Journal of Mathematical Biology manuscript No.
(will be inserted by the editor)

Interdependency and hierarchy of exact and approximate epidemic models on networks

Timothy J Taylor · **Istvan Z Kiss**

Received: date / Accepted: date

Abstract Over the years numerous models of *SIS* (susceptible → infected → susceptible) disease dynamics unfolding on networks have been proposed. Here, we discuss the links between many of these models and how they can be viewed as more general motif-based models. We illustrate how the different models can be derived from one another and, where this is not possible, discuss extensions to established models that enables this derivation. We also derive a general result for the exact differential equations for the expected number of an arbitrary motif directly from the Kolmogorov/master equations and conclude with a comparison of the performance of the different closed systems of equations on networks of varying structure.

Keywords Network · Epidemic · Kolmogorov/master equations · exact models

Mathematics Subject Classification (2000) 92D25 · 92D30 · 92D40 · 00A71 · 00A72

1 Introduction

Modeling the spread of infectious diseases requires an understanding of not only disease characteristics but also an understanding of the community (be it a hospital, school, town, etc) in which it pervades. An important consideration in modelling

Timothy Taylor is funded by a PGR studentship from the MRC, and the Departments of Informatics and Mathematics at the University of Sussex.

Timothy J Taylor
Centre for Computational Neuroscience and Robotics, University of Sussex, Falmer, Brighton BN1 9QH, UK

Istvan Z Kiss (corresponding author)
School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9QH, UK
E-mail: I.Z.Kiss@sussex.ac.uk

the spread of diseases is thus the contact structure on which disease transmission happens. Whereas traditional approaches ([2, 5]) assume little or no topological structure, recent work ([13–15]) has tried to incorportate the underlying linkages between entities in the population and study how these links facilitate the spread of the disease. For a continuous-time stochastic disease transmission model on an arbitrary network it is possible ([11]), to write down the relevant Kolmogorov/master equations and thus model it as a continuous time Markov chain that fully describes the movement between all possible system states. Unfortunately the complexity of the model comes from the size of the state space and the number of equations scales exponentially as a^N, where a is the number of different states a node can be in and N is the network size. One widely used resolution to this complexity is to create individual-based simulation models and investigate the system behaviour directly. Even though increasing computational power makes simulations an increasingly attractive proposition they lack analytic tractability. Whilst this is not always a hindrance, when the system displays a rich range of behaviour (e.g. oscillations, bistability) it may not be feasable to obtain a global overview of the effects of different parameter values and thus the more analytic approach is needed. For this reason, low-dimensional systems of differential equations ([13, 6, 15]) are sought provided that these can approximate the exact solution. By reducing the problem to a smaller system of equations it is easier to study the bifurcation structure of the model and gain a greater understanding of the full spectrum of behaviour. The challenge is then finding the set of equations that best approximate the solution of the Kolmogorov equations.

Given that here we focus on epidemic models, usually such models are formulated in terms of the expected values of the number of infected and/or susceptible individuals or some other motif in the network such as the expected number of infected and/or susceptible individuals of different degrees (the number of connections a node has). Such models range from classic meanfield [1] to pairwise [13], heterogenous pairwise [6], effective degree [15, 16] and individual-level models [20] to name a few. Whilst these models seem to use different approaches their derivation is based on the same conceptual framework, namely they begin by choosing a base-motif (e.g. a node, a link and the two nodes it connects, a node and all its links). These base-motifs are then used to formulate equations for the different possible states that they can achieve (e.g. for the expected number of motifs in different states or the probability that a specified motif in the network is in a certain state). These equations generally involve not only the base-motif itself, but larger or extended motifs of which they are usually part of. These larger motifs in turn depend on more complex motifs and a closure is needed in order to obtain a self-contained system of equations of reasonable size. Importantly the base-motif determines not only the complexity of the model (the larger the motif the greater the number of states it can be in) but also how much of the network topology can be captured. Interestingly differential equations for smaller motifs that are part of the base-motif should, in theory, be recoverable from the original differential equation. To this end the main focus of the paper is the consideration of various simple models of disease dynamics and the relations between them. We also consider which models are derivable directly (subject to a suitable closure) from the Kolmogorov/master equations and can thus be referred to as exact.

We begin in section 2 with an introduction of some of the more common approaches to modelling disease dynamics on networks, considering meanfield ([1]), pairwise ([13]), heterogeneous pairwise ([6]) and the effective degree ([15]) model formulations. In section 3 we formulate an exact version of the effective degree model and then illustrate how the pairwise model can then be recovered from this new set of equations. We are, however, unable to recover the heterogenous pairwise model from the exact effective degree and this motivates, in section 4, an extension of this which incorporates further network topology into the ODEs. From this extension we then show how it is then possible to recover the heterogeneous pairwise equations. Once the links between the models have been established, in section 5 we show how the unclosed version of the models can be derived directly from the Kolmogorov equations. This is done by proving that as long as the heuristic equations for any motif are written following a certain set of rules they will always be exact. We conclude, in section 6 with a brief comparison of the models and discuss under what circumstances they perform best, in the sense of being close to simulation results.

2 Models of disease dynamics

In this paper we focus on susceptible $\to$ infected $\to$ susceptible (*SIS*) disease dynamics on networks but note that all of the following models can be adapted for other disease (e.g. *SIR* and/or contact tracing) or non-disease (e.g. evolutionary [8]) dynamics. With this in mind we use τ as the per-link transmission rate between susceptible and infected nodes and γ as the recovery rate of an infected individual. Both infection and recovery are modelled as independent poisson processes. As a starting point we give a short summary of ODE-based models that are either exact or an approximation of the true dynamics resulting from the full system based on the Kolmogorov/master equations, where these are solvable, or based on simulation.

