Namely, as indicated in section 3.3.1, we start from networks with degree-dependent weights and compare R_0 and final epidemic size corresponding to this against those from networks with the same topology and same weight distribution but with weights assigned at random, and weighted networks where all weights are equal to the average weight from the original network, $\langle w\rangle_{dd} = q_1 w_1 + q_2 w_2 + q_3 w_3$. Fig. 3.5 (top panel) shows clearly that R_0 is maximised when all weights are equal, and that networks with randomly distributed weights allow for a larger R_0 value compared to the case of networks where degrees and weights are inversely correlated. This observation can be made rigorous. We start by noting that R_0 for the case of equal weights, based on Eq. (3.12), is given by,

$$R_0^{av} = ((l-1)P_e(l) + (h-1)P_e(h))\frac{\tau\langle w\rangle_{dd}}{\tau\langle w\rangle_{dd} + \gamma}. \tag{3.24}$$

Similarly, based on Eq. (3.12), the basic reproduction ratio is given by

$$R_0^{rw} = ((l-1)P_e(l) + (h-1)P_e(h))\left(q_1\frac{\tau w_1}{\tau w_1 + \gamma} + q_2\frac{\tau w_2}{\tau w_2 + \gamma} + q_3\frac{\tau w_3}{\tau w_3 + \gamma}\right).$$

First, we want to show that $R_0^{rw} \leq R_0^{av}$. Noting that $\varphi(w) = \frac{\tau w}{\tau w + \gamma}$ is a concave function on $w \in [0, \infty)$, as $\varphi'' < 0$, then using Jensen's inequality under the condition $q_1 + q_2 + q_3 = 1$, yields

$$\begin{aligned} q_1\varphi(w_1) + q_2\varphi(w_2) + q_3\varphi(w_3) &\leq \varphi(q_1 w_1 + q_2 w_2 + q_3 w_3), \\ q_1\frac{\tau w_1}{\tau w_1 + \gamma} + q_2\frac{\tau w_2}{\tau w_2 + \gamma} + q_3\frac{\tau w_3}{\tau w_3 + \gamma} &\leq \frac{\tau(q_1 w_1 + q_2 w_2 + q_3 w_3)}{\tau(q_1 w_1 + q_2 w_2 + q_3 w_3) + \gamma}. \end{aligned}$$

Hence, we can conclude that $R_0^{rw} \leq R_0^{av}$, with equality when all weights are equal. Moreover, it is easy to see that when $w_1 = w_2 = w_3 = w$, we have

$$R_0^{rw} = R_0^{av} = ((l-1)P_e(l) + (h-1)P_e(h))\frac{\tau w}{\tau w + \gamma}.$$

In Appendix A, we also provide a rigorous proof for the observation that $R_0^{dd} \leq R_0^{rw}$. Hence the following inequality holds

$$R_0^{dd} \leq R_0^{rw} \leq R_0^{av}. \tag{3.25}$$

We note that while the proof of $R_0^{rw} \leq R_0^{av}$ does not rely on the negative correlation between degree and weight, the proof of the second inequality, $R_0^{dd} \leq R_0^{rw}$, makes use of this information. The final epidemic size can be computed semi-analytically using the approach developed in the context of edge-based modelling. Namely, we use Eqs. (3.13 - 3.15) for the randomly-distributed and fixed weights case, and Eqs. (3.20 - 3.23) for the degree-dependent weighted network case. In both situations, we treat the equations as maps which we then numerically iterate to find their fixed points. The final epidemic size plots (see the bottom panel of Fig. 3.5) show that for the same R_0 value, the final epidemic size is largest on the original network with degree-dependent weights. This is a direct consequence of the relation between the R_0 values on the different networks, see Eq. (3.25). Namely, with all parameters being equal, R_0 is smallest on the original network. Hence, considering a fixed value of $R_0 (= R_0^{const})$ across the different networks requires a larger value of τ on the original network compared to the randomly distributed and fixed weights cases. This higher value is required to compensate for the negative correlation between degree and weights, which means that τ has to be disproportionately large to compensate for the smallest possible weights between highly connected nodes. This increase in τ has an automatic knock on effect of also improving transmission between poorly connected nodes with an overall increase in final epidemic size. It is worth noting the complete reversion of order between the top and bottom panel of Fig. 3.5. The same figure shows that the random and uniform average weight cases lead to an identical functional relation between final epidemic size and R_0. In Appendix B, we provide a simple, formal proof for this observation.

Final epidemic size comparisons

To explore the potential of the various models to capture the final epidemic size, we compare outputs from the semi-analytic approach with long-time results from simulations and the long-time solution of the pairwise model. To stress test the robustness of the model, we use two additional weight functions, namely $w_{ij} = 1/(i+j)^{1/2}$ and $w_{ij} = 1/\ln(i+j)$. Numerical results presented in Fig. 3.6 exhibit excellent agreement across all models and for the three different weight functions. As opposed to Fig. 3.5,

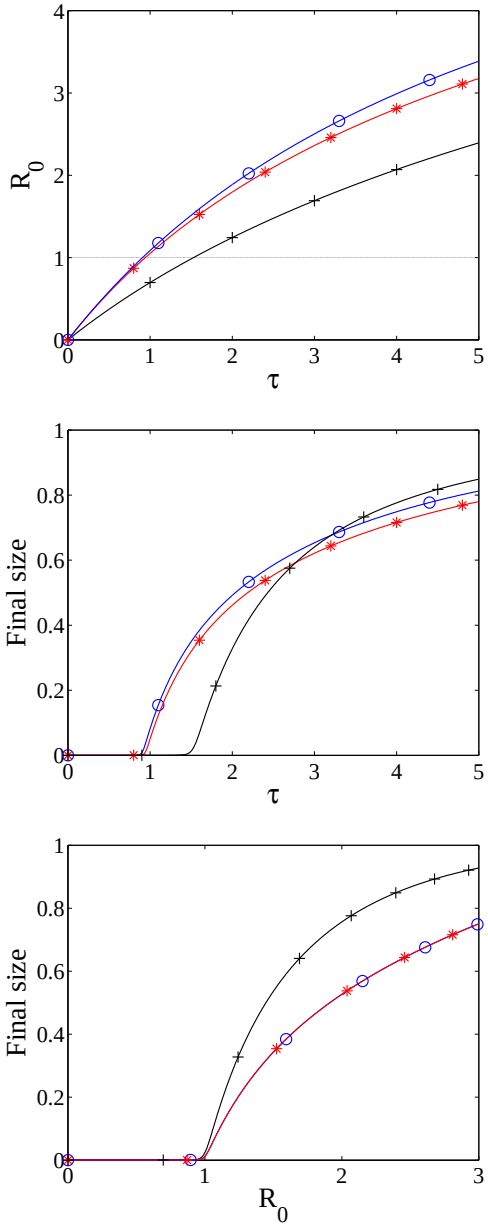


Figure 3.5: Basic reproductive ratio R_0 and final epidemic size for heterogeneous weighted networks. The parameters values are $\rho = 0.0001$, $P(l) = 0.8$, $P(h) = 1 - P(l)$, $l = 3$, $h = 13$ and $\gamma = 1$. Degree-dependent weighted networks (black line and $(+)$), networks with random weight distribution (red line and $(\star)$), and networks with all weights equal (blue line and $(\circ)$). All networks have the same average weight $\langle w \rangle_{dd} = q_1 w_{ll} + q_2 w_{lh} + q_3 w_{hh}$, where the weight function is $w_{ij} = 1/(i \times j)^{1/2}$.

here we use a higher number of initially infected nodes ($I_0 = 50$ out of $N = 1000$) to avoid early stochastic extinction in simulations. The plots in Fig. 3.6 show a similar trend with that observed in Fig. 3.5 (see middle panel).

A notable feature is the changeover in the size of the final epidemic size from being larger on networks with randomly distributed weights (for smaller values of τ) to the epidemic affecting a higher fraction of the population on networks with degree-dependent weights (for larger τ values). Intuitively this can be explained as follows. For degree-dependent weights, the transmissibility amongst, from or to highly connected nodes is penalised by small edge weights, with the smallest weights on high-to-high nodes connections. However, nodes that are less well connected can receive and transmit the infection more readily. We now discuss separately the cases of small and large τ:

1. For small values of τ, the random redistribution of weights will lead to links between, from or to highly connected nodes to be more likely to transmit, and this will lead to a larger final epidemic size. Transmission between poorly connected nodes will suffer but, infection involving highly connected nodes dominates for small values of τ.

2. As the value of τ increases the effect of small weights is less significant (i.e. transmission rate is the product of weight and the value of τ). Thus, disease spreads more readily across the whole network. However, redistributing links at random will improve an already appropriate transmission between highly-connected nodes (i.e. edge weights will always be greater or equal than for the degree-dependent weight case) but, at the expense of seeing smaller weights between less well connected nodes that are more abundant in the network.

The arguments above are confirmed by numerical simulations (not shown here), whereby the number of poorly connected, susceptible nodes at large times is greater in the case of random weights. All the effects above become less marked for the two additional weight functions. This is due to the two additional weight functions giving rise to higher edge weights, and thus a more efficient transmission with the epidemic affecting a large proportion of the network.

3.4.2 Numerical analysis of pairwise- and edge-based models

The numerical analysis part focuses around comparisons between the 'original' degree-dependent weighted networks and the two null models. Namely, we consider the network

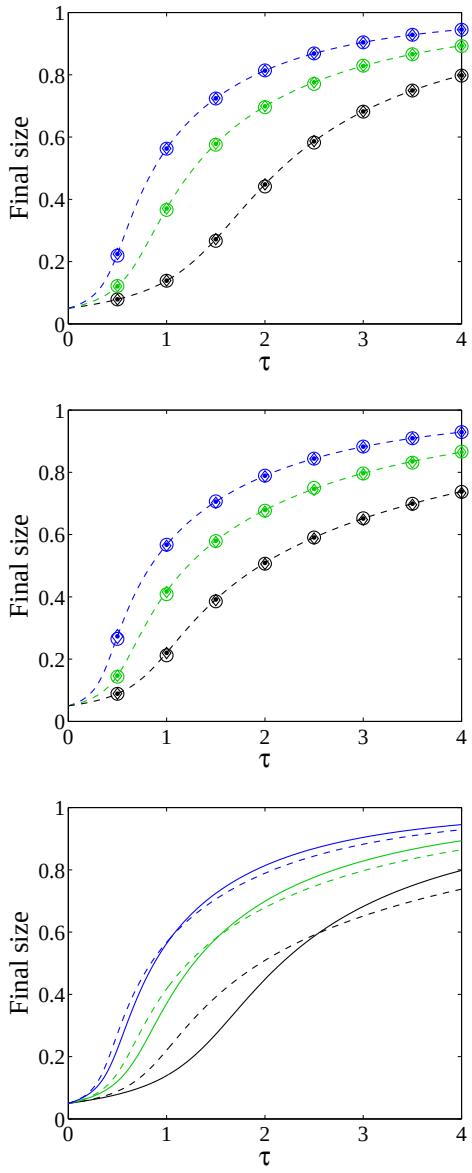


Figure 3.6: Final epidemic size for heterogeneous weighted networks with different weight functions: $w_{ij} = 1/\ln(i+j)$ (blue), $w_{ij} = 1/(i+j)^{1/2}$ (green) and $w_{ij} = 1/(i \times j)^{1/2}$ (black) (or top to bottom in each figure). The dash lines correspond to $R(\infty) = 1 - \psi(\theta(\infty))$ with $\rho = 0.05$ (equivalent to $I_0 = 50$ out of $N = 1000$ in simulations) , $\psi(\theta(\infty))$ corresponds to Eqs. (3.13-3.15) and Eqs. (3.20-3.23) from top to middle panel, respectively. The markers correspond to $\tau = 0.5, 1.0, ..., 4$ for simulation ($\circ$), pairwise ($\diamond$) and edge-based ($\bullet$). All numerical tests use $N = 1000$, $P(l) = 0.8$, $P(h) = 1 - P(l)$, $I_0 = 50$, $l = 3$, $h = 13$ and $\gamma = 1$, and simulations are averaged over 50 different network realisations and 50 simulations on each of these. The top and middle panel represent degree-dependent networks and networks with random weight distribution but with the same average weight as in the degree-dependent case $\langle w \rangle_{dd} = q_1 w_{ll} + q_2 w_{lh} + q_3 w_{hh}$, respectively. The bottom panel is simply the superposition of the top and middle panel, with continuous and dashed lines for degree-dependent and random weights, respectively.

with the same weight distribution but with the weights distributed at random, and the case of all weights equal to the average weight. For all cases we use a network where nodes can be of either a low or high degree, i.e. degrees of two types only. In Fig. 3.7, we present time evolution plots for the prevalence. There are several important observations that can be made. Firstly, the agreement between the pairwise, edge-based and simulation model is excellent for different parameter values and weight function combinations. Secondly, the distribution of weights has a significant impact on the time evolution of the epidemic with the homogenous/equal link-weight case giving rise to the fastest growing epidemic (see top panel of Fig. 3.7 for the strongest effect). The difference between the randomly distributed and equal weights cases is not significant, and both lead to fast epidemics compared to the original network model, where the epidemic is slower but lasts longer. All the features above become less pronounced if either the transmission rate, τ, increases (see the bottom panel of Fig. 3.7) or if the weights are of different magnitude. Both $w_{ij} = 1/(i+j)^{1/2}$ and $w_{ij} = 1/\ln(i+j)$ produce weights that have higher values when compared to the original $w_{ij} = 1/(i \times j)^{1/2}$ case. This explains the smaller differences in the middle and bottom panel of Fig. 3.7.

The marked difference in the time evolution of the epidemics can be explained intuitively by noting that on networks with degree-dependent weights, and especially when weights and degrees are inversely correlated, the important role played by highly connected nodes is negated by small link weights which makes transmission less likely. The slow initial growth in prevalence shows that the epidemic is 'struggling' to infect the highly connected nodes of the network, where link weights are the lowest. The transmission process is mainly capturing nodes that are less well connected with this process being favoured by larger link-weights. This effect fades away as the value of τ increases.

3.4.3 The principle of formally proving model equivalence

Our numerical results show remarkable agreement between the pairwise and the EBCM models, see Figs. 3.5-3.7. A careful analysis (is in a separate publication [77]) shows that while the two models appear to make different assumptions, they are in fact equivalent. We will give some insight into why this occurs. The central observation is that with both models, we will show that when considering two neighbours u and v, in our calculation of whether v has infected u it is rigorously possible to ignore whether any other neighbours have previously infected u.

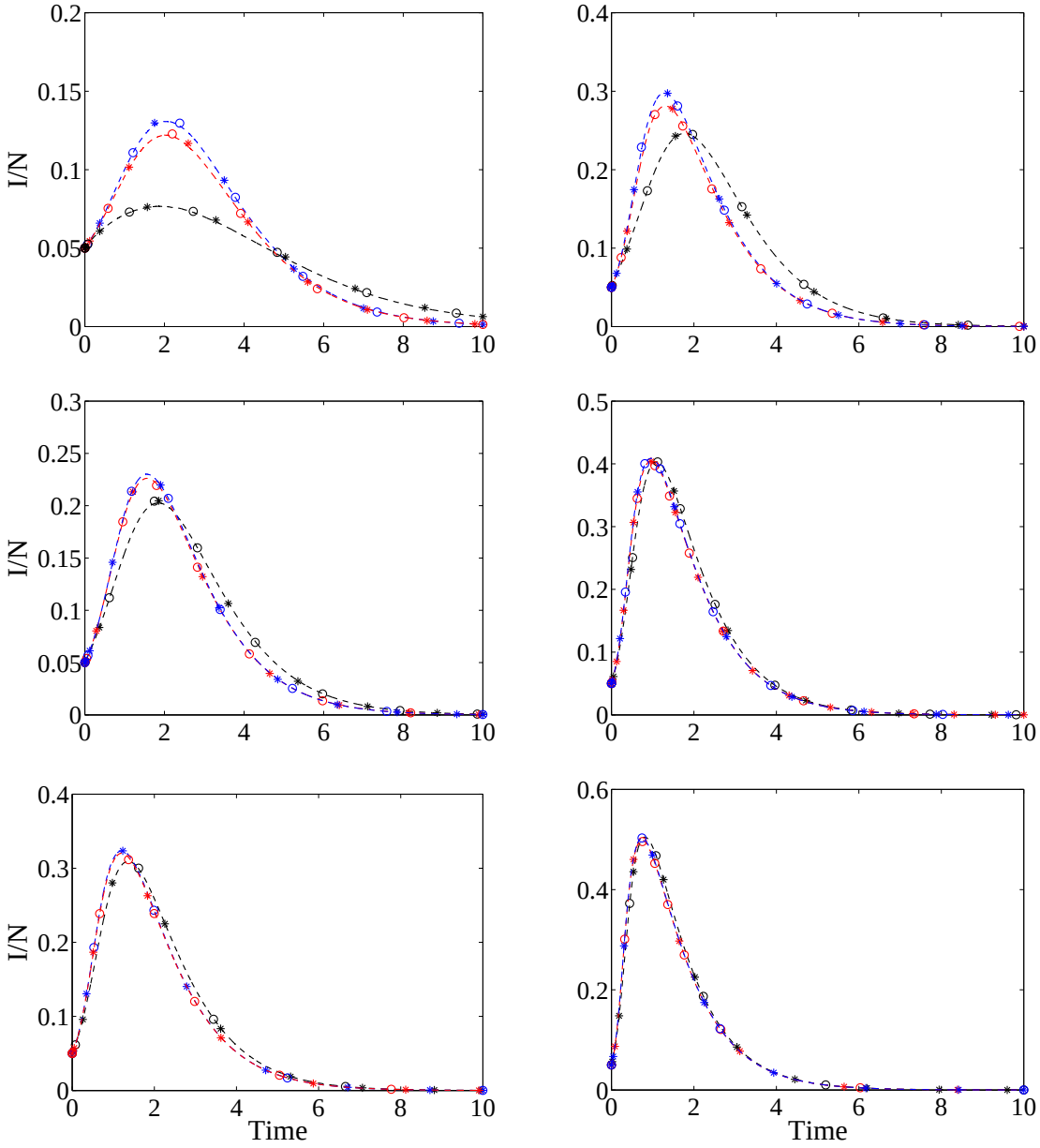


Figure 3.7: The infection prevalence (I/N) from heterogeneous weighted networks (simulation: dashed line, pairwise: ($\circ$), and edge-based: ($\star$)). All numerical tests use $N = 1000$, $P(l) = 0.8$, $P(h) = 1 - P(l)$, $I_0 = 50$, $l = 3$, $h = 13$, $\gamma = 1$, and simulations are averaged over 50 different network realisations and 50 simulations on each of these. Degree-dependent weighted networks (black), networks with randomly distributed weights (red) and networks with equal weights (blue). All networks have the same average weight $\langle w \rangle_{dd} = q_1 w_{ll} + q_2 w_{lh} + q_3 w_{hh}$. From top to bottom: $w_{ij} = 1/(i \times j)^{1/2}$, $w_{ij} = 1/(i + j)^{1/2}$ and $w_{ij} = 1/\ln(i + j)$, and left and right with $\tau = 2$ and $\tau = 4$, respectively.