2.1 Pairwise and the resulting simple compartmental model

In order to focus on the underlying network of contacts, we introduce the pairwise model first ([13,19]). The main idea of this model is to develop the hierarchical dependence of lower order moments (e.g. expected number of susceptible $[S]$ and infected $[I]$ nodes) on higher ones (e.g. expected number of pairs with one susceptible and one infected node, $[SI]$) and to derive appropriate models that correctly account for these. As already suggested, the expected number of pairs will depend on larger motifs, in this case these being the expected number of triples denoted by $[ABC]$, where $A,B,C \in \{S,I\}$ and B is connected to A and C. Using this notation the equations governing the evolution of the disease dynamics at the level of singles and pairs are given by

$$\frac{d}{dt}[I] = -\gamma[I] + \tau[SI], \tag{1}$$

$$\frac{d}{dt}[SS] = -2\tau[ISS] + 2\gamma[IS], \tag{2}$$

$$\frac{d}{dt}[SI] = \tau([ISS] - [ISI] - [IS]) + \gamma([II] - [IS]), \tag{3}$$

$$\frac{d}{dt}[II] = 2\tau([ISI] + [IS]) - 2\gamma[II]. \tag{4}$$

Importantly we note that these equations are unclosed as no equations are given for the evolution of the triples. The standard closure (in the absence of clustering) makes the assumption that the status of pairs are statistically independent of one another and then

$$[ABC] \approx [AB](n-1)\frac{[BC]}{n[B]},$$

where n is the average degree of the network. When we use this closure we say we have closed “at the level of triples” . In order to derive the classic mean-field model a closure at the level of paris can be applied, namely, $[SI]$ can be approximated as

$$[SI] \approx n[S]\frac{[I]}{N}$$

and upon using Eq. (1), the classic mean-field model can be recovered

$$\frac{d}{dt}[I] = -\gamma[I] + \tau n[S]\frac{[I]}{N}, \tag{5}$$

where the widely used transmission rate from the compartmental model,[1], is $\beta = \tau n$.

It is also important to note that the unclosed equations above (Eqs. (1-4)) can be derived directly from the state-based Kolmogorov equations and for this reason we refer to these equations as exact. Whilst a proof for the exactness of these equations was given in [21], in section 5 we provide a more general proof that allows us to write down exact equations for, not just pairs, but any motif structure. We also note that an alternative approach was used by Sharkey in [20], to prove that the standard pairwise equations were exact for models with susceptible $\to$ infected $\to$ recovered (SIR) disease dynamics.

2.2 Heterogeneous pairwise model

Whilst the pairwise equations perform well in capturing disease dynamics on networks that are well described by their average degree, the closure assumption fails when greater heterogeneity is introduced. More precisely, whilst the pairwise equations above are exact for an arbitrary network before a closure, these do not guarantee that with the current choice of singles and pairs (i.e. $[S]$ could be further divided to account for heterogeneity in degree) a valid closure could be found for any network. Indeed, to account for greater heterogeneity Eames *et al.* [6] further developed the pairwise model by taking into account not just the state of nodes and pairs but also the degrees of the nodes. By using $[A^n]$ to represent expected number of nodes of type A with degree n and with similar notation for pairs and triples, they were able to formulate the following set of unclosed equations

$$\frac{d}{dt}[S^n] = \gamma[I^n] - \tau\sum_q [S^n I^q], \tag{6}$$

$$\frac{d}{dt}[I^n] = -\gamma[I^n] + \tau\sum_q [S^n I^q], \tag{7}$$

$$\frac{d}{dt}[S^n S^m] = -\tau\sum_q \left([S^n S^m I^q] + [I^q S^n S^m]\right) + \gamma\left([S^n I^m] + [I^n S^m]\right), \tag{8}$$

$$\frac{d}{dt}[S^n I^m] = \tau\sum_q \left([S^n S^m I^q] - [I^q S^n I^m]\right) - \tau[S^n I^m] - \gamma[S^n I^m] + \gamma[I^n I^m], \tag{9}$$

$$\frac{d}{dt}[I^n I^m] = \tau\sum_q \left([I^n S^m I^q] + [I^q S^n I^m]\right) + \tau[I^n S^m] + \tau[S^n I^m] - 2\gamma[I^n I^m]. \tag{10}$$

Again assuming the statistical independence of pairs and absence of clustering, Eames et. al, [6], suggest the following approximations of triples

$$[B^n C^m D^p] \approx [B^n C^m](m-1)\frac{[C^m D^p]}{m[C^m]}.$$

2.3 The effective degree model

In [15], Lindquist *et al.* introduced the effective degree model for *SIS* (and also *SIR*) dynamics on a network (an equivalent model formulation was also proposed by Marceau et al. [16]). In this model they consider not only the state of a node (S or I), but also the number of the immediate neighbours in the various potential states. This is done by writing the following set of equations for all the possible star-like motifs in the network where $S_{s,i}$ ($I_{s,i}$) represents the expected number of susceptible (infected) nodes with s susceptible and i infected neighbours,

$$\begin{aligned}\dot{S_{s,i}} =& -\tau i S_{s,i} + \gamma I_{s,i} + \gamma[(i+1)S_{s-1,i+1} - iS_{s,i}] \\ &+ \tau\frac{\sum_{k=1}^{M}\sum_{j+l=k} jlS_{j,l}}{\sum_{k=1}^{M}\sum_{j+l=k} jS_{j,l}}[(s+1)S_{s+1,i-1} - sS_{s,i}],\end{aligned} \tag{11}$$

$$\begin{aligned}\dot{I_{s,i}} =& \tau i S_{s,i} - \gamma I_{s,i} + \gamma[(i+1)I_{s-1,i+1} - iI_{s,i}] \\ &+ \tau\frac{\sum_{k=1}^{M}\sum_{j+l=k} l^2 S_{j,l}}{\sum_{k=1}^{M}\sum_{j+l=k} jI_{j,l}}[(s+1)I_{s+1,i-1} - sI_{s,i}],\end{aligned} \tag{12}$$

with $1 \leq s+i \leq M$, where M is the maximum degree and the equations are suitably adjusted on the boundaries. It is important to note that this model is not exact as a closure has been already applied. Namely the infection of a node's susceptible neighbours is based on a population-level approximation. To illustrate this more precisely we borrow the notation of the pairwise model and make two observations

$$\frac{\sum_{k=1}^{M}\sum_{j+l=k} jlS_{j,l}}{\sum_{k=1}^{M}\sum_{j+l=k} jS_{j,l}} = \frac{[ISS]}{[SS]},$$
$$\frac{\sum_{k=1}^{M}\sum_{j+l=k} l^2S_{j,l}}{\sum_{k=1}^{M}\sum_{j+l=k} jI_{j,l}} = \frac{\sum_{k=1}^{M}\sum_{j+l=k} l(l-1)S_{j,l}+lS_{j,l}}{\sum_{k=1}^{M}\sum_{j+l=k} jI_{j,l}} = \frac{[ISI]+[SI]}{[SI]} = \frac{[ISI]}{[SI]}+1.$$

These means that the infection pressure on the susceptible neighbours of the central node is equal to the population level average taken from all the possible star-like configurations rather then from the extended star structures that would account exactly for these infections.

3 Recovering the pairwise model from the effective degree

Whilst the pairwise and effective degree models seem different they are based on a similar approach. Both models work on approximating the evolution of different motifs in the network; individuals and links in the pairwise model and star-like structures in the effective degree. For both models, but more clearly for the pairwise, the models begin with a starting or base motif (e.g. nodes) for which an evolution equation is required. This will of course depend on an extended motif, typically the base motif extended by the addition of an extra node (e.g. pairs). This dependency on higher order motifs continues, for example, with pairs depending on triples, and then triples depending on quadruplets (four nodes connected by a line, i.e. $A-B-C-D$, or a star with a centre and three spokes, i.e. $A-\overline{B-C\quad D}$). Hence, the models only differ in the choice of the base motif and then potentially in the way in which the systems are closed to curtail the dependency on higher order motifs. Since, here we are mainly interested in exact models, that is before a closure is applied, we begin by conjecturing an exact version of the effective degree model and show how starting from this the exact pairwise model can be derived.

3.1 Exact effective degree

Based on the ideas presented above, we extend the star-like base motif to reveal the dependence on higher order motifs and conjecture that this unclosed version of the effective degree model is exact. We begin by introducing a variable to count the expected number of infecteds connected to a node's susceptible neighbours. This is done by introducing two new terms, $[ISS_{s',i'}]$ and $[ISI_{s',i'}]$. For the term $[ISI_{s',i'}]$ (and similarly for $[ISS_{s',i'}]$) the S in the middle is actually used to represent the susceptible neighbours of the central I from the motif with composition $I_{s'i'}$ (i.e. the I node with neighbourhood (s', i') is the centre of the star, while S is a susceptible spoke). The I (on the left-hand side), in turn, represents the infective neighbours of these susceptibles' and within this count, in the case of $[ISI_{s',i'}]$, we also include the originating central I. The exact effective degree model can then be written as

$$\begin{aligned} \dot{S_{s,i}} =& -\tau i S_{s,i} + \gamma I_{s,i} + \gamma[(i+1)S_{s-1,i+1} - iS_{s,i}] \\ &+ \tau\,[ISS_{s+1,i-1}] - \tau\,[ISS_{s,i}]\,, \end{aligned} \tag{13}$$

$$\begin{aligned} \dot{I_{s,i}} =& \tau i S_{s,i} - \gamma I_{s,i} + \gamma[(i+1)I_{s-1,i+1} - iI_{s,i}] \\ &+ \tau\,[ISI_{s+1,i-1}] - \tau\,[ISI_{s,i}]\,. \end{aligned} \tag{14}$$

Fig. 1 shows the possible transitions captured by this model.