The EBCM approach proceeds by starting with the initial problem of calculating the proportion of the population that is in each state. By assuming that the population-scale dynamics are deterministic, we can conclude that this must equal the probability that a random individual is in each state. So we transition to the equivalent problem of choosing a random individual u and calculating its probability of being in a given state. We seek to calculate the probability that a random neighbor v of u has transmitted infection to u. This is complicated by the fact that u might first transmit to v. However, we note that preventing u from transmitting to v after infection of u does not alter the probability that u is susceptible, infected, or recovered. Thus we find another equivalent problem: to calculate the probability that u is in each state given that it is prevented from transmitting to its partners. This sequence of arguments means that as we calculate whether v has transmitted to u, we can ignore whether or not another neighbor has already transmitted to u.

In the pairwise model, we look at the equations for the rates of change of $[S_k S_{k'}]$, $[S_k I_{k'}]$, and $[S_k R_{k'}]$ in Eq. (3.3). In each equation, there is a term on the right hand side which represents infection of the S_k individual by a partner other than the k' individual. After substituting our closure relation, each of these terms looks like $-[S_k S_{k'}]f$, $-[S_k I_{k'}]f$, and $-[S_k R_{k'}]f$ where

$$f = -\tau \frac{k-1}{k} \frac{w_{kq} \sum_q [I_q S_k]}{[S_k]} = \frac{k-1}{k} \frac{[\dot{S}_k]}{[S_k]}.$$

So each of equations is of the form $\dot{x} = -xf + y$ where the y terms represent other effects. By moving the xf term to the left hand side, we can use an integrating factor which yields a differential equation for the new variable xe^F where $\dot{F} = f$. The y terms remain in the equation, multiplied by e^F, but the term that represented infection of the S_k individual by a partner other than the k' individual has been eliminated. If we follow this change of variables and perform a few more simplifications, it is possible to arrive at the EBCM equations.

3.5 Discussion

In this paper we have shown that the pairwise and edge-based compartmental models can be successfully extended to specific cases of weighted networks and studied the non-trivial case of non-independence between weights and nodal degrees. In particular, we assumed that the link weight is inversely proportional to the degrees of the nodes

that it connects. This model has been compared to two null models where for both the network topology remains the same and only the distribution of weights changes. First, we considered the case when the original weights are 'lifted off' the edges and redistributed at random, thus making weights and nodal degrees independent, and secondly, the networks with all weights equal has been considered.

The results show that the negative correlation between weights and nodal degrees can negate the important role played by highly connected nodes in standard epidemic models on non-weighted graphs, and that weight heterogeneity but with the same overall average or total weight, reduce the value of R_0. The relation between final epidemic size and R_0, as expected, is determined by the model structure and, in this case, the same R_0 value leads to the biggest final epidemic size on degree-dependent weighted networks.

An important by-product of our analysis is the issue around model equivalence. This aspect emerged from the numerical evaluation and comparison of pairwise, edge-based and simulation models. The excellent agreement between all three, but especially, the agreement between pairwise and the edge-based model leads us to consider whether the two models are indeed equivalent. While, here we only present the basic idea of a formal proof, in [77] we will present detailed arguments to show the relationship between these models and other models for SIR epidemics on networks. We believe that in a model 'rich' environment, this part of our study and future work, as well as of others in the community [54], are important in trying to reconcile as much as possible different modelling approaches and to identify model hierarchies, as well as to pinpoint model efficiencies in terms of generating analytical or semi-analytical results.

Acknowledgements

P. Rattana acknowledges funding for her Ph.D. studies from the Ministry of Science and Technology, Thailand. JCM was supported in part by: 1) the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security and the Fogarty International Center, National Institutes of Health, and 2) the Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health under Award Number U54GM088558 from the National Institute Of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General

Medical Sciences or the National Institutes of Health. We thank Prof Péter L. Simon for pointing out a viable approach to proving that $R_0^{dd} \leq R_0^{rw}$.

3.6 Appendices

3.6.1 Appendix A: Proof of $R_0^{dd} \leq R_0^{rw}$

We wish to provide a formal proof that $R_0^{dd} \leq R_0^{rw}$. This amounts to showing that

$$\begin{aligned} R_0^{dd} &= \frac{R_1 + R_2 + \sqrt{(R_1 - R_2)^2 + 4F}}{2} \\ &\leq ((l-1)P_e(l) + (h-1)P_e(h))\,(q_1 r_1 + q_2 r_2 + q_3 r_3) \\ &= R_0^{rw}. \end{aligned}$$

We introduce the following notation: $x = P_e(h)$, $y = P_e(l)$. Then $y = 1 - x$ (with both $x, y \in [0,1]$) and

$$q_1 = y^2,\ q_2 = 2xy,\ q_3 = x^2,\ a = (h-1)xr_3,\ b = (l-1)yr_1,\ d = 1 - \frac{r_2^2}{r_1 r_3}.$$

We will make use of the following straightforward inequalities:

1. $r_3 \leq r_2 \leq r_1$,
2. $d \leq 0 \leftrightarrow r_1 r_3 \leq r_2^2$,
3. $(h-1)r_3 \geq (l-1)r_1$.

We also note that $r_3 \leq r_2$ implies that $(h-1)r_2 \geq (h-1)r_3 \geq (l-1)r_1$. These can be simply checked and formally proven by plugging in the corresponding expressions and performing some standard algebraic manipulation to reach some equivalent inequalities that trivially hold.

The l.h.s. of the inequality is the positive root of the quadratic polynomial

$$\lambda^2 - \lambda(a+b) + abd = 0,$$

where $abd \leq 0$, since d is negative while a and b are positive. Hence, the roots of this polynomial are denoted by $\lambda_2 < 0 < \lambda_1$. Then the following has to be proved,

$$\lambda_1 \leq [(h-1)x + (l-1)y](x^2 r_3 + 2xyr_2 + y^2 r_1).$$

First we give an upper estimate of λ_1. Using the formula for λ_1 and the inequality $\sqrt{1+x} \leq 1 + x/2$, one obtains

$$\lambda_1 = \frac{a+b+\sqrt{(a+b)^2 - 4abd}}{2} = \frac{a+b+(a+b)\sqrt{1 - 4abd/(a+b)^2}}{2} \leq a + b - \frac{abd}{a+b}.$$

It is also easy to show that $a+b \geq (l-1)r_1$. This can be done by considering $a+b = (h-1)r_3x + (l-1)(1-x)r_1 = [(h-1)r_3 - (l-1)r_1]x + (l-1)r_1$ as a function of x. Due to $(h-1)r_3 \geq (l-1)r_1$, the function above is monotone increasing, and since $x \in [0,1]$, the function will attain its minimum at $x=0$, and the minimum is $(l-1)r_1$. Using this in the inequality for λ_1 yields

$$\lambda_1 \leq a+b-\frac{abd}{a+b} \leq a+b-\frac{abd}{(l-1)r_1}.$$

Thus it is enough to prove that

$$a+b-\frac{abd}{(l-1)r_1} \leq [(h-1)x+(l-1)y](x^2r_3+2xyr_2+y^2r_1).$$

Let the difference of the l.h.s and the r.h.s be

$$f(x) = [(h-1)x+(l-1)y](x^2r_3+2xyr_2+y^2r_1) - (h-1)xr_3 - (l-1)yr_1 + d(h-1)r_3xy.$$

Then it is enough to prove that for all $x \in [0,1]$ we have $f(x) \geq 0$. Since $y = 1-x$, it is easy to see that $f(x)$ is a cubic polynomial and $f(0) = 0$, $f(1) = 0$. Hence, it is enough to prove that $f'(0) > 0$ and $f'(1) < 0$. Simple algebra shows that

$$r_1 f'(0) = (r_1 - r_2)\{r_1(h-l) + [(h-1)r_2 - (l-1)r_1]\} \geq 0,$$

based on that $r_1 \geq r_2$, $l \leq h$ and $(h-1)r_2 \geq (l-1)r_1$. The inequality $f'(1)$ develops as follows,

$$\begin{aligned} r_1 f'(1) &= (l-1)r_1(r_1-r_3) + (h-1)(r_2^2 - 2r_2r_1 + r_1r_3) \\ &\leq (h-1)r_3(r_1-r_3) + (h-1)(r_2^2 - 2r_2r_1 + r_1r_3), \end{aligned}$$

and this can be rearranged to give

$$r_1 f'(1) \leq (h-1)(r_2-r_3)(r_2+r_3-2r_1) \leq 0,$$

since $r_3 \leq r_2 \leq r_1$. Thus the original inequality holds.

3.6.2 Appendix B : Proof of the invariance of the final size and R_0 relation

First, let us consider the final epidemic size corresponding to networks with random weight distribution

$$R^{rw}(\infty) = 1 - (1-\rho)(P(l)\theta^l_{rw}(\infty) + P(h)\theta^h_{rw}(\infty)), \tag{3.26}$$

where $\theta_{rw}(\infty) = q_1\theta_{w_1}(\infty) + q_2\theta_{w_2}(\infty) + q_3\theta_{w_3}(\infty)$. Substituting Eqs. (3.13-3.15) into $\theta_{rw}(\infty)$ and using Eq. (3.12), we have

$$\begin{aligned}\theta_{rw}(\infty) &= q_1\frac{\gamma + (1-\rho)\tau_{rw}\ w_1\left[P_e(l)\theta_{rw}^{l-1}(\infty) + P_e(h)\theta_{rw}^{h-1}(\infty)\right]}{\tau_{rw}w_1 + \gamma} \\ &+ q_2\frac{\gamma + (1-\rho)\tau_{rw}w_2\left[P_e(l)\theta_{rw}^{l-1}(\infty) + P_e(h)\theta_{rw}^{h-1}(\infty)\right]}{\tau_{rw}w_2 + \gamma} \\ &+ q_3\frac{\gamma + (1-\rho)\tau_{rw}w_3\left[P_e(l)\theta_{rw}^{l-1}(\infty) + P_e(h)\theta_{rw}^{h-1}(\infty)\right]}{\tau_{rw}w_3 + \gamma} \\ &= \frac{\zeta - R_0^{rw}}{\zeta} + \frac{R_0^{rw}}{\zeta}(1-\rho)\left[P_e(l)\theta_{rw}^{l-1}(\infty) + P_e(h)\theta_{rw}^{h-1}(\infty)\right]. \end{aligned} \tag{3.27}$$

Next, the final epidemic size corresponding to networks with all weights equal to the average weight is

$$R^{av}(\infty) = 1 - (1-\rho)(P(l)\theta_{av}^{l}(\infty) + P(h)\theta_{av}^{h}(\infty)). \tag{3.28}$$

Similarly, based on Eqs. (3.4-3.5) and Eq. (3.24), and using that the average weight $w_{av} = \langle w\rangle_{dd} = q_1w_1 + q_2w_2 + q_3w_3$, $\theta_{av}(\infty)$ can be writhen as,

$$\begin{aligned}\theta_{av}(\infty) &= \frac{\gamma + (1-\rho)\tau_{av}(q_1w_1 + q_2w_2 + q_3w_3)\left[P_e(l)\theta_{av}^{l-1}(\infty) + P_e(h)\theta_{av}^{h-1}(\infty)\right]}{\tau_{av}(q_1w_1 + q_2w_2 + q_3w_3) + \gamma} \\ &= \frac{\zeta - R_0^{av}}{\zeta} + \frac{R_0^{av}}{\zeta}(1-\rho)\left[P_e(l)\theta_{av}^{l-1}(\infty) + P_e(h)\theta_{av}^{h-1}(\infty)\right]. \end{aligned} \tag{3.29}$$

Now, we start by assuming that $R^{rw}(\infty) = R^{av}(\infty)$, then Eqs. (3.26) & (3.28) leads to

$$\theta_{rw}(\infty) = \theta_{av}(\infty) = \theta \tag{3.30}$$

due to the function $f(x) = ax^l + bx^h$ being strictly monotonically increasing on our domain of interest $x \in [0,1]$, and beyond. Using Eq. (3.30) and Eqs. (3.27) & (3.29) yields

$$\frac{\zeta - R_0^{av}}{\zeta} + \frac{R_0^{av}}{\zeta}(1-\rho)\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right] = \frac{\zeta - R_0^{rw}}{\zeta} + \frac{R_0^{rw}}{\zeta}(1-\rho)\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right],$$

$$\begin{aligned}(R_0^{av} - R_0^{rw})\left(1 - (1-\rho)\left[P_e(l)\theta^{l-1} + P_e(h)\theta^{h-1}\right]\right) &= 0, \\ R_0^{av} &= R_0^{rw}.\end{aligned}$$

Chapter 4

Paper 3: Impact of constrained rewiring on network structure and node dynamics

P. Rattana[1], L. Berthouze[2,3] and I.Z. Kiss[1]

[1] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK

[2] Centre for Computational Neuroscience and Robotics, University of Sussex, Falmer, Brighton BN1 9QH, UK

[3] Institute of Child Health, London, University College London, London WC1E 6BT, UK

4.1 Abstract

In this paper, we study an adaptive spatial network. We consider a susceptible-infected-susceptible (SIS) epidemic on the network, with a link or contact rewiring process constrained by spatial proximity. In particular, we assume that susceptible nodes break links with infected nodes independently of distance and reconnect at random to susceptible nodes available within a given radius. By systematically manipulating this radius we investigate the impact of rewiring on the structure of the network and characteristics of the epidemic. We adopt a step-by-step approach whereby we first study the impact of rewiring on the network structure in the absence of an epidemic, then with nodes assigned a disease status but without disease dynamics, and finally running network and epidemic dynamics simultaneously. In the case of no labelling and no epidemic dynamics, we provide both analytic and semi-analytic formulas for the value of clustering achieved in the network. Our results also show that the rewiring radius and the network's initial structure have a pronounced effect on the endemic equilibrium, with increasingly large rewiring radiuses yielding smaller disease prevalence.

4.2 Introduction

The spread of infectious diseases on social networks and theoretical contact structures mimicking these has been the subject of much research [24, 29, 59, 82]. In general, most work in this area is aimed at understanding the impact of different network properties on how diseases invade and spread and how to best control them. Topological properties of nodes and edges can be exploited in order to minimise the impact of epidemics. For example, it is well known that isolating or immunising highly connected nodes or cutting edges or links with high betweenness centrality is far more efficient than selecting nodes and edges at random [2, 50]. When global information is scarce, acquaintance immunisation [20] provides an effective way to significantly reduce the spread of an epidemic. More recently, dynamic and time-evolving network models motivated by real data or simple empirical observations [44, 45, 46, 67, 101, 103, 106] have offered a different modelling perspective with important implications for how and when epidemics can spread or can be effectively controlled. It is widely accepted that during an epidemic the risk of becoming infected leads to social distancing with individuals either losing links or simply rewiring [18, 38, 45, 47]. Such action can in fact be seen as an emerging control strategy. In simple dynamic network models, contacts between susceptible and

infectious individuals can be broken, and new ones be established. This is usually implemented by susceptible individuals breaking high-risk contacts and rewiring to exclusively susceptible individuals or in a random way, or through random link addition and deletion [63]. It has been shown that this adaptive mechanism has a strong impact on both epidemic dynamics and network structure.

Another major development is the consideration of spatial or geometric networks [10], where nodes are embedded in space. This is especially the case for real networks where geographical or spatial location is key. For example, mobile phone, power grid, social contacts and neuronal networks are all embedded in space with location and proximity being a key component to how contacts are realised. This feature gives special properties to the network and allows to distinguish between nodes based on spatial proximity. For example, Dybiec et al. [31] proposed a modified susceptible-infected-recovered (SIR) model using a local control strategy where nodes are distributed on a one-dimensional ring, two-dimensional regular lattice, and scale-free network. While infection could spread on the whole network, including shortcuts, control could only act over a 'control network' composed of mainly local links but with neighbourhoods of varying size, e.g. including local neighbours one, two, or more links away. They presented simulation results showing how the effectiveness of the local control strategy depends on neighbourhood size, and they explored this relationship for a variety of infection rates.

In order to make rewiring more realistic, it is possible to combine dynamic or adaptive networks with a spatial component, where nodes are given specific locations [85], such that the rewiring may take these locations into account when identifying candidate nodes for rewiring. For example, Yu-Rong et al. [114] considered a network with a spatial component, where the rewiring strategy was such that when an SI link is cut, the S individual will reconnect, with some probability p, to random individuals irrespective of distance, and to close-by or neighbouring individuals with probability $1-p$. It was found that a higher value of the rewiring rate led to a lower final epidemic size whereas a smaller value of probability p resulted in a slower epidemic spread.

In this study, we investigate an susceptible-infected-susceptible (SIS) epidemic spreading on adaptive networks. Any susceptible node can avoid contact with infected nodes by cutting its links to infectious nodes and by rewiring them to other susceptible nodes. However, we make the assumption that individuals may not be able to avoid connecting to individuals who are in the same community (e.g., social circles such as

family, friends, or workplace acquaintances). That is, while the network is rewired adaptively, the rewiring is restricted to susceptibles who are in the same 'local' (to be defined later) area. The use of a square domain with periodic boundaries gives rise to a natural distance between nodes and this is used to determine the local area around nodes.

Since we anticipate that the size of local areas or neighborhoods will affect the rewiring, we carry out systematic numerical investigations of adaptive networks where rewiring is locally constrained. We adopt a step-by-step approach whereby we first study the impact of rewiring on the network structure in the absence of an epidemic, then with nodes assigned a disease status but without disease dynamics, and finally running network and epidemic dynamics simultaneously. In the case of no labelling and no epidemic dynamics, we provide both analytic and semianalytic formulas for the value of clustering achieved in the network in relation to the size of the local area.

The paper is structured as follows. In Sec. 4.3, we describe the construction of spatial networks to which constrained rewiring is applied, as well as the algorithm by which edges for rewiring are selected. We also present the impact of rewiring on degree distribution and clustering when rewiring operates in the absence of an epidemic (Secs. 4.3.1-4.3.2, respectively) and when the nodes are labelled (Sec. 4.3.3). Section 4.4 describes the epidemic model with constrained rewiring, as well as numerical simulations of both homogeneous and heterogeneous networks. In Sec. 4.5 we conclude the paper with a discussion of our results and possible further extensions of our work.

4.3 Adaptive network model with locally-constrained rewiring

In this section, the simplest adaptive network model with constrained rewiring is presented. Node placement and network construction are described by the following simple rules:

(a) N nodes are placed uniformly at random on a square $S = [0, X] \times [0, Y]$, such that each node i will have coordinates $0 \leq x_i \leq X$ and $0 \leq y_i \leq Y$, respectively, and $\forall i = 1, 2, \ldots, N$.

(b) Local area of radius R: If the Euclidian distance between nodes i and j is less than or equal to R, nodes i and j are said to be in the same local area and can become connected during the rewiring process.