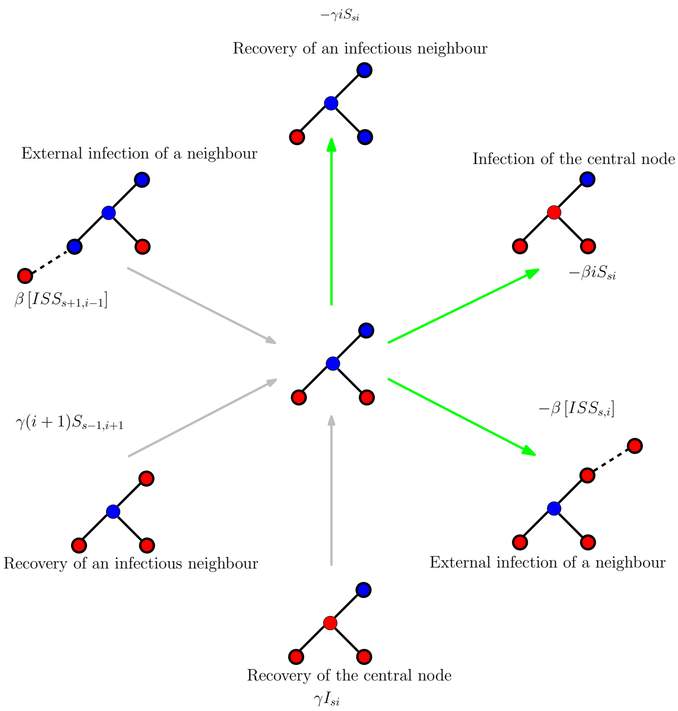


Fig. 1 Illustration of the transitions into and out of the $S_{2,1}$ class. Susceptible nodes are given in blue and infective nodes in red. Transitions into the class are shown in grey and transitions out in green. The corresponding terms of the general equation are also given.

3.2 Recovering the pairwise equations

The star-like composition of the effective degree model allows us to recover the pairwise equations via careful summations. The full derivation of the pairwise model is given in Appendix 1, whilst here we only illustrate the derivation of the individuals (trivial but given for completeness) and the $[II]$ pairs,

$$\frac{d}{dt}[S] = \sum_{s,i} \dot{S_{s,i}} = \gamma[I] - \tau[SI],$$
$$\frac{d}{dt}[I] = \sum_{s,i} \dot{I_{s,i}} = -\gamma[I] + \tau[SI],$$

where most terms from the original effective degree equations cancel and we have used that $\sum_{s,i} iS_{s,i} = [SI]$ and $\sum_{s,i} I_{s,i} = [I]$. For $[II]$ the following equality holds

$$\begin{aligned}
\frac{d}{dt}[II] =& \sum_{s,i} i\dot{I_{s,i}} \\
=& \tau \sum i^2 S_{s,i} - \gamma \sum iI_{s,i} + \gamma \sum i(i+1)I_{s-1,i+1} - \gamma \sum i^2 I_{s,i} \\
&+ \tau \sum i[ISI_{s+1,i-1}] - \tau \sum i[ISI_{s,i}] \\
=& \tau \sum i(i-1)S_{s,i} + \tau \sum iS_{s,i} - \gamma[II] \\
&+ \gamma[III] - \gamma \sum i(i-1)I_{s,i} - \gamma \sum iI_{s,i} \\
&+ \tau \sum (i-1)[ISI_{s+1,i-1}] + \tau \sum [ISI_{s+1,i-1}] - \tau \sum i[ISI_{s,i}] \\
=& \tau[ISI] + \tau[IS] - \gamma[II] + \gamma[III] - \gamma[III] - \gamma[II] + \tau[ISI] + \tau[IS] \\
=& 2\tau([ISI] + [IS]) - 2\gamma[II],
\end{aligned}$$

where we have used that $\sum_{s,i} iI_{si} = [II]$, $\sum_{s,i}(i-1)[ISI_{s+1,i-1}] = \sum i[ISI_{s,i}]$ and that $\sum [ISI_{s+1,i-1}] = [ISI] + [SI]$. These all follow from the definition of the pairwise model and the definition of the new extended motifs from the exact effective degree model. We note that this result does indeed correspond to that of the given pairwise model.

4 Higher order models

Whilst we can recover the pairwise equations from the exact effective degree model we note that the same is not possible with the heterogeneous pairwise equations. This motivates an extension of the effective degree model where the degrees of neighbouring nodes are also taken in to account. Again we conjecture that this model can, in theory, be derived from the exact Kolmogorov equations and thus refer to it as exact.

4.1 Exact effective degree with neighbourhood composition

We extend the exact effective degree model to include the number of neighbours of the central nodes' neighbours. We begin by defining the following notation

$$\begin{aligned} s' &= (s_1, s_2, \ldots, s_M), \\ i' &= (i_1, i_2, \ldots, i_M), \\ |s'| &= s_1 + s_2 + \ldots + s_M, \\ |i'| &= i_1 + i_2 + \ldots + i_M, \end{aligned}$$

where s_j (i_j) represents the number of susceptible (infective) neighbours of degree j. We now define $S_{s'i'}$, ($I_{s'i'}$) as the number of susceptible (infective) nodes with neighbouring nodes whose own degrees are given by the entries in s' and i'. We can now write the extended ODEs in the following form

$$\begin{aligned} \dot{S_{s,'i'}} =& -\tau|i'|S_{s,'i'} + \gamma I_{s',i'} + \gamma \sum_{k=1}^{M} (i'_k + 1) S_{s'_{k-},i'_{k+}} - \gamma|i'|S_{s',i'} \\ &+ \tau \sum_{k=1}^{M} \left[IS^k S_{s'_{k+},i'_{k-}} \right] - \tau \left[ISS_{s',i'} \right], \end{aligned} \tag{15}$$

$$\begin{aligned} \dot{I_{s',i'}} =& \tau|i'|S_{s',i'} - \gamma I_{s',i'} + \gamma \sum_{k=1}^{M} (i'_k + 1) I_{s'_{k-},i'_{k+}} - \gamma|i'|I_{s',i'} \\ &+ \tau \sum_{k=1}^{M} \left[IS^k I_{s'_{k+},i'_{k-}} \right] - \tau \left[ISI_{s',i'} \right]. \end{aligned} \tag{16}$$

Here $s'_{k-} = (s_1, s_2, \ldots, s_k - 1, \ldots, s_M)$ and $s'_{k+} = (s_1, s_2, \ldots, s_k + 1, \ldots, s_M)$ with a similar definition for i'_{k-} and i'_{k+}. With a small modification to the exact effective degree notation terms such as $\left[IS^k S_{s'_{k+},i'_{k-}} \right]$ are taken to represent number of infectious contacts of the susceptible neighbours of degree k.

4.2 Model recovery

Here we show how, from the extended effective degree model, we can recover the heterogenous pairwise model. It is also straightforward to show, and thus omitted here, that the extended effective degree leads to the simpler exact effective degree. In turn, it also follows easily that both the exact effective degree and heterogenous pairwise models reduce to the standard pairwise model. This hierarchy of recovery is illustrated in Fig. 2.

4.2.1 Recovering the heterogeneous pairwise model from the extended effective degree

As earlier we make use of careful summation to recover the model. The full derivation is provided in Appendix 2 so here we just provide the derivation at the individual level and of the $[I^l I^n]$ pairs. For singles the following identities hold,

$$\frac{d}{dt}[S^n] = \sum_{|s'|+|i'|=n} \dot{S_{s',i'}} = \gamma[I^n] - \tau[S^n I],$$
$$\frac{d}{dt}[I^n] = \sum_{|s'|+|i'|=n} \dot{I_{s',i'}} = -\gamma[I^n] + \tau[S^n I],$$

where most terms from the original effective degree cancel and we have used that

$$\sum_{|s'|+|i'|=n} I_{s',i'} = [I^n] \quad \text{and} \quad \sum_{|s'|+|i'|=n} |i'| S_{s',i'} = [S^n I].$$

For the $[I^l I^n]$ pair we obtain

$$\begin{aligned}
\frac{d}{dt}\left[I^l I^n\right] = &\sum_{|s'|+|i'|=n} i'_l \dot{I_{s',i'}} \\
=&\tau \sum i'_l |i'| S_{s',i'} - \gamma \sum i'_l I_{s',i'} + \gamma \sum i'_l \sum_{k=1}^{M} (i'_k+1) I_{s'_{k-},i'_{k+}} \\
&- \gamma \sum i'_l |i'| I_{s',i'} + \tau \sum i'_l \sum_{k=1}^{M} \left[IS^k I_{s'_{k+},i'_{k-}}\right] - \tau \sum i'_l \left[ISI_{s',i'}\right] \\
=&\tau \sum i'_l \left(|i'|-1\right) S_{s',i'} + \tau \sum i'_l S_{s',i'} - \gamma \left[I^l I^n\right] \\
&+ \gamma \left[I^l I^n I\right] - \gamma \sum i'_l \left(|i'|-1\right) I_{s',i'} - \gamma \sum i'_l I_{s',i'} \\
&+ \tau \sum i'_l \sum_{k \neq l} \left[IS^k I_{s'_{k+},i'_{k-}}\right] + \tau \sum \left(i'_l - 1\right) \left[IS^l I_{s'_{l+},i'_{l-}}\right] \\
&+ \tau \sum \left[IS^l I_{s'_{l+},i'_{l-}}\right] - \tau \sum i'_l \left[ISI_{s',i'}\right] \\
=&\tau \left[I^l S^n I\right] + \tau \left[I^l S^n\right] - 2\gamma \left[I^l I^n\right] \\
&+ \gamma \left[I^l I^n I\right] - \gamma \left[I^l I^n I\right] + \tau \left[IS^l I^n\right] + \tau \left[S^l I^n\right] \\
=&\tau \left[I^l S^n I\right] + \tau \left[I^l S^n\right] - 2\gamma \left[I^l I^n\right] + \tau \left[IS^l I^n\right] + \tau \left[S^l I^n\right] \\
=&\tau \sum_q \left(\left[I^l S^n I^q\right] + [I^q S^n I^m]\right) + \tau \left[I^l S^n\right] + \tau \left[S^l I^n\right] - 2\gamma \left[I^l I^n\right].
\end{aligned}$$

Again, we note that this result corresponds to previously given heterogenous pairwise model.

5 Exactness of the models

In the previous sections we have at times referred to a set of ODEs as being exact. This terminology implies that the ODEs can be derived directly from the Kolmogorov

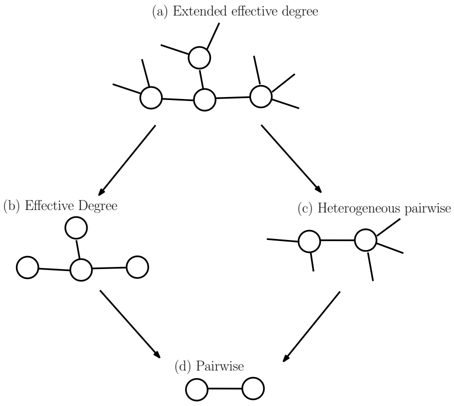


Fig. 2 Illustration of the hierarchial structure of model recovery. Links that are known are given by lines and knowledge of a nodes status is given by circles. The upper level (a) represents the extended effective degree ODEs. The status of the central node is known along with that of it's neighbours and also their degrees. The secondary level is given by (b), the effective degree model where there is no knowledge of neighbours' degrees and (c), the heterogenous pairwise model where the number of pairs of nodes and their relative degree is know. The final level shown, (d), is known as the standard pairwise model, [13], where the status of individual nodes and pairs is used.