All results in this paper are derived by considering $S = [0, \sqrt{N}] \times [0, \sqrt{N}]$, and internodal distances are calculated using periodic boundary conditions. With this choice, the density of nodes is exactly one node per unit area. Moreover, if the radius of the local area is R, then the circle, on average, will hold $n = \pi R^2$ nodes. Or if one wishes to control the expected number of nodes in a local area, then the radius is given by $R = \sqrt{n/\pi}$. Obviously, if $R \geq \sqrt{2N}/2$, the effect of spatial constraint is nonexistent as each node i has $N - 1$ potential neighbours to connect to. In what follows we will use either n, the expected number of nodes in a local area, or R, the radius of that area, as the control parameter of the rewiring process.

4.3.1 Rewiring at random within local areas and impact of the local area radius

We now investigate how changing the radius, which defines the local area for rewiring, affects the network structure. Here, in order to gain a better understanding of the rewiring algorithm, we study the network dynamics alone, in the absence of any dynamics of the nodes and without labelling nodes. Starting from the original idea of cutting a link between a susceptible node S and an infectious node I, and rewiring the susceptible to another S node randomly chosen among the set of all susceptible nodes [45], we consider two scenarios for implementing locally constrained rewiring. Specifically, we explore two different edge selection mechanisms:
(1) *Link-based selection:* a SI link is chosen at random (with equal probability), after which, the susceptible node S in the link is rewired to a randomly chosen available susceptible node S.
(2) *Node-based selection:* a susceptible node S is chosen at random and, if connected to an infectious node I, is rewired to a randomly chosen available susceptible S.

Unlike the node-based selection mechanism, the link-based selection mechanism favours highly connected nodes and therefore these two selection mechanisms have the potential to lead to networks with different properties. Note that, in both cases, once a prospective link or node has been identified, rewiring happens according to the local constraint, that is, rewiring happens only if at least one susceptible node S is available in the local area. Otherwise, rewiring is not performed. The total number of edges is kept constant throughout the simulations, and rewiring is not allowed if it leads to self-connections or multiple connections.

To begin to consider the impact of the network dynamics and show how it depends on

the choice of selection algorithm and size of local area, we consider two different starting conditions: (a) homogeneous and (b) heterogeneous Erdős-Rényi networks with average connectivity $\langle k \rangle = 10$. Then, when $R = \sqrt{2N}/2$ or $n = N$, the network will be in the situation where $\langle k \rangle \ll n$, whereas when $R = \sqrt{6/\pi}$, we will have $\langle k \rangle > n$. In one simulation step, only two outcomes are possible: the rewiring is successful (one link has been cut and a new 'local' link has been created) or the rewiring fails, as there are no suitable nodes in the local area. The latter tends to be more likely when the number of nodes in the local area is close to, or smaller than, the average connectivity, as this means that after a few successful steps, new links would lead to multiple or repeat connections, which are not allowed. The simulations or rewiring steps are performed until network characteristics such as degree distribution and clustering have stabilised.

Fig. 4.1 shows the average or expected degree distribution at steady state for both link-based and node-based selection methods. The good agreement between simulation and binomial distribution, when $R = \sqrt{2N}/2$, confirms that the degree distribution has not changed for the random network, but has changed significantly for homogeneous network with both selection methods leading to a heterogeneous network.

Starting from homogeneous and heterogeneous networks leads to different outcomes, with the difference most pronounced at the peak of the degree distribution when $R = \sqrt{6/\pi}$. Namely, the peak of the degree distribution when using link-based selection is higher than that obtained when using node-based selection, and the peak when starting from heterogeneous networks is less than that starting from homogeneous network. These differences can be explained as follows.

For small local areas, where the average number of nodes is smaller than the average degree or connectivity, the rewiring will not be able to rewire all original links such that the final, stable distribution remains relatively close to the original or starting distribution. Hence, starting with a homogenous network with distribution $p(k) = \delta(k - \langle k \rangle)$, i.e. $p(\langle k \rangle) = 1$, will lead to a network with a distribution that will maintain a high peak around $\langle k \rangle$. The heterogenous network has a much lower peak to start with, namely $p(\langle k \rangle) = \binom{N-1}{\langle k \rangle} p^{\langle k \rangle}(1-p)^{N-1-\langle k \rangle}$, where $p = \langle k \rangle/(N-1)$, and thus further limited rewiring will flatten the distribution further.

A similar explanation holds for the difference in the peak when the starting network is the same but the selection method differs. This is a result of the selection algorithm, and we will consider the case when the starting network is homogenous. Some nodes with connectivity higher than k will emerge quickly and these will be favourably picked

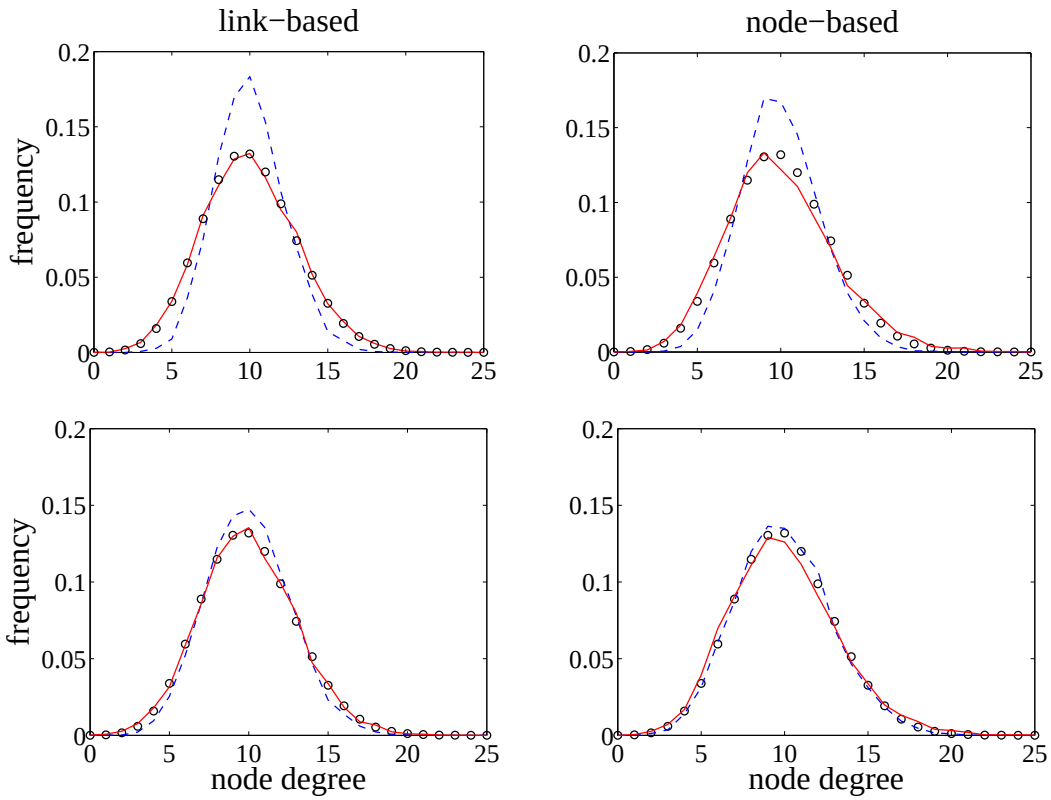


Figure 4.1: The average degree distribution at the end of simulations starting from homogeneous (top) and heterogeneous (bottom) networks compared with the binomial distribution $X \sim B(N-1, \langle k \rangle/(N-1))$ (black circles, corresponding to an Erdős-Rényi random network with N nodes and connectivity $\langle k \rangle$). The left and right panels correspond to link- and node-based selection, respectively. The plots show the average of 100 simulations with $R = \sqrt{2N}/2$ (red solid line) and $R = \sqrt{6/\pi}$ (blue dashed line), with $N = 100$ and $\langle k \rangle = 10$.

for rewiring when the link-based algorithm is used. However, this will only lead to conserving the nodes' degree, and rewiring will only lead to an increase in the maximal degree in the network if the target of the rewiring is itself one of the already highly connected nodes. This becomes very limiting and leads to little growth in degree, and thus to limited flattening of the distribution or decrease in its peak. This is exacerbated when the rewiring is limited by fewer available nodes than the average connectivity.

The size of the local area has a significant effect on the number of nodes in the area. If we consider small values of R, such as $R = \sqrt{6/\pi}$ and $\langle k \rangle > n$, then a typical node will connect to almost all nodes within the local area during the rewiring process. In other words, while the rewiring process is happening, the small number of nodes in

the area will become well connected and will lead to the formation of triangles, and thus increasing levels of clustering. In the extreme case with only three nodes in the local area, a triangle will quickly form. When the average connectivity is similar to the number of nodes in a local area, the rewiring process will create a significant number of closed loops of length 3, which will have a significant impact on the spread of a disease. To quantify this effect in a more rigorous way, we measure clustering in the network for local areas of different sizes as well as its evolution in time. Clustering can simply be calculated as the ratio of the number of triangles to connected triples, open or closed. This can be computed by simple operations on the adjacency matrix of the network as follows:

$$C = \frac{N_{\text{triangles}}}{N_{\text{triples}}} = \frac{\text{trace}(G^3)}{\|G^2\| - \text{trace}(G^2)},$$

where $G = (g_{ij})_{i,j=1,2,\ldots N} \in \{0,1\}^{N^2}$ and $g_{ij} = 1$ if there is a connection between node i and node j and $g_{ij} = 0$ otherwise.

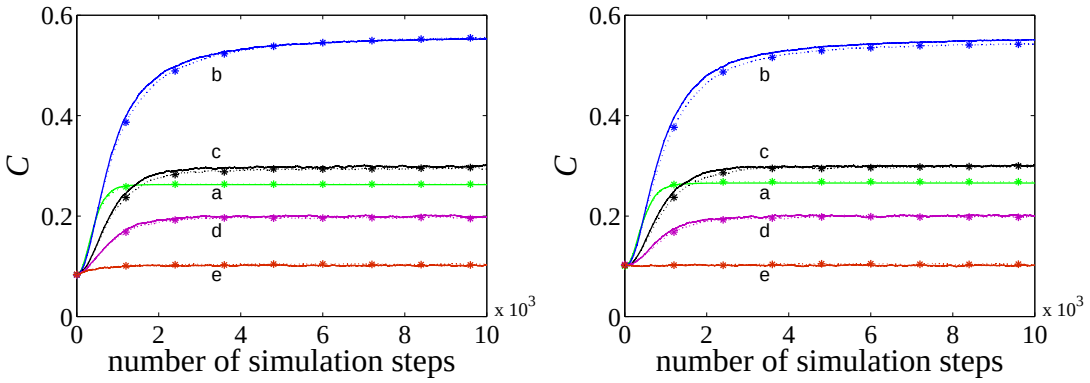


Figure 4.2: Evolution of clustering during rewiring, starting from homogeneous (left) and heterogeneous (right) networks. The plots show the average of 100 simulations with $R = \sqrt{6/\pi}$, $\sqrt{10/\pi}$, $\sqrt{20/\pi}$, $\sqrt{30/\pi}$ and $R = \sqrt{2N}/2$ (green (a), blue (b), black (c), purple (d) and red (e) lines, respectively), where the solid and dotted ($\star$) lines correspond to link- and node-based selection, with $N = 100$ and $\langle k \rangle = 10$.

Fig. 4.2 shows the evolution of clustering during rewiring for a range of radii R, and with both selection methods, as above. As expected, smaller values of R, but such that $\langle k \rangle \ll n$ still holds, lead to higher levels of clustering. However, when R is such that $\langle k \rangle \gg n$, clustering decreases as rewiring will be limited by the low number of potential targets for rewiring in local areas. This means that many long-range links from the original network will be conserved, and thus clustering is pushed to smaller values. Both

selection methods produce similar results in both clustering and preferential mixing for a variety of R values, with both homogeneous and heterogeneous starting networks.

It is observed that across all values of radius R, given enough time, clustering stabilises. This begs the question of how the rewiring process operates throughout the simulation, especially for large R. In Fig. 4.3, we examine how the the number of successful rewiring events depends on the simulation step when using node-based selection for both homogeneous and heterogeneous networks. As expected, with a small value of R, the rewiring process evolves quickly to a stable equilibrium, whereas, for a large value of R, it continues throughout the simulation. Interestingly, for large values of R, even when there are still prospective links or nodes to be rewired, clustering of the network is no longer affected (see Fig. 4.2 and Fig. 4.3 where $R = \sqrt{20/\pi}$, $\sqrt{30/\pi}$). Intuitively, this can be explained as follows. Since there are many available target nodes to rewire to in a local area, a node, with say k contacts, proceeds to randomly connect to k nodes within its local area. If the local area is not extremely large, and for relatively dense networks, this process will lead to an initial increase in clustering. Since the area holds more candidates for rewiring than the number of neighbours a node has, link rewiring will continue and other nodes from the same area will be chosen. However, this will lead to no significant further increase in clustering, except small movements around the equilibrium value.

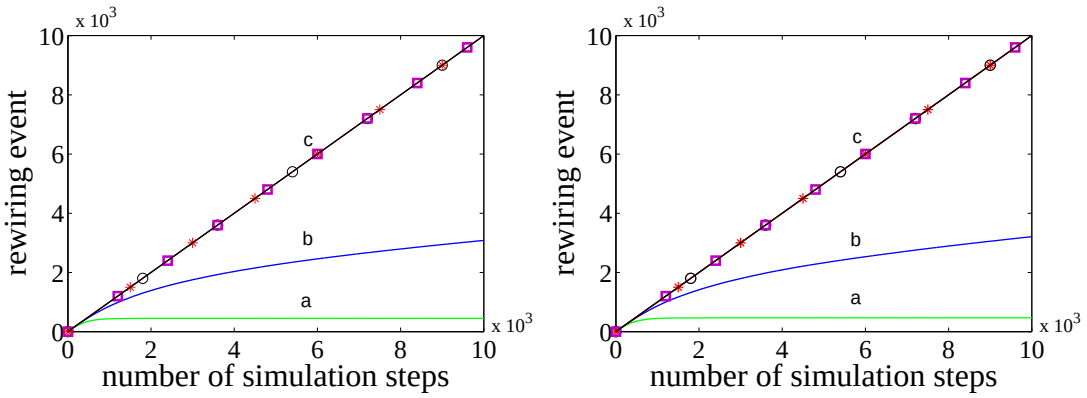


Figure 4.3: Evolution of the rewiring process, starting from homogeneous (left) and heterogeneous (right) networks with node-based selection. The plots show the average of 100 simulations with $R = \sqrt{6/\pi}$, $\sqrt{10/\pi}$, $\sqrt{20/\pi}$, $\sqrt{30/\pi}$ and $R = \sqrt{2N}/2$ (green (a), blue (b), black (c, o), purple (c, $\square$) and red (c, $\star$) lines, respectively), with $N = 100$ and $\langle k \rangle = 10$.

4.3.2 Computing clustering

A. $n \ll \langle k \rangle$: small areas but high degree

We aim to derive an analytical approximation for clustering by concentrating on the case when, on average, the number of nodes in a circle of radius R is less than the average degree in the network. In addition, we consider the situation when all possible links have been rewired. Due to having limited options for rewiring locally, we can assume that at the end of the rewiring process almost all local connections have been realised. We will focus on a typical node and its neighbours within distance R and beyond, noting that two nodes within a circle of radius R are not necessarily at a distance of less than R from each other.

Let us introduce some notation. Let B be the number of nodes within a radius R from a given node, and not including the node at the centre. B itself is a random variable. Let k be the degree of the node at the centre of the circle (k is therefore also a random variable). To compute the clustering of the central node we seek to establish the number of links between the neighbours of the node. We break this down into links between neighbours who are within the circle, links between internal and external neighbours, and finally links between nodes that are exclusively outside the circle. Counting multiplicatively, the total number of possible triangles is:

$$B(B-1) + 2B(k-B) + (k-B)(k-B-1) = k(k-1).$$

We now set out to find the probability of connections existing between the three different types of edges. First, we work out the probability of two interior nodes being connected. This can be done by considering a circle of radius R and then an arbitrary point within it. The probability that the second node will be within distance R from the initial node will be proportional to the overlap area $A_{overlap}$ between the original circle and the circle of radius R centred around the first random point. Hence, the probability that the distance between the two random points within the circle is less then R is simply

$$P(d < R) = \frac{A_{overlap}}{\pi R^2}.$$

To determine $A_{overlap}$, we first work out the density function for the distance of the first point from the centre. However, when placing nodes at random in a circle, the uniform random number has to be scaled with the $\sqrt{\cdot}$ function. Effectively, a good or valid random choice for the distance from the centre is not a uniform random number in (0,1), X, times R but $\sqrt{X(0,1)}R$. This means that the density function for the

distance from the centre of a randomly and uniformly placed node is $\rho(r) = \frac{2r}{R^2}$. This integrates to 1 for r going from 0 to R. Knowing the distance r between the two points, we average the well known area for the intersection of two circle of radii R and with distance r between their centres, that is,

$$A_{overlap}(r, R) = 2R^2 \cos^{-1}(\frac{r}{2R}) - \frac{1}{2}r\sqrt{4R^2 - r^2}.$$

Hence, the probability that two nodes within a circle of radius R are less than R apart is given by

$$q = \frac{1}{\pi R^2}\int_0^R A_{overlap}(r, R)\frac{2r}{R^2}dr, \tag{4.1}$$

and the number of triangles that are forming between interior nodes is $B(B-1)q$.

We now focus on the probability of links existing between the remaining non-connected interior-interior nodes (of which there are $B(B-1)(1-q)$), as well as between interior-exterior (i.e. $2B(k-B)$) and exterior-exterior (i.e. $(k-B)(k-B-1)$) nodes. In general, we can state that if the distance between two nodes is less than R then at the end of the simulation they will have formed a link. The probability that the distance between two randomly placed nodes is less than R is the ratio between the area of the circle or local area with respect to the total area. Thus, with probability $\frac{\pi R^2}{N}$, two nodes are less than R apart and are connected with probability 1. With probability $1 - \frac{\pi R^2}{N}$, these nodes will be more than R away and therefore will be connected by the long-range links that remain at the end of the rewiring process. However, the average number of such links is $(k-B)N$ with short-range links accounting for BN. Thus assuming that long-range links are distributed at random across all possible long-range pairs, we get that the probability of such a link existing is

$$p_{lr} = \frac{(k-B)N}{N(N-1)(1-\frac{\pi R^2}{N})}.$$

Hence, a random pair of nodes forms a link with probability

$$\frac{\pi R^2}{N} + (1 - \frac{\pi R^2}{N})p_{lr} = \frac{k+1}{N-1} - \frac{B+1}{N(N-1)} \sim \frac{k+1}{N-1},$$

since $\frac{B+1}{N(N-1)}$ is likely to be small. However, surprisingly, this value is very close to what is the initial probability of a link existing when the network is connected up according to the Erdős Rényi model. In this case, the probability of a link existing is $\frac{k}{N-1}$ which is also the measure of clustering for the initial network since all links are placed at random and thus where a node has two neighbours, the probability of them being connected is

$C = \frac{k}{N-1}$. However, at the end of the rewiring process we get that clustering should be well approximated by

$$\begin{aligned} C_L &= \frac{B(B-1)q + p_{lr}B(B-1)(1-q)}{k(k-1)} \\ &+ \frac{(\frac{k+1}{N-1} - \frac{B+1}{N(N-1)})\,[2B(k-B)]}{k(k-1)} \\ &+ \frac{(\frac{k+1}{N-1} - \frac{B+1}{N(N-1)})\,[(k-B)(k-B-1)]}{k(k-1)}. \end{aligned} \tag{4.2}$$

We expect that when clustering is high, the $B(B-1)q$ term dominates. We can also suggest a simpler formula for C, namely, one that assumes that almost all interior neighbours of a central node will become connected and the contribution from other pair types towards clustering is small. On the one hand, this overestimates clustering when looking at connections between interior nodes, as these could be apart by more than distance R. On the other hand, it underestimates clustering as some interior-exterior and exterior-exterior nodes could still be connected. This formula gives

$$C_a = \frac{B(B-1)}{k(k-1)}.$$

Both formulas above work on average or expected values. As noted previously, k and B can be treated as random variable with some distribution. An analytic or semi-analytic expression for these would make it possible to numerically evaluate our two approximations and compare them to clustering measured from simulations.