equations which describe the evolution of the epidemic through the full state space $\mathcal{S}$ (on a network of size N, $\mathcal{S} = \{S, I\}^N$). In [21] the exactness of the pairwise equations was rigorously proven but no other motif structures were considered. In section 3.1, we conjectured that the newly defined exact effective degree model is derivable from the Kolmogorov equations. Due to the structure of the motifs used in the effective degree model a mechanistic proof (as in [21]) may be difficult and intricate to implement. Instead we will prove that a heuristic formulation of the ODEs for any motif structure is indeed exact providing they are written following rigorous bookkeeping This derivation of the evolution equations for an arbitrary motif, directly from the Kolmogorv equations, will be based on an extension of ideas presented in [11] and [21] and using the notation defined in Table 1.

We should note that in what follows a motif of connected nodes will only ever be counted once. In a network of size N and in a motif, m, with k nodes this singular counting can be understood in the following way. We consider each of the $\binom{N}{k}$ unique sets of k nodes between 1 and N. Then for each set whose nodes are isomorphic in topological structure and status to the motif m, we simply increase the counter of such motifs by one. This formalism is unlike that used in the standard pairwise model where an SS link would contribute a value of two to the $[SS]$ count. However, the two resultant sets of equations are equivalent in the sense that the different ways of counting can easily be recovered by using a simple mapping between the two. For this reason, whilst we prove that the following theorem is correct, it's intricacy and generality means a certain amount of care is needed when interpreting the resultant terms. Table 2 summarises the additional notation needed for the newly introduced motif approach. The result for a general motif is given in the following theorem.

Table 1 Notation for matrix representation of the Kolmogorov equations (Table from [21]).

Variable	Definition
N	Number of nodes in the network
$G=(g_{ij})\in\{0,1\}^{N^2}$, $i,j=1,2,\ldots,N$	Adjacency matrix with $g_{ij}=1$ if nodes i and j are connected and $g_{ij}=0$ otherwise. The network is bi-directional and has no self loops such that $G=G^T$ and $G_{ii}=0$, $\forall\ i$.
τ	Rate of infection per (S,I) edge.
γ	Rate of recovery.
$\mathcal{S}=\{S,I\}^N$	State space of the network, with nodes either susceptible, S, or infected, I and $\lvert S\rvert = 2^N$.
$\mathcal{S}^k=\{\mathcal{S}^k_1,\mathcal{S}^k_2,\ldots,\mathcal{S}^k_{c_k}\}$	The $c_k=\binom{N}{k}$ states with k infected individuals in all possible configurations, with $k=0,1,\ldots,N$.
$X^k_j(t)$	Probability of being in state $\mathcal{S}^k_j$ at time t, where $k=0,1,\ldots,N$ and $j=1,2,\ldots,c_k$.
$X^k(t)$	$X^k(t)=\left(X^k_1(t),X^k_2(t),\ldots,X^k_{c_k}(t)\right)^T$.
$A^k_{i,j}$	Rate of transition from $\mathcal{S}^{k-1}_j$ to $\mathcal{S}^k_i$, where $k=0,1,\ldots,N$, $i=1,2,\ldots,c_k$ and $j=1,2,\ldots,c_{k-1}$. Note that only one individual is changing (i.e. in this case an S node changes to an I through infection).
$C^k_{i,j}$	Rate of transition from $\mathcal{S}^{k+1}_j$ to $\mathcal{S}^k_j$, where $k=0,1,\ldots,N$, $i=1,2,\ldots,c_k$ and $j=1,2,\ldots,c_{k+1}$. Note that only one individual is changing (i.e. in this case an I node changes to an S through infection).
$B^k_{i,j}$	Rate of transition out of $\mathcal{S}^k_j$, where $B^k_{i,j}=0$ if $i\neq j$ with $k=0,1,\ldots,N$ and $i,j=1,2,\ldots,c_k$.

Theorem 1 *The equation for the expected number ($|\mathcal{M}|$) of motifs of type $\hat{m}$, given by*

$$\begin{aligned}|\dot{\mathcal{M}}| =&\tau\mathcal{N}^{SI}_{in}(\hat{m}^-,\hat{m})+\tau\mathcal{N}^{SI}_{ex}(\hat{m}^-,\hat{m})-\tau|\mathcal{M}|N^{SI}_{in}(\hat{m})-\tau\mathcal{N}^{SI}_{ex}(\hat{m})\\ &+\gamma\mathcal{N}^{I}(\hat{m}^+,\hat{m})-\gamma|\mathcal{M}|N^{I}(\hat{m})\end{aligned} \tag{17}$$

is derivable directly from the exact Kolmogorov equations.