B. $n \gg \langle k \rangle$: large areas but low degree

Let us use the same definition of B and k as in the previous section, but here $B > k$. Our analysis will focus on a typical node out of the k nodes in the area. Since the probability of two nodes within a circle of radius R being connected is q and there are $B-1$ nodes in total available to form links, clustering should be approximated by

$$C_R = q\frac{k-1}{B-1}. \tag{4.3}$$

This formula works on the assumption that the centre node forms triangles only within its local area since $B > k$.

Both Eq. (4.2) and Eq. (4.3) are shown in Fig. 4.4. Here, we present only the case of homogeneous networks with node-based selection, due to the two rewiring methods

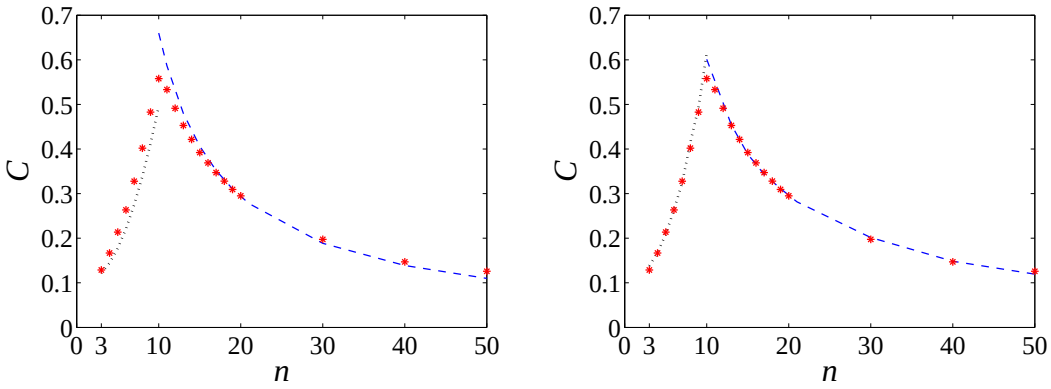


Figure 4.4: Clustering at the end of simulations starting from homogeneous networks with node-based selection. Simulation results (red $\star$) are compared with analytic formulas (Eq. (4.2) (black dotted) and Eq. (4.3) (blue dashed)), with $k = \langle k \rangle$. In the left panel we use the formulas with average (B, k) values. In the right panel we use (B, k)'s joint distribution computed from simulation. The plots show the average of 100 simulations with $N = 100$ and $\langle k \rangle = 10$.

giving very similar clustering values; see Fig. 4.2. The left panel of Fig. 4.4 uses average (B, k) values such that all centre nodes have B nodes within a radius R and have degree k. However, this is an approximation since in reality B and k are random parameters and have a joint distribution. When accounting for this heterogeneity by computing the joint parameter distribution from simulation, the agreement significantly improves as shown in the right panel of Fig. 4.4. Here, we randomly choose 5% of N nodes to be centre nodes and count the true values of B and k to compute the clustering.

While the analytic formulas for the clustering values are derived for the limiting cases of $n \ll \langle k \rangle$ and $n \gg \langle k \rangle$, a close examination of Fig. 4.4 reveals that agreement with simulation is maintained close to the $n \simeq \langle k \rangle$ regime. Moreover, the same figure shows that the maximum value of clustering is achieved for $n \sim \langle k \rangle$. By using this value in the analytic formulas, i.e., $B = n - 1 = \langle k \rangle - 1$, and by neglecting the small terms leads to

$$C_L = q\left(1 - \frac{1}{\langle k \rangle}\right) \text{ and } C_R = q\left(1 + \frac{1}{\langle k \rangle - 2}\right),$$

which shows that clustering will be dominated by the probability q that two nodes within a circle of radius R are less than a distance R apart. The value of q is independent of R and it is $q \sim 0.587$, as confirmed by our figure. While, C_L underestimates and C_R overestimates clustering at $n = \langle k \rangle$, it is worth noting that using $n = \langle k \rangle + 1$ or

$B = \langle k \rangle$ in both formulas, i.e., Eq. (4.2) and Eq. (4.3), we get

$$C_L = C_R = q.$$

Hence, we can conclude that clustering can be maximised if the expected number of nodes in the local area is very close or identical to the expected degree of a node. Such a setup will ensure that all potential neighbours can be drawn from inside a local area, and clustering will be dominated by the probability q that two nodes within a circle of radius R are less than a distance R apart.

For large n, $n \to N$, the reasoning that led to working out q breaks down, since for large R values almost all nodes are in the same unique area. This effectively means that $q \to 1$ and thus $C_R \to \frac{\langle k \rangle - 1}{N-2} \simeq \frac{\langle k \rangle - 1}{N-1}$ (for large N), which is the value of clustering in a random network.

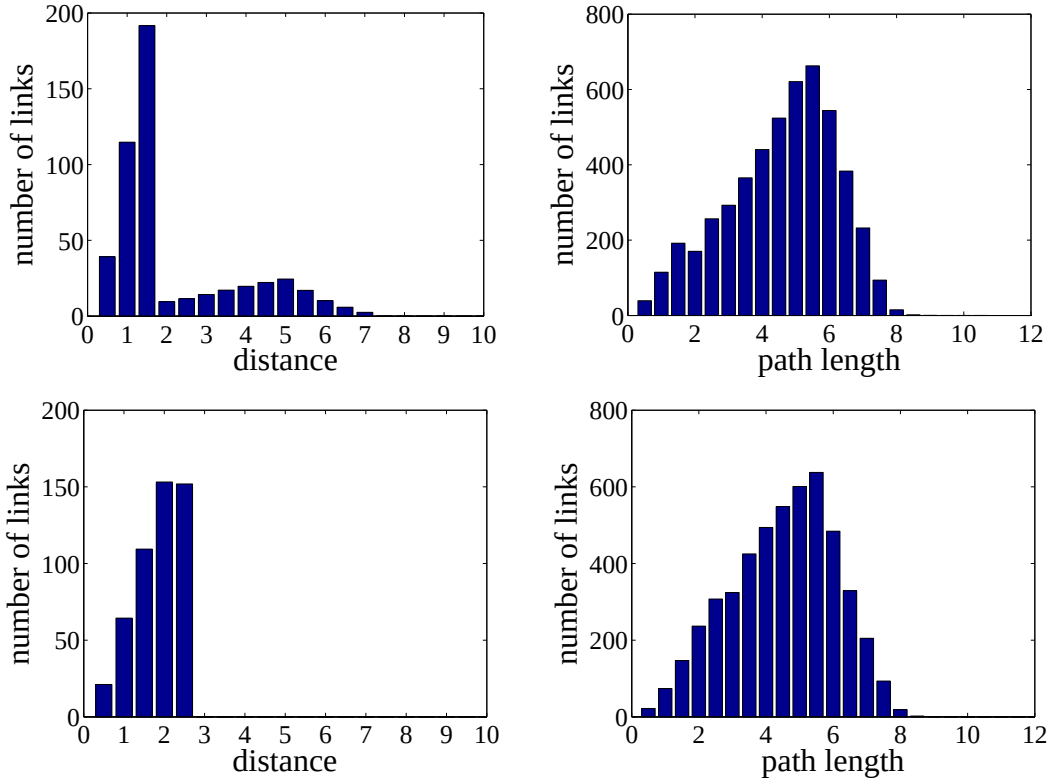


Figure 4.5: Distribution of distance between i and j if $g(i,j) = 1$ and distribution of path length at the end of simulations starting from homogeneous networks with node-based selection. The plots show the average of 100 simulations for $n = 7$ (top) and $n = 18$ (bottom) with $N = 100$ and $\langle k \rangle = 10$.

From Fig. 4.4, we note that networks with the same level of clustering can be gen-

erated with both $n \ll \langle k \rangle$ and $n \gg \langle k \rangle$. This begs the interesting question of whether structural differences exist in these networks. We examined a number of network characteristics, including path length distribution and distribution of true link lengths. Fig. 4.5 shows the distribution of distance for all links as well as the distribution of path length, for $n = 7$ and $n = 18$. As expected, with a large value of n, the rewiring will be able to rewire all links. Thus, the final network has all its links with length less than or equal to the value of R (see distribution of distance in Fig. 4.5 when $n = 18$). The final networks show a slight difference in mean path length, $L(n = 7) \approx 4.33$ and $L(n = 18) \approx 4.26$, even though their distributions of distance are significantly different. To further highlight the different network structures, Fig. 4.6 shows the small-worldness index of each final network as a function of n. This index is obtained by computing the ratio of C/L divided by the ratio of C_r/L_r where C_r and L_r are the clustering and mean path length respectively of the equivalent randomised network.

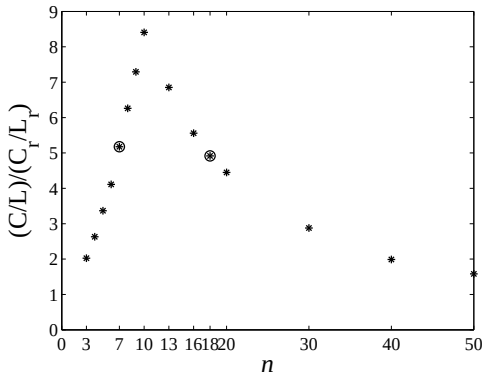


Figure 4.6: The small-worldness index $(C/L)/(C_r/L_r)$ at the end of simulations starting from homogeneous networks with node-based selection. The plots show the average of 100 simulations with $N = 100$ and $\langle k \rangle = 10$.

C. Comparisons to random geometric networks

In this section, we focus on properties of networks after rewiring has finished and in the particular case of n being close to $\langle k \rangle$. Assuming that all links can be rewired locally, all edges will have length of at most R. It is apparent that this description is closely related to that of random geometric graphs (RGGs) [23, 93]. Hence, it is worth considering how closely the two are related or whether these can be considered equivalent for some appropriately chosen parameter values.

Let us give a brief introduction to RGGs. A two-dimensional random geometric graph can be constructed by placing N nodes at random on the unit square, and assuming that each node is the centre of a circle of radius d. Nodes whose circles intersect or at least touch become connected [23, 10]. Thus, the probability p that two randomly chosen nodes are connected is equal to

$$p = \pi D^2,$$

where $D = 2d$. The average degree of a RGG is $\langle k \rangle \simeq pN$, and its degree distribution is binomial,

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \tag{4.4}$$

and well described by the Poisson distribution when N is large. The clustering coefficient of a RGG is

$$C_{RGG} = \frac{2}{D^2} \int_0^D \rho(x) x dx,$$

where $\rho(x)$ is the overlap area of two circles of radius D with distance x between their centres. Following [23], clustering is calculated as follows:

$$C_{RGG} = \frac{2}{D^2} \int_0^D x \frac{(\theta(x) - \sin\theta(x))}{\pi} dx = 1 - \frac{3\sqrt{3}}{4\pi} \simeq 0.587, \tag{4.5}$$

where $\theta(x) = 2cos^{-1}(x/2D)$.

Here, in the case of our dynamic networks, N nodes are randomly placed on an area given by $S = [0, \sqrt{N}] \times [0, \sqrt{N}]$. This is followed by a rewiring that allows connections only to nodes that are at most a distance R away. Hence, $d = R/2$ and $D = R$. If we were to follow the RGG rules, then the probability of two random nodes being connected would be $p = \frac{\pi R^2}{\sqrt{N}\sqrt{N}}$. If these nodes were then to be connected according to the RGG convention, ignoring the dynamic network, the average degree would be $\langle k \rangle = p(N-1) \simeq pN$. Thus to achieve a desired average degree, as in our starting network, one needs to set R according to

$$R = \sqrt{\frac{\langle k \rangle}{\pi}},$$

which for our specific case of $\langle k \rangle = 10$ gives $R = \sqrt{\frac{10}{\pi}}$.

From the above conditions, we expect that the stabilised dynamic network, when $R = \sqrt{10/\pi}$ or $n = 10$, is equivalent to a RGG network. We also note that in this case, when n is close to $\langle k \rangle$, the rewiring will almost surely be completed successfully,

i.e., all edges at the end of rewiring will be at most of length R. Fig. 4.7 shows the expected degree distribution at steady state for both link-based and node-based selection methods. As expected, we obtain excellent agreement between the degree distribution of the simulated networks and binomial distributions, when $p = \pi R^2/N$. This confirms that our final networks lead to the same degree distribution as that of RGG networks.

In order to explore the agreement between our and random geometric networks, we also consider the clustering value. In Sec.**B**, we have shown that when $n = \langle k \rangle$ then clustering is equal to q. To find the value of q, we use the overlap area between two circles of radius R and with distance r between their centers. This is given by $A_{overlap} = 2R^2 \cos^{-1}(\frac{r}{2R}) - \frac{1}{2}r\sqrt{4R^2 - r^2}$ or $A_{overlap} = R^2(\theta - \sin\theta)$, where $\theta(r) = 2cos^{-1}(r/2R)$. Substituting the latter into Eq. (4.1) yields

$$q = \frac{1}{\pi R^2}\int_0^R R^2(\theta - \sin\theta)\frac{2r}{R^2}dr = \frac{2}{\pi R^2}\int_0^R r(\theta - \sin\theta)dr. \tag{4.6}$$

This shows that the clustering values in the final rewired and RGG networks are identical, i.e., Eq. (4.6) is equivalent to Eq. (4.5). This is also confirmed from simulation results which yield $C_{n=10} = 0.587 = C_{RGG}$. While this confirms our results, we point out that RGGs appears only as a special 'limit' of the proposed dynamic network model, namely when $n = \langle k \rangle$.

4.3.3 Rewiring within local areas with SI labelling

To get closer to the full model (i.e., coupled epidemic dynamics and rewiring) and to gain more insights into the properties of the adaptive network, we now consider the scenario in which each node is assigned a disease status. Using the analogy of simple epidemic models, such as the SIS model, nodes are labelled at random as susceptible, S nodes, with probability p_s, and infected, I nodes, with probability $p_i = 1 - p_s$. We consider the network when the rewiring mechanism makes use of node labels, but without the full epidemic dynamics. This means that while the numbers of S and I are constant, the number of each type of links changes depending on type; namely, the number of SI decreases, the number of SS links increases and the number of II remains constant, thus changing the structure of the network. Provided that $S_0 = p_s N$ and $I_0 = (1 - p_s)N$, the initial link counts for SS, II and SI links are $S_0^2\langle k \rangle/2N$, $I_0^2\langle k \rangle/2N$ and $S_0 I_0 \langle k \rangle/N$, respectively, where each link is uniquely counted. When one of the SI links is cut and a new SS link is formed, it is obvious that the total number of

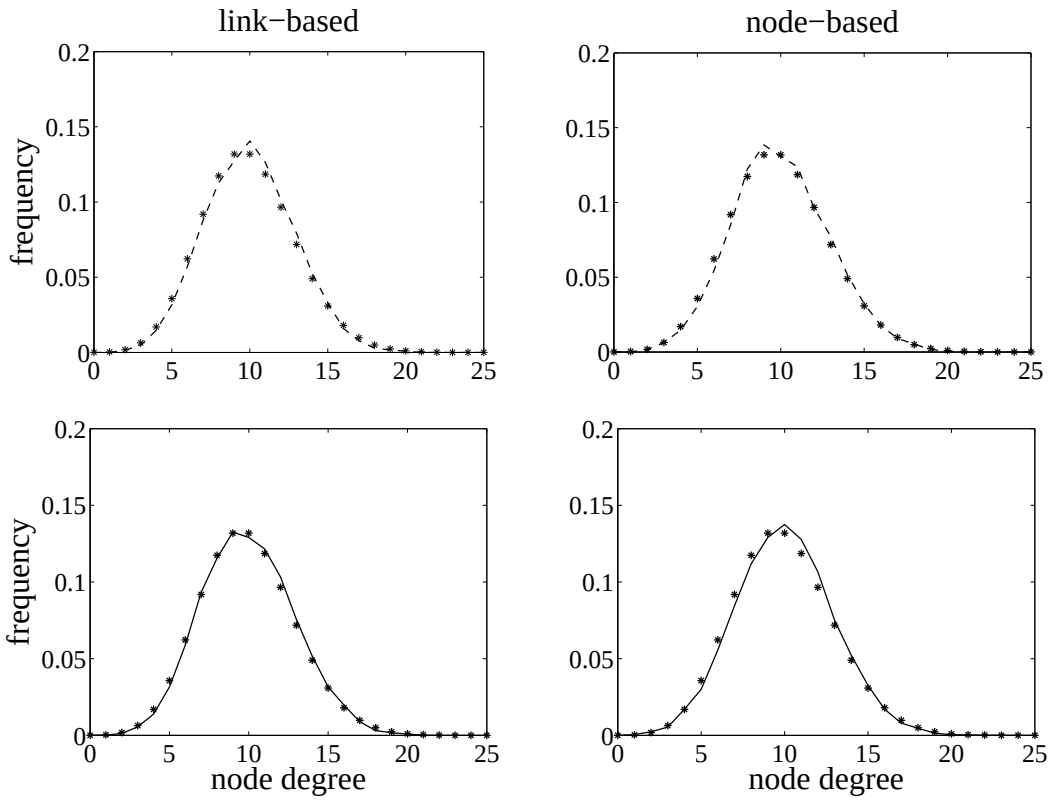


Figure 4.7: The average degree distribution at the end of simulations starting from homogeneous (top, dashed line) and heterogeneous (bottom, solid line) networks with $N = 100$, $\langle k \rangle = 10$ and $R = \sqrt{10/\pi}$, compared with Eq. (4.4) ($\star$) where $p = \pi R^2/N$. The left and right panels correspond to link- and node-based selection, respectively. The plots show the average of 100 simulations.

SS links increases relative to the (decreasing) number of SI links, and therefore, most S nodes in the network will evolve higher degrees.