5.1 Proof of Theorem 1

For a detailed description of writing the Kolmogorov equations for an arbitrary graph we refer the reader to [11]. Here we only provide a brief description making use of the notation defined in Table 1 to allow us to illustrate the proof. Setting $X=(X^1,X^2,\ldots,X^N)^T$, the epidemic evolution through the state space is given by

$$\dot{X}=PX, \tag{18}$$

Table 2 Additional notation for matrix representation of the Kolmogorov equations

Variable	Definition
$\hat{m}$	An arbitrary motif encompassing both topology and status of nodes (e.g. an $S-I$ edge or a star like structure such as $I_{3,0}$). The arbitrary motif we are consdering which will encompass both topology and status of nodes.
$\hat{m}^{+}$	Represents the different motifs with the same structure as $\hat{m}$ but with a susceptible node of $\hat{m}$ having become infected.
$\hat{m}^{-}$	Represents the different motifs with the same structure as $\hat{m}$ but with with an infective node of $\hat{m}$ having become susceptible.
$M_{k,j}$	Set of $\hat{m}$ motifs in configuration state $\mathcal{S}_j^k$. Defining the i^{th} element of $M_{k,j}$ as $\hat{m}_{k,j}^i$ gives $M_{k,j}=\{\hat{m}_{k,j}^1,\hat{m}_{k,j}^2,\ldots,\hat{m}_{k,j}^{\lvert M\rvert}\}$.
$M_{k,j}^{+}$	The set of motifs, in configuration state $\mathcal{S}_j^k$, with the same topology as $\hat{m}$ but with 1 more infective and 1 less susceptible. Defining the i^{th} element of $M_{k,j}^{+}$ as $\hat{m}_{k,j}^{i+}$ gives $M_{k,j}^{+}=\{\hat{m}_{k,j}^{1+},\hat{m}_{k,j}^{2+},\ldots,\hat{m}_{k,j}^{\lvert M_{k,j}^{+}\rvert+}\}$.
$M_{k,j}^{-}$	The set of motifs, in configuration state $\mathcal{S}_j^k$, with the same topology as $\hat{m}$ but with 1 less infective and 1 more susceptible. Defining the i^{th} element of $M_{k,j}^{-}$ as $\hat{m}_{k,j}^{i-}$ we have $M_{k,j}^{-}=\{\hat{m}_{k,j}^{1-},\hat{m}_{k,j}^{2-},\ldots,\hat{m}_{k,j}^{\lvert M_{k,j}^{-}\rvert-}\}$.
$N_{\hat{m}}(\mathcal{S}_j^k)$	Number of $\hat{m}$ motifs in state $\mathcal{S}_j^k$, with $k=0,1,\ldots,N$ and $j=1,2,\ldots,c_k$.
$N_{in}^{SI}(\hat{h})$	Number of SI links within the motif $\hat{h}$.
$\mathcal{N}_{in}^{SI}(\hat{h})$	Expected total number of SI links within all motifs of type $\hat{h}$
$N_{in}^{SI}(\hat{h},k)$	Number of SI links within the motif $\hat{h}$, along which, were an infection to occur, would result in a motif of type k.
$\mathcal{N}_{in}^{SI}(\hat{h},k)$	Expected total number of SI links within all motifs of type $\hat{h}$, along which, were an infection to occur, would result in a motif of type k.
$N_{ex}^{SI}(\hat{h})$	Number of SI links where the S is contained within the motif $\hat{h}$ and the I is external to it.
$\mathcal{N}_{ex}^{SI}(\hat{h})$	Expected total number of SI links to all motifs with structure $\hat{h}$, where the S is contained within the motif $\hat{h}$ and the I external to it.
$N_{ex}^{SI}(\hat{h},k)$	Number of SI links where the S is contained within the motif $\hat{h}$ and the I is external to it, along which, were an infection to occur, would result in a motif of type k.
$\mathcal{N}_{ex}^{SI}(\hat{h},k)$	Expected total number of SI links to all motifs with structure $\hat{h}$, where the S is contained within the motif $\hat{h}$ and the I external to it, along which, were an infection to occur, would result in a motif of type k.
$N^{I}(\hat{h})$	Number of I nodes within motif $\hat{h}$.
$N^{I}(\hat{h},k)$	Number of I nodes within motif $\hat{h}$, whose recovery lead to a motif of type k.
$\mathcal{N}^{I}(\hat{h},k)$	Expected total number of Is within motifs of type $\hat{h}$, whose recovery lead to a motif of type k.

where

$$P = \begin{pmatrix} B^0 & C^0 & 0 & 0 & 0 & 0 \\ A^1 & B^1 & C^1 & 0 & 0 & 0 \\ 0 & A^2 & B^2 & C^2 & 0 & 0 \\ 0 & 0 & A^3 & B^3 & C^3 & 0 \\ 0 & 0 & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & 0 & A^N & B^N \end{pmatrix}.$$

We now have the following equations for the state space probabilities.

$$\begin{aligned} \dot{X}^0 &= B^0 X^0 + C^0 X^1, \\ \dot{X}^k &= A^k X^{k-1} + B^k X^k + C^k X^{k+1} \quad \text{for } k = 1 \dots (N-1), \\ \dot{X}^N &= A^N X^{N-1} + B^N X^N. \end{aligned}$$

From [11], we also know that the entries of the matrix B are zero except on the diagonals, where we find that

$$\begin{aligned} B^k_{jj} &= -\sum_{i=1}^{c_{k+1}} A^{k+1}_{i,j} - \sum_{i=1}^{c_{k-1}} C^{k-1}_{i,j} \\ &= -\tau N_{SI}(S^k_j) - k\gamma. \end{aligned} \tag{19}$$