This adaptive rewiring rule can lead to the network dividing into two sub-networks: one containing only S nodes and SS connections, and the other I nodes with II connections. Of course, this is not unique to the introduction of local rewiring constraints, i.e., $R < \sqrt{2N}/2$. Further, it should be noted that it is possible that not all SI links are cut. This can happen when there is a very small number of S nodes compared to a large number of I nodes or when the local neighbourhood or radius is very small. In this case, not all SI links can be cut since reconnection would lead to multiple links, which we do not allow.

To simplify the dynamics of the adaptive network, we start with $S_0 = 80\%$ of N and $I_0 = 20\%$ of N, and we allocate node labels at random. As previously, an SI link

is chosen at random, and the S node within this link reconnects to another S node in its local area, provided that such a node exists. Otherwise, the rewiring step is abandoned and a new SI link is selected. The simulation or rewiring is complete when either all SI links have been rewired or the remaining links cannot be rewired due to a lack of available S nodes in the local areas.

Impact of rewiring on the degree distribution of the network

To explore the impact of the rewiring dynamics (whereby only SS links can be formed) on network degree, we consider changes in degree distribution when starting with either homogeneous or heterogeneous networks.

(a) *Heterogeneous networks:*

When starting from a heterogeneous network at time $t = 0$, the network has a degree distribution given by the binomial distribution, namely, $p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$, where $p = \langle k \rangle / (N-1)$, and the average degree of both susceptible and infected nodes is equal to $\langle k \rangle$. We assume that the degree distribution of S and I nodes remain random throughout the simulation, and is binomial. First, let us consider the degree distribution of S nodes. We start by calculating the average degree of S nodes at time t. Let us define $\Delta k_S(t)$ as the rate of change of the average degree of S nodes, and assume that $\Delta k_S(t)$ depends on the number of SI links that are being cut at time t. Since the average degree of S nodes at the end of the simulations (when all SI links have been cut) is given by $(1 + i_0)\langle k \rangle$ [45], where $i_0 = I_0/N$, $\Delta k_S(t)$ can be computed as

$$\Delta k_S(t) = \Big[(1 + i_0)\langle k \rangle - \langle k \rangle\Big] \frac{[SI]^{cut}(t)}{[SI]_0} = i_0 \langle k \rangle \frac{[SI]^{cut}(t)}{[SI]_0},$$

where $[SI]_0$ is the initial number of SI links and $[SI]^{cut}(t)$ is the total number of SI links that have been cut up to time t. Then, as we know that all S nodes have degree $\langle k \rangle$ at $t = 0$, and the degree can only increase by Δk_S due to the rewiring process, we can calculate the average degree of a S node as

$$\begin{aligned} \langle k_S \rangle(t) &= \langle k \rangle + \Delta k_S(t) \\ &= \langle k \rangle + i_0 \langle k \rangle \frac{[SI]^{cut}(t)}{[SI]_0} \\ &= \Big[1 + i_0 \frac{[SI]^{cut}(t)}{[SI]_0}\Big] \langle k \rangle. \end{aligned}$$

Therefore, the degree distribution of a susceptible node can be written as

$$P(S = a)_t = \binom{N-1}{a} p_S^a (1 - p_S)^{N-1-a}, \tag{4.7}$$

where $a = 0, 1, 2, ..., N-1$ and $p_S = \frac{\langle k_S \rangle(t)}{N-1}$.

We can use the same methodology to derive $\Delta k_I(t)$, the average degree, and the degree distribution of I nodes. However, the degree of I nodes can only decrease by Δk_I and using the average degree of I nodes, $i_0\langle k \rangle$, when all SI links have been cut [45], we get

$$\begin{aligned}
\langle k_I \rangle(t) &= \langle k \rangle - \Delta k_I(t) \\
&= \langle k \rangle - \Big[\langle k \rangle - i_0\langle k \rangle\Big] \frac{[SI]^{cut}(t)}{[SI]_0} \\
&= \langle k \rangle - s_0\langle k \rangle \frac{[SI]^{cut}(t)}{[SI]_0} \\
&= \Big[1 - s_0 \frac{[SI]^{cut}(t)}{[SI]_0}\Big]\langle k \rangle.
\end{aligned}$$

Therefore, the degree distribution of an infected node can be written as

$$P(I = a)_t = \binom{N-1}{a} p_I^a (1 - p_I)^{N-1-a}, \tag{4.8}$$

where $a = 0, 1, 2, ..., N-1$ and $p_I = \frac{\langle k_I \rangle(t)}{N-1}$.

(b) *Homogeneous networks:*

We now focus on homogeneous networks for which the degree distribution of the network at time $t = 0$ is $p(k) = 1$, and the average degree of both susceptible and infected nodes is equal to k. Since we apply a random rewiring process, we assume that the network will evolve towards a random network with a binomial distribution, for both S and I nodes. As before, we assume that the average degree of S nodes increases by Δk_S, and the average degree of I nodes decreases by Δk_I, which depends on how many SI links are cut. In the case of S nodes, all S nodes start with exactly k links and their degree will increase to $k+1$, $k+2$, $k+3$, ..., $k+S_0-1$. Similarly, all I nodes start with k links and their degree will be decreased to $k-1$, $k-2$, $k-3$, ..., 0. So we have

$$\Delta k_S(t) = i_0 k \frac{[SI]^{cut}(t)}{[SI]_0},$$

and the degree distribution of a susceptible node can be written as

$$P(S = a)_t = \binom{S_0 - 1}{a} p_S^a (1 - p_S)^{S_0-1-a}, \tag{4.9}$$

where $a = 0, 1, 2, ..., S_0 - 1$, $\langle k_S \rangle(0) = k$, $\langle k_S \rangle(t) = \Delta k_S(t)$ and $p_S = \frac{\langle k_S \rangle(t)}{S_0-1}$.

In the case of I nodes, using the same approach as for heterogeneous networks yields

$$\langle k_I \rangle(t) = k - \Delta k_I(t) = \Big[1 - s_0 \frac{[SI]^{cut}(t)}{[SI]_0}\Big] k,$$

and, therefore, the degree distribution of an infected node can be written as

$$P(I = a)_t = \binom{k}{a} p_I^a (1 - p_I)^{k - k_I}, a = 0, 1, 2, ..., k, \tag{4.10}$$

where $p_I = \frac{\langle k_I \rangle(t)}{k}$.

Starting with the no-constraint scenario, $R = \sqrt{2N}/2$, Fig. 4.8 (left panel) confirms that the network has split into two disconnected networks, where the mean degrees of susceptible and infected nodes at the end of the simulations are given by $\langle k_S \rangle = (1 + i_0)\langle k \rangle$ and $\langle k_I \rangle = i_0 \langle k \rangle$ where $s_0 + i_0 = 1$. This is true when starting from either homogeneous or heterogeneous networks. As expected, the degree of S nodes can only increase, while the degree of I nodes strictly decreases. Starting with a homogeneous network, there is no S node with a degree less than $\langle k \rangle$, and the maximum degree of I nodes is at most $\langle k \rangle$ because all nodes have the same initial degree k. For both homogeneous and heterogeneous networks, there are disconnected I nodes at the end of the simulation, but, as discussed previously, this may result from the fact that $\langle k \rangle$ is not very high.

For small local areas, e.g., $R = \sqrt{6/\pi}$, where the average number of nodes in a local area is smaller than the average degree, the rewiring is restricted by the limited number of available S nodes. Therefore, the network evolves quickly to a stable equilibrium. This is clearly shown in Fig. 4.9 in which the evolution of clustering for $R = \sqrt{6/\pi}$ stops (due to all rewiring being complete) before that of other (larger) radii R.

These results are not solely dependent on the spatial constraint, but also on the number of initial SI links. Fig. 4.10 shows the clustering at the end of the simulations for a range of radii R and I_0 values. Starting with either homogeneous or heterogeneous networks produces similar results in clustering for a variety of R and I_0 values. As expected, the maximum clustering values for all sets of parameters n and I_0 are not higher than the maximum clustering value for networks with no node labelling, obtained previously (see Fig 4.4). A small number of initial S nodes leads to a small number of successful rewiring events (see Fig. 4.10 where $I_0 = 80$). This means that a larger value of R is needed in order to find available S nodes before cutting SI links, and therefore, we find that clustering increases as the value of R grows larger.

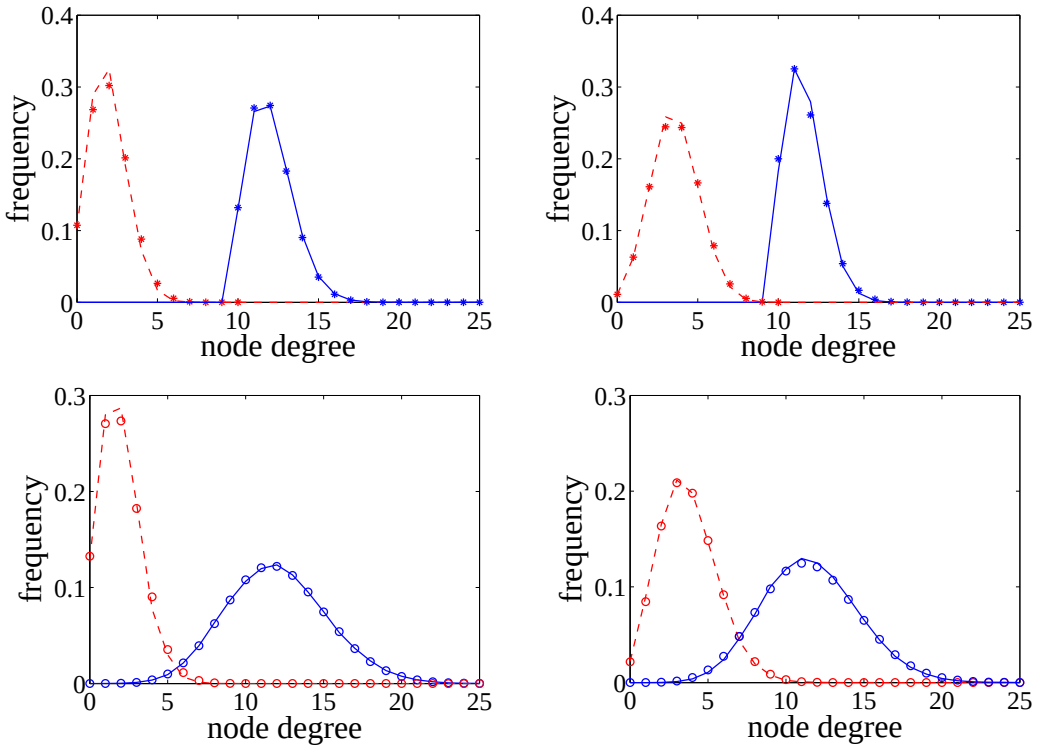


Figure 4.8: Average degree distribution of all S (blue solid line) and I (red dashed line) nodes at the end of simulations, when starting from homogeneous (top) and heterogeneous (bottom) networks with node-based selection. The plots correspond to the average of 1000 simulations with $N = 100$, $I_0 = 20$, $S_0 = N - I_0$, and $\langle k \rangle = 10$. In the left panel, $R = \sqrt{2N}/2$. In the right panel, $R = \sqrt{6/\pi}$. The blue and red ($\star$) markers correspond to Eq. 4.9 and Eq. 4.10, respectively. The blue and red ($\circ$) markers correspond to Eq. 4.7 and Eq. 4.8, respectively. We note that our analytic derivation needs the number of SI links that have been cut by the end of the rewiring process. This is taken from the simulation.

4.4 SIS models with constrained rewiring

In the previous section, we showed that the spatially constrained rewiring plays an essential role in determining network structure in the absence of any node dynamics. In this section, we extend this work by combining the dynamics of the network with the dynamics on the network in the form of the simple SIS model. The simulations are carried out on both homogeneous and heterogeneous networks, with a fixed size of N nodes and average degree of $\langle k \rangle$ links per node. The epidemic dynamics is specified in terms of infection and recovery events. The rate of transmission across an SI link is

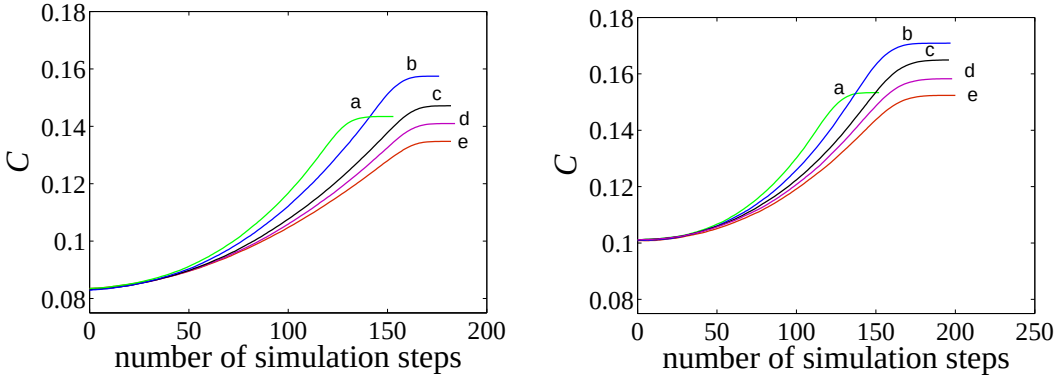


Figure 4.9: Evolution of clustering during rewiring, starting from homogeneous (left) and heterogeneous (right) networks. The plots correspond to the average of 1000 simulations with $N = 100$, $I_0 = 20$, $S_0 = N - I_0$, and $\langle k \rangle = 10$. Data for R values of $\sqrt{6/\pi}$, $\sqrt{10/\pi}$, $\sqrt{20/\pi}$, $\sqrt{30/\pi}$ and $\sqrt{2N}/2$ are shown in green (a), blue (b), black (c), purple (d) and red (e), respectively.

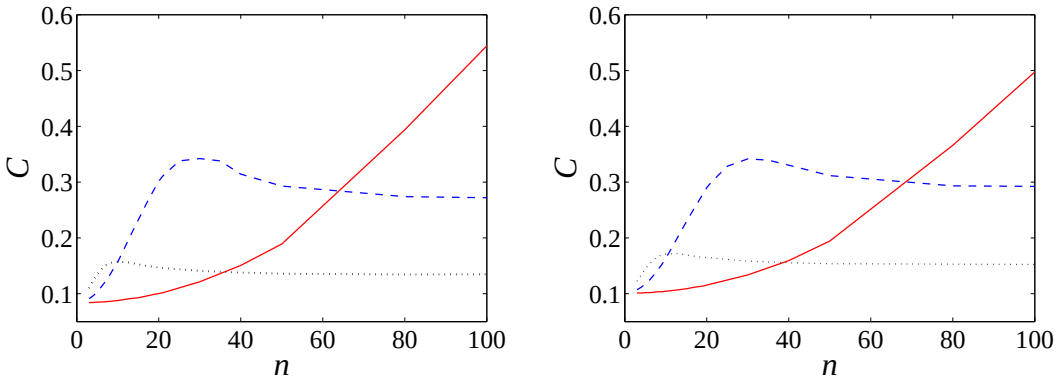


Figure 4.10: Final value of clustering when starting from homogeneous (left) and heterogeneous (right) networks. The plots correspond to the average of 1000 simulations with $N = 100$, and $\langle k \rangle = 10$. Data are shown for $I_0 = 20$ (black dotted line), $I_0 = 50$ (blue dashed line), $I_0 = 80$ (red solid line) with $S_0 = N - I_0$.

denoted by τ. Infected individuals recover independently of each other at rate γ. The network dynamics is specified in terms of rewiring events which affect SI links. Here, we make the assumption that the rewiring of an SI link depends on the number of susceptible nodes available for rewiring in the local neighbourhood of the S node that wishes to break its link to an I node and rewire to a susceptible one. It is natural to assume that the rewiring rate is proportional to the number of available S nodes that can accept new connections. For all SI links, this is achieved by using a rewiring rate equal to hw, where h is the number of available susceptible nodes within S's local area. We also assume that all processes are independent Poisson processes.

Simulations rely on synchronous updating with a small time step Δt, which guarantees that at most one event happens per iteration. Only three different types of event are possible during one time step Δt: (a) infection of a susceptible S node can occur with probability $1 - exp(-k\tau\Delta t)$, where k is the number of I neighbours, (b) an infectious I node can recover with probability $1 - exp(-\gamma\Delta t)$, and (c) a SI link can be rewired with probability $1 - exp(-hw\Delta t)$, as long as $h > 0$. This guarantees that rewiring only happens if viable candidates for rewiring exist and that the number of links in the network is constant throughout the simulation.

Given that the main focus of our study is the role of the spatially constrained rewiring, we will investigate the impact of the R (or n) values on whether epidemics die out and/or the endemic state becomes established. Specifically, we use the following definition to characterise the impact of the expected number of nodes in a local area or size of local area:

Definition 1. *n^* is the critical value of the expected number of nodes in a circle-like local area such that any greater value of n leads to disease extinction before a time T, or the end of the simulation, whichever comes first.*

The time evolution of infection on adaptive networks with constrained rewiring is shown in Fig. 4.11. Here, all simulations use the following parameter values: $N = 100$, $\langle k \rangle = 10$, $\gamma = 1$ and final simulation time $T = 100$. Simulations are started with infectious nodes chosen at random. The controlling effect of the local area radius R or expected number of nodes in a local area n is clear to see. As expected, with a small value of n, the network dynamics does not play a significant role in the control of epidemic spread for either homogeneous or heterogeneous networks. The small value of n affects the network dynamics in that the rewiring process can only happen briefly at the outset of the simulation and then stops while the epidemic dynamics continues

throughout the simulation.

Larger values of n, however, creates ideal conditions for rewiring and this can continue throughout the simulation, resulting in breaking many SI links. This scenario leads to a slowing down of the spread of the epidemic and a reduced infection prevalence. This is confirmed by Fig. 4.11, which shows small levels of infection prevalence for $n = 15$ and $n = 20$. The same figure also shows smaller and smaller endemic levels when the rewiring radius passes through the critical expected number n^*, namely, $n^* = 26$ for homogeneous networks and $n^* = 29$ for heterogeneous networks.

To further understand the relationship between the critical value n^* and the disease parameters, we systematically varied the infectious and rewiring rates (with fixed recovery rate) and recorded the corresponding critical n^* value. Fig. 4.12 shows the resulting surface for both homogeneous and heterogeneous networks, where τ varies from 0.15 to 3.5 in steps of 0.05 and w varies from 0.05 to 0.35 in steps of 0.05.

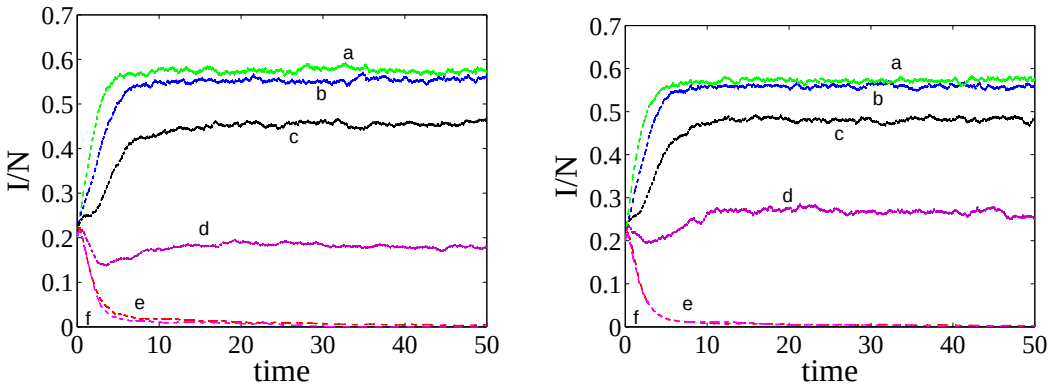


Figure 4.11: Infection prevalence (I/N) starting from homogeneous (left) and heterogeneous (right) networks. The plots correspond the average of 200 simulations with $N = 100$, $I_0 = 20$, $S_0 = N - I_0$, $\langle k \rangle = 10$, $\gamma = 1$, $\tau = 0.25$, $w = 0.2$. Data are shown for n values of 5 (green - a), 10 (blue - b), 15 (black - c), 20 (purple - d), critical value $n^* = 26$ for homogeneous network and $n = 27$ (red - e and pink - f, left panel), and critical value $n^* = 29$ and $n = 30$ (red - e and pink - f, right panel).

Increasing values of n increase the rewiring rate hw, since h will be higher due to more targets for the rewiring being available. This in turn leads to an active rewiring process which results in an overall decrease in the endemic equilibrium or in the extinction of the epidemic.

It is found that when n is large, the starting configuration of the network affects the endemic equilibrium in so far as starting with a homogeneous network leads to a smaller

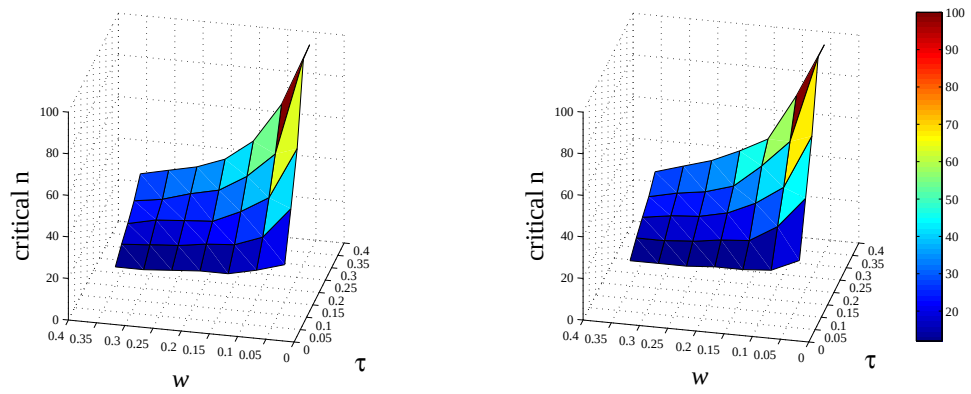


Figure 4.12: Critical n as a function of τ and w, starting from homogeneous (left) and heterogeneous (right) networks, and with $N = 100$, $I_0 = 20$, $S_0 = N - I_0$, $\langle k \rangle = 10$ and $\gamma = 1$.

epidemic, at a given n, than when starting with a heterogeneous network. Typically, the critical n^* is higher for heterogeneous networks, meaning that rewiring needs to be less constrained in order to curtail the epidemic. In general, for all n values, the epidemic will spread faster on heterogenous networks early on in the epidemic, when the link rewiring is still limited. However, as the networks are rewired, this effect is weakened as the homogenous network will become more heterogenous and will become more similar to the networks started with heterogenous degree distributions. Nevertheless, the critical threshold differs between homogenous and heterogenous networks, which may reflect a buildup of structural correlations or differences which may differentially affect the endemic prevalence.

In Fig. 4.13, we present the final clustering value for a range of radii R for both the full model and the model with no epidemic or labelling. The simulation results show that the impact of changing the radius on network structure is similar in both cases. Specifically, high values of n, but with $\langle k \rangle \gg n$ (the region to the left of the vertical line), result in higher levels of clustering, whereas when n is such that $\langle k \rangle \ll n$ (see the region to the right of the vertical line), clustering decreases, irrespective of which network is used. It is worth noting that the analysis of the dynamic network model alone, without labelling or epidemic, gives a clear insight into how the structure of the network changes. Observations from this analysis still hold in the full model, but as expected, the clustering of networks in the full model is less than in the network-only model since labelling reduces the number of nodes that can be used when rewiring. Higher clustering values in the full model are due to the epidemics dying out quickly

with no further rewiring, and thus with the network displaying a clustering value that is close to the values observed in the starting network. For the network-only model or for full-blown epidemics, however, the network will be fully randomised.

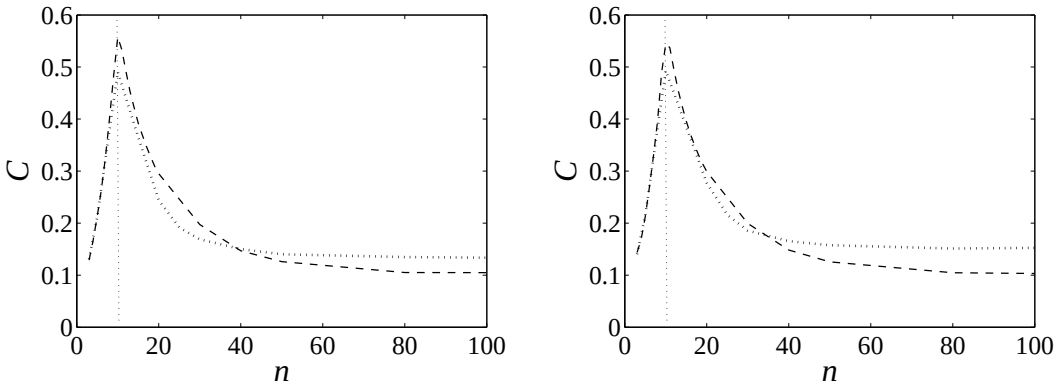


Figure 4.13: Final value of clustering starting from homogeneous (left) and heterogeneous (right) networks with $N = 100$ and $\langle k \rangle = 10$. The dashed line shows the clustering in a network without any dynamics of the nodes and without node labelling. The dotted line denotes the clustering when the full model, couple epidemic dynamics and rewiring, is considered, with $I_0 = 20$, $S_0 = N - I_0$, $\gamma = 1$, $\tau = 0.25$ and $w = 0.2$.

4.5 Discussion

The present study explored the effect of spatially constrained rewiring on an SIS epidemic unfolding on an adaptive network. Specifically, the dynamics of the network was achieved by breaking links and reconnecting to nodes within a local area. A step-by-step approach was taken in which the network dynamics was studied first in the absence of disease dynamics, then with node labelling but no dynamics, and finally with both network and node dynamics. Two different starting networks were used and analysed. In all models, a range of radii R, giving circular neighbourhoods, within which to rewire, was considered and shown to provide the means to control epidemic outbreaks. Spatially limited rewiring provides a more realistic mechanism than choosing partners to rewire to from the entire population. It is highly likely that in most situations, rewiring will be limited to a small subpopulations or set of individuals.

Our study provided a detailed analysis of the impact of constrained rewiring on the structure of the network. In particular, we were able to give analytic and semianalytic results for degree distribution and clustering. These showed excellent agreement with

simulations, and we have revealed that it is possible to generate networks with the same mean path length and the same clustering but significantly different distribution of real link lengths. This comes in support to the findings of [102] that networks with the same clustering can have substantially different higher-order network structure. This needs further investigation, possibly using more complex node dynamics to reveal how subtle differences in the network structure may impact on the outcome of dynamical processes supported by the network.

Further results provided analytical formulas for the degree distributions of susceptible and infected nodes which again showed good agreement with simulation results. These also confirmed that starting from a heterogeneous network, and when R is equal to $\sqrt{2N}/2$ or in the absence of spatial constraints for rewiring, the average degree of S and I nodes are $(1 + i_0)\langle k\rangle$ and $i_0\langle k\rangle$, respectively, which is in line with [45].

Finally, we have shown that even constrained rewiring can serve as a potent control measure. We highlighted that the expected number n in a typical local area is a key parameter which influences the network dynamics and can determine whether disease dies out or becomes endemic. Extensions to the methodology presented in this paper include considering other forms of constrained rewiring, e.g., network models where locality is not just defined in terms of spatial distance but possibly some more abstract or general metric, and understanding how this impacts network structure and processes, other than epidemics, taking place on the network.

Acknowledgements

P.R. was supported by a Ph.D. scholarship from the Ministry of Science and Technology of the Royal Thai Government.

Chapter 5

Paper 4: Comment on "A BINOMIAL MOMENT APPROXIMATION SCHEME FOR EPIDEMIC SPREADING IN NETWORKS" in U.P.B. Sci. Bull., Series A, Vol. 76, Iss. 2, 2014

Istvan Z. Kiss[1] & Prapanporn Rattana[1]

[1] School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK

5.1 Abstract

In this short comment we report on our test of the generalisation proposed by Shang in [104]. Shang in [104] claims to generalise previous results developed by Kiss and Simon in [66] and Nagy, Kiss and Simon in [84]. However, our tests show that the proposed generalisation performs poorly for all networks proposed by Shang, except for heterogenous networks with high average degree. While the binomial closure gives good results, in that the solution of the Kolmogorov equations, with the newly proposed rates, agrees well with the closed system, the agreement with simulation is extremely poor.

5.2 Introduction

Kiss and Simon in [66] considered the susceptible-infected-suceptible (SIS) dynamics on a fully connected network with N nodes. The model was formulated in terms of the master equation given by

$$\dot{p_k}(t) = a_{k-1}p_{k-1}(t) - (a_k + c_k)p_k(t) + c_{k+1}p_{k+1}(t), \tag{KE}$$

where $p_k(t)$ is the probability that there are k infectious nodes at time $t \geq 0$, with $k = 0, 1, 2, \ldots, N$. Furthermore, the rates of infection, a_k, and rates or recoveries, c_k are given by

$$a_k = \tau k(N-k), \quad c_k = \gamma k \quad \text{for} \quad k = 0, 1, \ldots, N \quad \text{with} \quad a_{-1} = c_{N+1} = 0.$$

All infection and recovery processes are modelled as independent Poisson processes. The infection rates encode all the information about the network, and the rate of recovery is simply a rate corresponding to pooled Poisson processes. Kiss and Simon in [66] show that rather than solving this full system, it is possible to derive a low-dimensional ODE based on the assumption that the number of infectious nodes is binomially distributed. Namely, it is assumed that $p_k(t)$ is distributed binomially, i.e. $\mathcal{B}(n, p)$, where n and p depend on time.

More precisely, the low-dimensional ODE is formulated for the first moment of the distribution, and this will also involve the second moment and the third. However, due to the assumption that $p_k(t)$ is binomially distributed, it is possible to express the third moment in terms of the first and second. This then yields an ODE system with

2 equations only. We briefly focus on deriving equations for the moment. Namely, for

$$y_j(t) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^j p_k(t) \quad \text{or} \quad Y_j(t) = \sum_{k=0}^{N} k^j p_k(t), \tag{5.1}$$

where $N^j y_j = Y_j$ with $j = 1, 2, \ldots$. Deriving evolution equations for these is straightforward. For example, the derivative of the first moment, and in a similar way for all other moments, can be given in function of higher-order moments upon using the Kolmogorov equations, Eq. (KE). The derivation for the first moment is outlined below,

$$\begin{aligned} \dot{Y}_1(t) &= \sum_{k=0}^{N} k\dot{p}_k = \sum_{k=0}^{N} k(a_{k-1}p_{k-1} - (a_k + c_k)p_k + c_{k+1}p_{k+1}) \\ &= \sum_{k=0}^{N} (ka_{k-1}p_{k-1} - ka_k p_k - kc_k p_k + kc_{k+1}p_{k+1}). \end{aligned}$$

By changing the indices of the summation, plugging in the corresponding expressions for the transition rates a_k and c_k, and taking into account that $a_{-1} = c_{N+1} = 0$ the following expression holds,

$$\dot{Y}_1(t) = \sum_{k=0}^{N} (\tau(k + k^2)(N - k) - \tau k^2(N - k) - k^2\gamma + (k^2 - k)\gamma)p_k.$$

Based on our notations, see Eq. (5.1), the equation above reduces to

$$\dot{Y}_1(t) = \tau N Y_1 - \tau Y_2 - \gamma Y_1. \tag{5.2}$$

We emphasise that this was possible due to the special form of the a_k coefficients, namely that these are quadratic polynomials in k. Using a similar procedure, the equation for the second moment Y_2 can be easily computed and is given by

$$\dot{Y}_2 = 2(\tau N - \gamma)Y_2 - 2\tau Y_3 + (\tau N + \gamma)Y_1 - \tau Y_2. \tag{5.3}$$

Equations (5.2) & (5.3) can be recast in terms of the density dependent moments y_js to give

$$\dot{y}_1 = (\tau N - \gamma)y_1 - \tau N y_2, \tag{5.4}$$

$$\dot{y}_2 = 2(\tau N - \gamma)y_2 - 2\tau N y_3 + \frac{1}{N}\left((\tau N + \gamma)y_1 - \tau N y_2\right). \tag{5.5}$$

The above equations are not closed or self-contained since the second moment depends on the third and an equation for this is also needed. It is easy to see that this dependence of the moments on higher moments leads to an infinite but countable number of

equations. Hence, a closure is needed and below we show that it is possible to express Y_3 as a function of Y_1 and Y_2. The first three moments of the binomial distribution can be specified easily in terms of the two parameters and are as follows,

$$\begin{aligned} Y_1 &= np && (5.6)\\ Y_2 &= np + n(n-1)p^2 && (5.7)\\ Y_3 &= np + 3n(n-1)p^2 + n(n-1)(n-2)p^3. && (5.8) \end{aligned}$$

Using Eqs. (5.6) & (5.7), n and p can be expressed in term of Y_1 and Y_2 as follows,

$$p = 1 + Y_1 - \frac{Y_2}{Y_1}, \qquad n = \frac{Y_1^2}{Y_1 + Y_1^2 - Y_2}. \tag{5.9}$$

Plugging the expressions for p and n, Eq. (5.9), into Eq. (5.8), the closure for the third moment is found to be

$$Y_3 = \frac{2Y_2^2}{Y_1} - Y_2 - Y_1(Y_2 - Y_1).$$

This relation defines the new closure, and in terms of the density dependent moments this is equivalent to

$$y_3 = \frac{2y_2^2}{y_1} - y_1 y_2 + \frac{1}{N}(y_1^2 - y_2).$$

Using the equation for the first moment, Eq. (5.4), the closure at the level of second moment yields the following approximate equation

$$\dot{x}_1 = (\tau N - \gamma)x_1 - \tau N x_1^2.$$

Using the equations for the first two moments, Eqs. (5.4) & (5.5), and the closure at the level of the third moment yields

$$\begin{aligned} \dot{x_1} &= (\tau N - \gamma)x_1 - \tau N x_2,\\ \dot{x_2} &= 2(\tau N - \gamma)x_2 - 2\tau N x_3 + \left(\left(\tau + \frac{\gamma}{N}\right)x_1 - \tau x_2\right), \end{aligned}$$

where

$$x_3 = \frac{2x_2^2}{x_1} - x_1 x_2 + \frac{1}{N}(x_1^2 - x_2).$$

Hence, we have derived two approximate system, with the first and second closed at the level of the second and third moment, respectively. It is in general true that the higher the moment at which the closure the more likely that the resulting approximate model performs well. We note that we used x instead of y to highlight that the closed systems, define in term of x, are only an approximation to the exact system given in terms of y.

The major challenge is generalising this to arbitrary networks is in finding a correct functional form for the infection rates a_k for any network in general. Kiss and Simon [66] have shown that for homogenous random networks and based on the random mixing argument a_k can be written as

$$a_k = \tau(N-k)\langle k\rangle \frac{k}{N-1},$$

where it is assumed that infectious nodes are distributed at random around susceptible nodes. Our numerical experiments also show that such a formula also performs well for Erdős-Rényi random networks. For other graphs no such immediate or intuitive formula exists.

Shang in [104] proposed that a_k in general could be written as

$$a_k = \frac{\tau k(N-k)\langle k^2\rangle}{\langle k\rangle(N-1)}, \quad c_k = \gamma k \quad \text{for} \quad k = 0, 1, \ldots, N \quad \text{with} \quad a_{-1} = c_{N+1} = 0, \tag{5.10}$$

where the network is given in terms of a degree distribution with $P(k)$ denoting the probability that a randomly chosen node has degree k, with $k = 0, 1, 2, \ldots, N-1$ for a network of size N. Moreover $\langle k\rangle = \sum kP(k)$ and $\langle k^2\rangle = \sum k^2P(k)$. While there is no explicit explanation for this, we can heuristically explain how such a formula could be arrived at. A newly infected node, under the assumption of random mixing will have degree l with probability $lP(l)/\langle k\rangle$. Hence, such a node has l onward connections and one such links leads to a susceptible node with probability $(N-k)/(N-1)$. Putting this together for a single node and averaging across all degrees gives

$$\sum_l \frac{lP(l)}{\langle k\rangle} \times l \times \frac{N-k}{N-1},$$

and upon multiplying this with k, the number of infectious nodes, yields

$$a_k = \frac{\tau k(N-k)\langle k^2\rangle}{\langle k\rangle(N-1)}.$$

Shang then used the same procedure as above to derive a set of 2 ODEs for these potentially more general infection term. His closed system yields

$$\dot{x}_1(t) = \left(\frac{\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} - \gamma\right) x_1 - \frac{\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} x_2, \tag{5.11}$$

$$\dot{x}_2(t) = \left(\frac{\tau\langle k^2\rangle(2N-1)}{\langle k\rangle(N-1)} - 2\gamma\right) x_2 - \frac{2\tau\langle k^2\rangle N}{\langle k\rangle(N-1)} x_3 + \left(\frac{\tau\langle k^2\rangle}{\langle k\rangle(N-1)} + \frac{\gamma}{N}\right) x_1, \tag{5.12}$$

where the same closure applies, namely

$$x_3 = \frac{2x_2^2}{x_1} - x_1 x_2 + \frac{1}{N}(x_1^2 - x_2).$$

5.3 Testing Shang's generalisation

To carry out our tests we used the same networks and parameters as given in Shang's paper [104]. We note that some of these choices are not natural, as the proposed network has a very low average degree, which in general makes it very difficult to obtain good mean-field like approximation for stochastic processes unfolding on sparse networks.

Table 5.1: Network models with degrees in the range $1 \leq k \leq 20$ for the truncated power laws and $k \in \{0, 1, 2, \ldots\}$ for the networks with Poisson degree distributions.

Network	Degree distribution	$\langle k \rangle$	$\langle k^2 \rangle$
Homogenous/regular	$P(4) = 1$	4	16
Bimodal	$P(2) = P(4) = 0.5$	3	10
Poisson	$P(k) = \langle k \rangle^k \frac{e^{-\langle k \rangle}}{k!}$	10	110
Truncated power law (a)	$P(k) = 0.673 k^{-2} e^{-k/30}$	2.0406	9.6613
Truncated power law (b)	$P(20 - k) = 0.673 k^{-2} e^{-k/30}$	17.9635	328.1197

5.3.1 Full versus reduced/closed ODEs

Here we show that solving the master equations, Eq. (KE), directly with the more general infection term, Eq. (5.10), gives good agreement with the solution of the closed/reduced system, Eqs. (5.11-5.12). In Fig. 5.1, we show that for a range of parameter values the agreement is excellent, and in line with what Shang found in [104], which simply means that the assumption of a binomial distribution for the number of infected individuals at a given time is a valid approximation. However, it does neither confirm nor invalidates the appropriateness of the choice of the new infection rate a_k, as proposed by Shang in [104]. Their appropriateness is tested via comparing the output from the master and / or reduced equations to the average of stochastic simulations and this is what we test next.

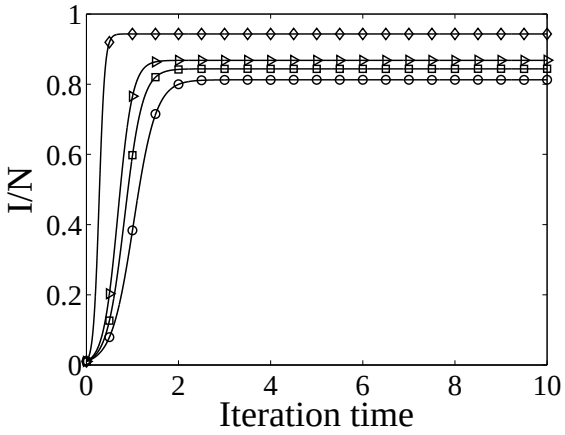


Figure 5.1: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Continuous lines represent the solution of the full equations, see Eq. (KE), while the solution of reduced model is given by Eqs. (5.11-5.12) for ($\square$) - homogeneous distribution $P(4) = 1$, ($\circ$) - bimodal distribution $P(2) = P(4) = 0.5$, ($\diamond$) - Poisson distribution with $\langle k \rangle = 10$, and ($\triangleright$) - truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$. For all cases there is excellent agreement between the full and reduced equations.

5.3.2 Comparison of Shang's generalisation to simulation

We first generate networks with the prescribed degree distribution by using the configuration method. This is followed by implementing the epidemic as a continuous-time Markov Chain on these networks. This is done by using a Gillespie-type approach [42, 43]. In this case, inter event times are chosen from an exponential distribution with a rate given by the sum of the rates of all possible events, followed by the choice of an event at random but proportionally to its rate.

We now move on to the crucial comparison of output based on the closed system to results from explicit stochastic network simulations. First, we validate our own simulations for the range of networks suggested by Shang in [104], see Table 5.1 for a summary. We use the pairwise [54], see Appendix 5.5.1, and effective-degree models [70], see Appendix 5.5.2, and as shown in Figs. 5.2 and 5.3, the agreement with our simulations is excellent. As pointed out before, the small disagreements are due to the very small average degree of the networks used in [104]. A small average degree is well-known to make the approximation with mean-field type models difficult. The

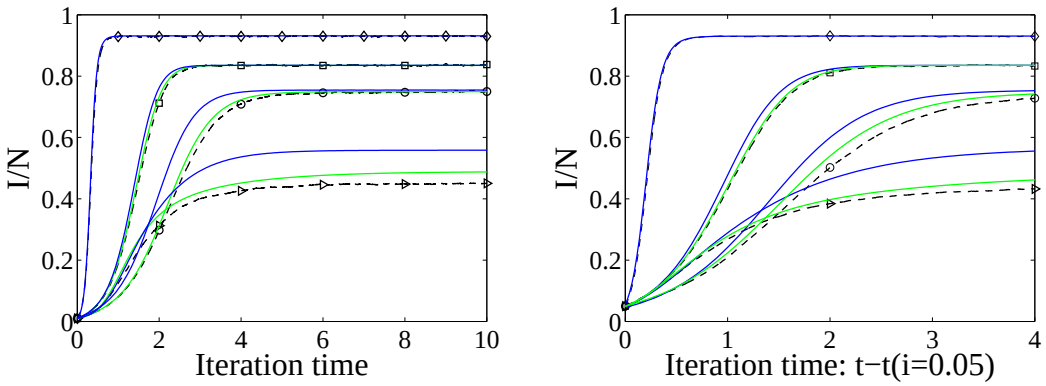


Figure 5.2: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these: homogeneous distribution $P(4) = 1$ ($\square$), bimodal distribution $P(2) = P(4) = 0.5$ ($\circ$), Poisson distribution with $\langle k \rangle = 10$ ($\diamond$) and truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$) (simulation: black dashed line, effective degree model: green line, compact pairwise model: blue line). We note that the effective degree model has not been implemented for networks with Poisson distribution due to the degrees being theoretically unbounded.

same figures show that the agreement improves as the average degree increases, see the case of networks with homogeneous and heterogeneous degree distributions with $\langle k \rangle = k = 4$ and $\langle k \rangle = 10$, respectively.

In Figs. 5.4 and 5.5, we plot the prevalence based on Shang's closed model, Eqs. (5.11-5.12), versus that from simulations. These plots show clearly that the agreement is poor, except for heterogenous networks with relatively large average degree and for networks with the inverted truncated power law distribution with very high degree as shown in Fig. 5.5. Our tests significantly differ from Shang's results and we infer that Shang's simulation method, which is not described in [104], is flawed or incorrectly implemented. We point out that the results concerning the full master equation and its reduction are correct and we were able to reproduce these. However, this alone neither leads to nor guarantees agreement with results based on simulations. In all our tests, and in line with Shang's work, we also attempted to time shift the prevalence, see the right panel in Fig. 5.4, but this did not lead to better agreement. Moreover, a close visual inspection shows clearly that there are fundamental differences between Shang's

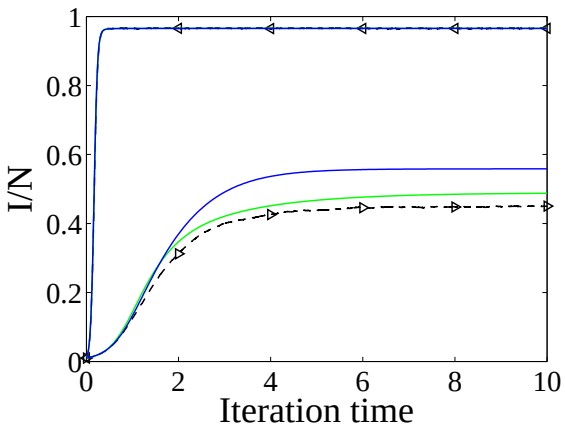


Figure 5.3: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. The networks have truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$) and degree inverted distribution ($\triangleleft$), i.e. $P(20-k) = 0.673k^{-2}\exp(-k/30)$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these (simulation: black dashed line, effective degree model: gree line, compact pairwise model: blue line).

closed model and simulation results and that no amount of time shifting will lead to a better agreement. For example, the equilibrium prevalence is very different and this again is in stark disagreement with Shang's results.

5.4 Discussion

It is our view that identifying general infectious terms a_k remains a major challenge as this is highly dependent on the structure of the network, parameters of the disease dynamics, and more importantly on the correlations that build up during the spreading process. It is unfortunate that this generalisation does not work and, as we showed in [84], it is possible to try and derive semi-analytical or numerical approximations for the infection rates. We conclude that Shang's simulation method is flawed and that Shang's generalisation is not valid. We look forward to any clarifications.

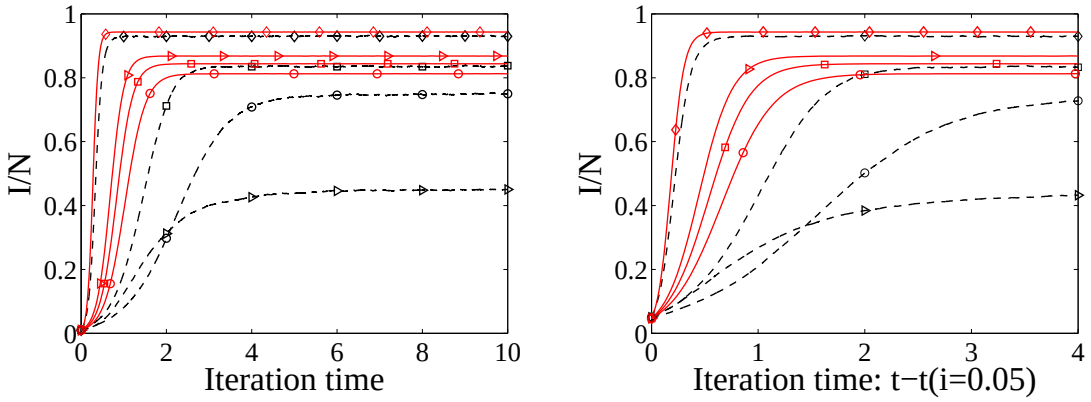


Figure 5.4: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these: homogeneous distribution $P(4) = 1$ ($\square$), bimodal distribution $P(2) = P(4) = 0.5$ ($\circ$), Poisson distribution with $\langle k \rangle = 10$ ($\diamond$) and truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$). Simulations are black dashed lines and results based on Shang's model, see Eqs. (5.11-5.12), are given by the red lines.

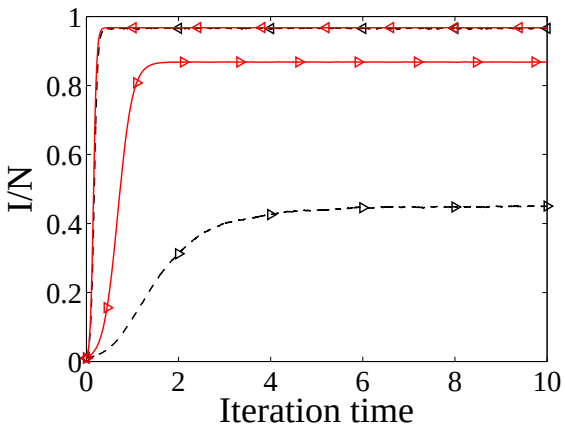


Figure 5.5: Time evolution of the fraction infected (I/N) based on networks with $N = 1000$ nodes, $I_0 = 10$ initial infectious nodes chosen at random, $\gamma = 1$ and $\tau = 1.6$. The networks have truncated power law distribution $P(k) = 0.673k^{-2}\exp(-k/30)$ for $1 \leq k \leq 20$ ($\triangleright$) and degree inverted distribution ($\triangleleft$), i.e. $P(20-k) = 0.673k^{-2}\exp(-k/30)$. Simulations are averaged over 20 different network realisations and 20 simulations on each of these. Simulations are black dashed lines and results based on Shang's model, see Eqs. (5.11-5.12), are given by red lines.

Acknowledgements

P. Rattana acknowledges funding for her Ph.D. studies from the Ministry of Science and Technology, Thailand.

5.5 Appendices

5.5.1 Appendix A: Compact pairwise model

House and Keeling [54] have successfully extended the general pairwise model of Eames and Kelling [33] to heterogeneous networks and for both SIR and SIS models. The reduced/compact pairwise SIS model is given by:

$$[\dot{S}_k] = \gamma([k] - [S_k]) - \tau[SI]\frac{k[S_k]}{\sum_l l[S_l]},$$

$$[\dot{SI}] = \tau[SI]\Big(\sum_k k[S_k] - 2[SI]\Big)\frac{\sum_l l(l-1)[S_l]}{(\sum_m m[S_m])^2} - (\tau+\gamma)[SI]$$

$$+\gamma\Big(\sum_k k([k] - [S_k]) - [SI]\Big),$$

where $[k]$ is the number of nodes of degree k. This system results from the standard pairwise model of Eames and Kelling [33] by using the following more compact closure

$$[A_k B] \approx [AB]\frac{k[A_k]}{\sum_l l[A_l]}.$$

We note that $[A_k]$ stands for the expected number of nodes of degree k across the whole network in state A, $[A_k B] = \sum_l [A_k B_l]$, where $[A_k B_l]$ represents the number of links of type $A - B$ when A has degree k and B has degree l. τ is the transmission rate and γ is the recovery rate.

5.5.2 Appendix B: Effective degree model

Lindquist et al. [70] formulated the SIS mean-field model base on the effective degree approach. This model is based on keeping track of the expected number of susceptible and infected nodes with all possible neighbourhood combinations, S_{si} and I_{si}, respectively. S_{si} represents the expected number of susceptible nodes that have s connections to other susceptible nodes and i connections to infected nodes, with similar argument for I_{si}.

Accounting for all possible transitions, the equations as formulated by Lindquist et

al. [70] are:

$$\acute{S_{si}} = -\tau i S_{si} + \gamma I_{si} + \gamma \Big[(i+1) S_{s-1,i+1} - i S_{si} \Big]$$

$$+ \frac{\sum_{k=1}^{M} \sum_{j+l=k} \tau j l S_{jl}}{\sum_{k=1}^{M} \sum_{j+l=k} j S_{jl}} \Big[(s+1) S_{s+1,i-1} - s S_{si} \Big],$$

$$\acute{I_{si}} = \tau i S_{si} - \gamma I_{si} + \gamma \Big[(i+1) I_{s-1,i+1} - i I_{si} \Big]$$

$$+ \frac{\sum_{k=1}^{M} \sum_{j+l=k} \tau l^2 S_{jl}}{\sum_{k=1}^{M} \sum_{j+l=k} j I_{jl}} \Big[(s+1) I_{s+1,i-1} - s I_{si} \Big],$$

for $\{(s,i) : s \geq 0, i \geq 0, s+i \leq M\}$, where M is the maximum node degree in the network.

Chapter 6

Discussion

This thesis presented work from the discipline of mathematical epidemiology and focused on modelling the spread of disease on networks, specifically for weighted and dynamic networks. In this final chapter, we conclude with a discussion of some of the results and present further extensions or future research ideas, as and when appropriate.

In Chapter 2, with the research paper titled "A Class of Pairwise Models for Epidemic Dynamics on Weighted Networks", we focused mainly on SIS and SIR epidemic models. These processes were run on weighted networks using pairwise approximation models [58, 94] and comparisons were made against individual-based network simulations. To evaluate the impact of different weight distributions on epidemic thresholds and dynamics, we investigated a simple weighted model where edges have random weights on an undirected, homogeneous network. For the SIR model, the basic reproductive ratio R_0 is derived based on both the network and pairwise models, by using the next generation matrix approach [5] and by using the approach introduced by Keeling [58] and Eames [32], respectively. The result of the study has shown an excellent agreement between simulation and pairwise models. The agreement remains valid for both SIS and SIR dynamics. Disagreement only occurs for extreme weight distributions, and we hypothesise that this is mainly due to the network becoming more modular with islands of nodes connected by links of low weight being bridged together by highly-weighted links. An analysis of R_0 for different weight distributions has illustrated that more heterogeneity across the weights leads to lower R_0, where average weight is constant. Further extensions of this study may consider the analysis of correlations between link weight and node degree. This direction has already been explored in the context of classic compartmental mean-field models based on node degree [55, 88]. Given that pairwise models extend to heterogeneous networks, such

an avenue can be further explored to include different types of correlations or other network-dependent weight distributions. Another theoretically interesting and practically important aspect is the consideration of different types of time delays, representing latency or temporary immunity [13], and the analysis of their effects on the dynamics of epidemics on weighted networks.

The simplest extensions to the distribution of weights are considered in Chapter 3 with the research paper titled "Pairwise and Edge-based Models of Epidemic Dynamics on Correlated Weighted Networks". Namely, we looked at SIR disease dynamics on heterogeneous weighted networks where the weights are randomly distributed and dependent on nodal degree, revealing the impact of different weight distributions and the correlations between link-weight and degree on epidemic dynamics. Our pairwise model in Chapter 3 [95] and edge-based compartmental model [78, 79], as well as simulations, are simultaneously developed and analysed. In this work, we assume that the link weight is inversely proportional to the degrees of the nodes that it connects. This model has been compared to two null models where for both the network topology remains the same and only the distribution of the weights changes. First, we considered the case where the original weights are 'lifted off' the edges and redistributed at random, thus making weights and nodal degrees independent, and secondly, we considered networks with all weights equal. The numerical results describing the evolution of the disease show remarkable agreement between the pairwise, edge-based compartmental and simulation models for all cases considered. The results show that the negative correlation between weights and nodal degrees can negate the important role played by highly connected nodes in standard epidemic models on non-weighted graphs, and that weight heterogeneity but with the same overall average or total weight, reduces the value of R_0. We furthermore measured the early growth rate, final epidemic size and R_0. The relation between final epidemic size and R_0 is determined by the model structure and, in our case study, the same R_0 value leads to the biggest final epidemic size on degree-dependent weighted networks. Finally, we illustrate that two seemingly different modelling approaches, namely the pairwise and the edge-based compartmental models, operate on similar assumptions and it is possible to formally link the two. Future work may now focus [77] on presenting detailed arguments to show the relationship between these models and other models for SIR epidemics on networks. We believe that in a model 'rich' environment, this part of our study and future work, as well as of others in the community [54], are important in trying to reconcile as many different modelling

approaches as possible and to identify model hierarchies, as well as to pinpoint model efficiencies in terms of generating analytical or semi-analytical results. Our work on the edge-based compartmental model focuses purely on SIR dynamics, since there is no equivalent for SIS dynamics yet. Future work should therefore aim at testing if edge-based compartmental models can be extended to SIS dynamics. If this turns out to be possible, it may lead to a model which is more amenable to deriving rigorous analytical results.

The papers associated with Chapters 2 and 3 focus on weighted and static networks. This led us to the interesting problem of increasing model realism by considering dynamic networks, where links change over time. In Chapter 4 with the research paper titled "Impact of constrained rewiring on network structure and node dynamics", we consider epidemic dynamics on dynamic networks. We explore the effect of spatially constrained rewiring on an SIS epidemic unfolding on an adaptive network. Specifically, the dynamics of the network is achieved by the assumption that susceptible nodes break links with infected nodes independently of distance, and reconnect at random to susceptible nodes available within a given radius R. Here, we assumed that nodes are placed at random on a square of length L with periodic boundary conditions and unit density. Two different starting networks were used and analysed, namely homogeneous and heterogeneous Erdős-Rényi networks. Following Gross et al. [45] a step-by-step approach is taken to investigate the dynamics of the network structure and disease dynamics on the network itself. We began by studying network dynamics in the absence of disease dynamics, followed by looking at the dynamics where there is dependence on individual statuses but these statuses do not change over time, and finally we study a coupling of both network dynamics and disease dynamics. In all models, a range of radii R, giving circular neighbourhoods within which to rewire, is considered and shown to provide the means to control epidemic outbreaks. We are able to give analytic and semi-analytic formulas for the value of clustering achieved in the network. These showed excellent agreement with simulations and we have revealed that it is possible to generate networks with the same mean path length and the same clustering but significantly different distribution of real link lengths. This needs further investigation, possibly using more complex node dynamics to reveal how subtle differences in the network structure may impact on the outcome of dynamical processes supported by the network. Further results provided analytical formulas for the degree distributions of susceptible and infected nodes which again showed good agreement with simulation

results. We were also able to show that the resulting networks, in certain regimes, are equivalent to the well-known random geometric graphs. Finally, we have shown that even constrained rewiring can serve as a potent control measure. We highlighted that the expected number of nodes in a typical local area is a key parameter which influences the network dynamics and can determine whether a disease dies out or becomes endemic. Extensions to the methodology presented in this study include considering other forms of constrained rewiring, e.g., network models where locality is not just defined in terms of spatial distance but possibly some more abstract, general metric, or community, and understanding how this impacts on the emerging network structure and epidemic or processes other than epidemics. Moreover, future work should consider the rewiring process on various networks, such as scale-free networks which are closer to the degree distributions resulting from some more realistic networks.

In Chapter 5, with the research paper titled "Comment on "A BINOMIAL MOMENT APPROXIMATION SCHEME FOR EPIDEMIC SPREADING IN NETWORKS" in U.P.B. Sci. Bull., Series A, Vol. 76, Iss. 2, 2014 ", we provide an extensive test and comment on a generalisation proposed by Shang in [104]. Shang presented a binomial moment approximation model for the study of SIS dynamics on networks with a variety of degree distributions, and it was claimed that numerical results were a good approximation for epidemics spreading on a range of configuration model networks [104]. However, our tests show that the proposed generalisation performs poorly for all networks proposed by Shang, except for heterogenous networks with high average degree. To support this statement, we also validated our simulation results by using the well-known pairwise [54] and effective degree models [70]. We conclude that Shang's simulation method is flawed and that Shang's generalisation is not valid.

Although both SIS and SIR dynamics are studied in this thesis, any future research should be based around understanding how the SIS dynamics can be best approximated using mean-field type models. Whilst these issues are well understood for an SIR epidemic, there remain many open questions for SIS disease dynamics. A potential good start could be to compare the performance of models such as (i) the compact pairwise model [54, 108], (ii) the effective degree model [70], (iii) the individual-based model proposed by Mieghem et al. [74], and (iv) the edge-based compartmental model [78] in approximating results based on individual-based stochastic network simulations. By investigating all these candidate models and quantifying their agreement with simulations, we could gain a better understanding of when and how disagreements arise and,

this in turn, may for example help to understand whether edge-based compartmental models can be extended to SIS dynamics. Further important extensions can be made for dynamic networks, where oscillations predicted by mean-field models are notoriously difficult to match by simulations. This is mainly due to the fact that the average of many individual-based stochastic network simulations can mask the true oscillatory behaviour. However, as shown in this thesis, model extensions have to be made with caution, as accounting for more complexity usually results in more complicated models which are more difficult to analyse. This can in turn then mask and preclude a deeper understanding.

Bibliography

[1] L. J. Abu-Raddad, P. Patnaik, and J. G. Kublin. Dual infection with HIV and Malaria Fuels the Spread of Both Diseases in Sub-Saharan Africa. *Science*, 314:1603–1606, 2006.

[2] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[3] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford, 1991.

[4] K. B. Athreya and P. E. Ney. *Branching Processes*. Dover Publications Inc. Mineola, New York, 2008.

[5] F. Ball and P. Neal. Network epidemic models with two levels of mixing. *Math. Biosci.*, 212:69–87, 2008.

[6] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA.*, 101:3747–3752, 2004.

[7] A. Barrat, M. Barthélemy, and A. Vespignani. Modeling the evolution of weighted networks. *Phys. Rev. E*, 70:066149, 2004.

[8] A. Barrat, M. Barthélemy, and A. Vespignani. Weighted Evolving Networks: Coupling Topology and Weight Dynamics. *Phys. Rev. Lett.*, 92:228701, 2004.

[9] A. Barrat, M. Barthélemy, and A. Vespignani. The effects of spatial constraints on the evolution of weighted complex networks. *J. Stat. Mech.*, page P05003, 2005.

[10] M. Barthélemy. Spatial networks. *Physics Reports*, 499:1–101, 2011.

[11] P. Beutels, Z. Shkedy, M. Aerts, and P. Van Damme. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiol. Infect.*, 134:1158–1166, 2006.

[12] K. B. Blyuss and Y. N. Kyrychko. On a basic model of a two-disease epidemic. *Appl. Math. Comp.*, 160:177–187, 2005.

[13] K. B. Blyuss and Y. N. Kyrychko. Stability and bifurcations in an epidemic model with varying immunity period. *Bull. Math. Biol.*, 72:490–505, 2010.

[14] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Phys. Rep.*, 424:175–308, 2006.

[15] M. Boguńa, L. F. Lafuerza, R. Toral, and M. Á. Serrano. Simulating non-Markovian stochastic processes. *Phys. Rev. E.*, 90:042108, 2014.

[16] T. Britton, M. Deijfen, and F. Liljeros. A weighted configuration model and inhomogeneous epidemics. *J. Stat. Phys.*, 145:1368–1384, 2011.

[17] T. Britton and D. Lindenstrand. Inhomogeneous epidemics on weighted networks. *Math. Biosci.*, 260:124–131, 2012.

[18] Y. Bu, S. Gregory, and H. L. Mills. Efficient local behavioral-change strategies to reduce the spread of epidemics in networks. *Phys. Rev. E.*, 88:042801, 2013.

[19] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Network Robustness and Fragility: Percolation on Random graphs. *Phys. Rev. Lett.*, 85:54685471, 2000.

[20] R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91:247901(4), 2003.

[21] V. Colizza, A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani. Modelling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Med.*, 4:95–110, 2007.

[22] B. S. Cooper, R. J. Pitman, W. J. Edmunds, and N. J. Gay. Delaying the international spread of pandemic influenza. *PLoS Med.*, 3:e212, 2006.

[23] J. Dall and M. Christensen. Random geometrix graphs. *Phys. Rev. E.*, 66:016121, 2002.

[24] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011:284909, 2010.

[25] M. Deijfen. Epidemics and vaccination on weighted graphs. *Math. Biosci.*, 232:57–65, 2011.

[26] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Chichester: Wiley, 2000.

[27] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R_0, in models for infectious diseases in heterogeneous populations. *J. Math. Biol.*, 28:365–382, 1990.

[28] K. Dietz and J. A. P. Heesterbeek. Daniel Bernoullis epidemiological model revisited. *Mathematical Biosciences*, 180:1–21, 2002.

[29] S. N. Dorogstev and J. F. F. Mendes. Evolution of networks. *Adv. Phys*, 51:1079, 2002.

[30] S. N. Dorogstev and J. F. F. Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press, Oxford, 2003.

[31] B. Dybiec, A. Kleczkowshi, and C. A. Gilligan. Controlling disease spread on networks with incomplete knowledge. *Phys. Rev. E.*, 70:066145, 2004.

[32] K. T. D. Eames. Modelling disease spread through random and regular contacts in clustered populations. *Theor. Popul. Biol.*, 73:104–111, 2008.

[33] K. T. D. Eames and M. J. Keeling. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *Proc. Natl. Acad. Sci. USA.*, 99:13330–13335, 2002.

[34] K. T. D. Eames, J. M. Read, and W. J. Edmunds. Epidemic prediction and control in weighted networks. *Epidemics.*, 1:70–76, 2009.

[35] W. J. Edmunds, C. J. O'Callaghan, and D. J. Nokes. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc. R. Soc. Lond. B*, 264:949–957, 1997.

[36] P. Erdös and A. Rényi. ON THE EVOLUTION OF RANDOM GRAPHS. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.

[37] S. Eubank, H. Guclu, V. S. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreak in realistic urban social networks. *Nature*, 429:180–184, 2004.

[38] S. Funk, E. Gilad, C. Watkins, and V. A. A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *PNAS*, 106:6872, 2009.

[39] D. Garlaschelli. The weighted random graph model. *New J. Phys.*, 11:073005, 2009.

[40] G. P. Garnett and R. M. Anderson. Sexually Transmitted Diseases and Sexual Behavior: Insights from Mathematical Models. *J. Infect. Dis.*, 174:S150S161, 1996.

[41] M. Gilbert, A. Mitchell, D. Bourn, J. Mawdsley, R. Clifton-Hadley, and W. Wint. Cattle movements and bovine tuberculosis in Great Britain. *Nature.*, 435:491–496, 2005.

[42] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[43] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

[44] T. Gross and B. Blasius. Adaptive coevolutionary networks: a review. *J. R. Soc. Interface*, 5:259–271, 2008.

[45] T. Gross, C. D. J. D'Lima, and B. Blasius. Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.*, 96:208701, 2006.

[46] T. Gross and H. Sayama. *Adaptive Networks: Theory, Models and Applications*. Springer Verlag Heidelberg, 2009.

[47] V. Hatzopoulos, M. Taylor, P. L. Simon, and I. Z. Kiss. Multiple sources and routes of information transmission: Implications for epidemic dynamics. *Math. Biosci.*, 231:197–209, 2011.

[48] H. W. Hethcote and J. A. Yorke. *Gonorrhea Transmission Dynamics and Control*. Springer Lecture Notes in Biomathematics. Springer, Berlin, 1984.

[49] T. D. Hollingsworth, N. M. Ferguson, and R. M. Anderson. Will travel restrictions control the international spread of pandemic infuenza? *Nat. Med.*, 12:497–499, 2006.

[50] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Phys. Rev. E.*, 65:056109, 2002.

[51] D. R. Hopkins. *The Greatest Killer: Smallpox in History*. The University of Chicago Press, 2002.

[52] T. House, G. Davies, L. Danon, and M. J. Keeling. A motif-based approach to network epidemics. *Bull. Math. Biol.*, 71:1693–1706, 2009.

[53] T. House and M. J. Keeling. The impact of contact tracing in clustered populations. *PLoS Comput. Biol.*, 6:e1000721, 2010.

[54] T. House and M. J. Keeling. Insights from unifying modern approximations to infections on networks. *J. Roy. Soc. Interface*, 8:67–73, 2011.

[55] J. Joo and J. L. Lebowitz. Behavior of susceptible-infected-susceptible epidemics on heterogeneous networks with saturation. *Phys. Rev. E.*, 69:066105, 2004.

[56] S. Karlin and H. M. Taylor. *A First Course in Stochastic Processes*. Academic Press, ISBN 0-12-398552-8, 1975.

[57] S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, ISBN 0-12-398650-8, 1981.

[58] M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B*, 266:859–867, 1999.

[59] M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *J. Roy. Soc. Interface*, 2:295–307, 2005.

[60] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2007.

[61] M. J. Keeling, M. E. J. Woolhouse, R. M. May, G. Davies, and B. T. Grenfellk. Modelling vaccination strategies against foot-and-mouth disease. *Nature*, 421:136–142, 2003.

[62] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond A.*, 115:700–721, 1927.

[63] I. Z. Kiss, L. Berthouze, T. J. Taylor, and P. L. Simon. Modelling approaches for simple dynamic networks and applications to disease transmission models. *Proc. R. Soc. A*, 468(2141):1332–1355, 2012.

[64] I. Z. Kiss, J. Cassell, M. Recker, and P. L. Simon. The impact of information transmission on epidemic outbreaks. *Math. Biosci.*, 225:1–10, 2010.

[65] I. Z. Kiss, D. M. Green, and R. R. Kao. The effect of contact heterogeneity and multiple routes of transmission on final epidemic size. *Math. Biosci.*, 203:124–136, 2006.

[66] I. Z. Kiss and P. L. Simon. New moment closures based on a priori distributions with applications to epidemic dynamics. *Bull. Math. Biol.*, 74:1501–1515, 2012.

[67] C. Lagorio, M. Dickison, F. Vazquez, L. A. Braunstein, P. A. Macri, M. V. Migueles, S. Havlin, and H. E. Stanley. Quarantine-generated phase transition in epidemic spreading. *Phys. Rev. E.*, 83:026102, 2011.

[68] R. Lambiotte, L. Tabourier, and J.-C. Delvenne. Burstiness and spreading on temporal networks. *Eur. Phys. J. B*, 86:1–4, 2013.

[69] C. Li and G. Chen. A comprehensive weighted evolving network model. *Physica A*, 343:288–294, 2004.

[70] J. Lindquist, J. Ma, P. van den Driessche, and F. H. Willeboordse. Effective degree network disease models. *J. Math. Biol.*, 62:143–164, 2011.

[71] R. J. Littman and M. L. Littman. Galen and the Antonine Plague. *Am. J. Philology*, 94(3):243–255, 1973.

[72] V. Marceau, P-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé. Adaptive networks: Coevolution of disease and topology. *Phys. Rev. E.*, 82:036116, 2010.

[73] R. M. May and A. L. Lloyd. Infection dynamics on scale-free networks. *Phys. Rev. E.*, 64:066112, 2001.

[74] P. V. Mieghem, J. Omix, and R. Kooij. Virus Spread in Networks. *IEEE/ACM*, 17:1–14, 2009.

[75] J. C. Miller. Spread of infectious disease through clustered populations. *J. Roy. Soc. Interface.*, 6:1121–1134, 2009.

[76] J. C. Miller. Epidemics on networks with large initial conditions or changing structure. *PLoS ONE*, 9(7):e101421, 2014.

[77] J. C. Miller and I. Z. Kiss. Epidemic Spread in Networks: Existing Methods and Current Challenges. *Math. Model. Nat. Phenom.*, 9(2):4–42, 2014.

[78] J. C. Miller, A. C. Slim, and E. M. Volz. Edge-based compartmental modelling for infectious disease spread. *J. Roy. Soc. Interface.*, 9:890–906, 2012.

[79] J. C. Miller and E. M. Volz. Edge-based compartmental modeling with disease and population structure. *Available at http://arxiv.org/abs/1106.6344*.

[80] C. E. Mills, J. M. Robins, and M. Lipsitch. Transmissibility of 1918 pandemic influenza. *Nature*, 432:904–906, 2004.

[81] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Struct Alg.*, 6:161–180, 1995.

[82] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, 26:521–529, 2002.

[83] D. M. Morens and R. J. Littman. Epidemiology of the Plague of Athens. *Transactions of the American Philological Association*, 122:271–304, 1992.

[84] N. Nagy, I. Z. Kiss, and P. L. Simon. Approximate master equations for dynamical processes on graphs. *Math. Model. Nat. Phenom.*, 9:43–57, 2014.

[85] M. Nekovee. Worm epidemics in wireless as hoc networks. *New Journal of Physics*, 9:189, 2007.

[86] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E.*, 66:016128, 2002.

[87] J. R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[88] R. Olinky and L. Stone. Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmission. *Phys. Rev. E.*, 70:030902(R), 2004.

[89] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Review of Modern Physics, arXiv preprint, arXiv:1408.2701*, 2014.

[90] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Phys. Rev. E*, 63:066117, 2001.

[91] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3202, 2001.

[92] D. L. Pendell, J. Leatherman, T. C. Schroeder, and G. S. Alward. The Economic Impacts of a Foot-And-Mouth Disease Outbreak: A Regional Analysis. *J. of Agricultural and Applied Economics*, 39:19–33, 2007.

[93] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[94] D. A. Rand. Correlation equations and pair approximations for spatial ecologies. *CWI Quarterly*, 12(3&4):329–368, 1999.

[95] P. Rattana, K. B. Blyuss, K. T. D. Eames, and I. Z. Kiss. A class of pairwise models for epidemic dynamics on weighted networks. *Bull. Math. Biol.*, 75(3):466–490, 2013.

[96] J. M. Read, K. T. D. Eames, and W. J. Edmunds. Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface*, 5:1001–1007, 2008.

[97] F. P. Retief and L. Cilliers. EPIDEMICS OF THE ROMAN EMPIRE, 27 BC - AD 476. *SAMJ*, 90:267–272, 2000.

[98] T. Rhodesa, M. Singerb, P. Bourgoisc, S. R. Friedmand, and S. A. Strathdee. The social structural production of HIV risk among injecting drug users. *Social Science & Medicine*, 61:1026–1044, 2005.

[99] S. Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316:1298–1301, 2007.

[100] S. Riley and N. M. Ferguson. Smallpox transmission and control: spatial dynamics in Great Britain. *Proc. Natl. Acad. Sci.*, 103:12637–12642, 2006.

[101] S. Risau-Gusman and D. H. Zanette. Contact switching as a control strategy for epidemic outbreaks. *J. Theor. Biol.*, 257:52, 2009.

[102] M. Ritchie, L. Berthouze, T. House, and I. Z. Kiss. Higher-order structure and epidemic dynamics in clustered networks. *J. Theor. Biol.*, 348:21–32, 2014.

[103] J. Saramaki and K. Kaski. Modelling development of epidemics with dynamic small-world networks. *J. Theor. Biol.*, 234:413, 2005.

[104] Y. Shang. A binomial moment approximation scheme for epidemic spreading in networks. *U.P.B. Sci. Bull. Series A*, 76:23–30, 2014.

[105] K. J. Sharkey, C. Fernandez, K. L. Morgan, E. Peeler, M. Thrush, J. F. Turnbull, and R. G. Bowers. Pair-level approximations to the spatio-temporal dynamics of epidemics on asymmetric contact networks. *J. Math. Biol.*, 53:61–85, 2006.

[106] L. B. Shaw and I. B. Schwartz. Fluctuating epidemics on adaptive networks. *Phys. Rev. E.*, 77:066101, 2008.

[107] P. Slack. The black death past and present. 2. Some historical problems. *Trans R Soc Trop Med Hyg*, 83:461–463, 1989.

[108] E. Volz. SIR dynamics in random networks with heterogeneous connectivity. *J. Math. Biol.*, 56:293–310, 2008.

[109] E. M. Volz, J. C. Miller, A. Galvani, and L. Ancel Meyers. Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics. *PLoS Comput Biol*, 7:e1002042, 2011.

[110] S. Wang and C. Zhang. Weighted competition scale-free network. *Phys. Rev. E.*, 70:066127, 2004.

[111] H. S. Wilf. *Generatingfunctionology*. Academic Press, Boston, Mass, USA, 1994.

[112] R. Yang, T. Zhou, Y.-B. Xie, Y.-C. Lai, and B.-H. Wang. Optimal contact process on complex networks. *Phys. Rev. E*, 78:066109, 2008.

[113] Z. Yang and T. Zhou. Epidemic spreading in weighted networks: An edge-based mean-field solution. *Phys. Rev. E*, 85:056106, 2012.

[114] S. Yu-Rong, J. Guo-Ping, and G. Yong-Wang. Epidemic propagation on adaptive coevolutionary networks with preferential local-world reconnection strategy. *Chin. Phys. B.*, 22:040205, 2013.