Where [11] focussed on individual and edge motifs here we focus on the derivation of evolution equations for the expected number of and arbitrary motif, $\hat{m}$. We begin by writing the exact equations for an arbitrary motif $\hat{m}$ based on the transition and recovery matrices. This yields,

$$\begin{aligned} |\dot{\mathcal{M}}| &= \sum_{k=0}^{N} N_{\hat{m}}(S^k)\dot{X}^k \\ &= N_{\hat{m}}(S^0)\left[B^0 X^0 + C^0 X^1\right] \\ &\quad + \sum_{k=1}^{N-1} N_{\hat{m}}(S^k)\left[A^k X^{k-1} + B^k X^k + C^k X^{k+1}\right] \\ &\quad + N_{\hat{m}}(S^N)\left[A^N X^{N-1} + B^N X^N\right] \\ &= \sum_{k=1}^{N} N_{\hat{m}}(S^k) A^k X^{k-1} + \sum_{k=0}^{N} N_{\hat{m}}(S^k) B^k X^k + \sum_{k=0}^{N-1} N_{\hat{m}}(S^k) C^k X^{k+1} \\ &= \sum_{k=0}^{N-1} N_{\hat{m}}(S^{k+1}) A^{k+1} X^k + \sum_{k=0}^{N} N_{\hat{m}}(S^k) B^k X^k + \sum_{k=1}^{N} N_{\hat{m}}(S^{k-1}) C^{k-1} X^k \\ &= \left[N_{\hat{m}}(S^1)A^1 + N_{\hat{m}}(S^0)B^0\right] X^0 \\ &\quad + \sum_{k=1}^{N-1}\left[N_{\hat{m}}(S^{k+1})A^{k+1} + N_{\hat{m}}(S^k)B^k + N_{\hat{m}}(S^{k-1})C^{k-1}\right] X^k \\ &\quad + \left[N_{\hat{m}}(S^N)B^N + N_{\hat{m}}(S^{N-1})C^{N-1}\right] X^N. \end{aligned} \tag{20}$$

Before continuing we note the following

$$
\begin{aligned}
B^N = B^N_{1,1} &= -\sum_{i=1}^{N} C^{N-1}_{i,1} = -\gamma N,\\
B^0 = B^0_{1,1} &= -\sum_{i=1}^{N} A^1_{i,1} = -\tau N_{SI}(S^0_1) = 0.
\end{aligned}
$$

Taking these and (19) into account and using the fact that B is only none zero on it's diagonal, we then obtain the following equation,

$$
\begin{aligned}
|\dot{\mathcal{M}}| =& N_{\hat{m}}(S^1)A^1X^0\\
&+ \sum_{k=1}^{N-1}\Big[N_{\hat{m}}(S^{k+1})A^{k+1} - \tau\Big(N_{\hat{m}}(S^k) * N_{SI}(S^k)\Big) - \gamma k N_{\hat{m}}(S^k) + N_{\hat{m}}(S^{k-1})C^{k-1}\,\Big]X^k\\
&+ \big[N_{\hat{m}}(S^{N-1})C^{N-1} - \gamma N N_{\hat{m}}(S^N)\big]X^N\\
=& \sum_{k=1}^{N-1}\Big[N_{\hat{m}}(S^{k+1})A^{k+1} - \tau\Big(N_{\hat{m}}(S^k) * N_{SI}(S^k)\Big)\Big]X^k - \sum_{k=1}^{N}\Big[\gamma k N_{\hat{m}}(S^k) - N_{\hat{m}}(S^{k-1})C^{k-1}\Big]X^k.
\end{aligned}
\tag{21}
$$

We note that the term containing X^0 vanishes because A^1 is a column vector with all zero entries. We now consider the summations involving the A and C matrices:

$$
\begin{aligned}
\left[N_{\hat{m}}(S^{k+1})A^{k+1}\right]_j =& \sum_{i=1}^{c_{k+1}} N_{\hat{m}}(S_i^{k+1})A_{i,j}^{k+1} \\
=& r_1\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node 1 becoming infected })\right. \\
&\left.-(\text{number of } \hat{m} \text{ lost by node 1 becoming infected })\right] \\
&+r_2\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node 2 becoming infected })\right. \\
&\left.-(\text{number of } \hat{m} \text{ lost by node 2 becoming infected })\right] \\
&+\dots \\
&+r_{N-k}\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node } (N-k) \text{ becoming infected })\right. \\
&\left.-(\text{number of } \hat{m} \text{ lost by node } (N-k) \text{ becoming infected })\right] \\
=& r_1\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^- \text{ where node 1 is susceptible}\right. \\
&\qquad \text{and where node } 1's \text{ infection would lead to a motif of type } \hat{m}) \\
&\left.-(\text{number of elements of } M_{k,j} \text{ where node 1 is susceptible })\right] \\
&+r_2\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^- \text{ where node 2 is susceptible}\right. \\
&\qquad \text{and where node } 2's \text{ infection would lead to a motif of type } \hat{m}) \\
&\left.-(\text{number of elements of } M_{k,j} \text{ where node 2 is susceptible })\right] \\
&+\dots \\
&+r_{N-k}\tau\left[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^- \text{ where node } (N-k) \text{ is susceptible}\right. \\
&\qquad \text{and where node } (N-k)'s \text{ infection would lead to a motif of type } \hat{m}) \\
&\left.-(\text{number of elements of } M_{k,j} \text{ where node } (N-k) \text{ is susceptible })\right],
\end{aligned}
$$

grouping the terms we obtain,

$$
\begin{aligned}
\left[N_{\hat{m}}(S^{k+1})A^{k+1}\right]_j =& \tau N_{SI}(S_j^k)N_{\hat{m}}(S_j^k) + \tau \sum_{i=1}^{|M_{k,j}^-|}\left[N_{in}^{SI}(\hat{m}_{k,j}^{i-},\hat{m}) + N_{ex}^{SI}(\hat{m}_{k,j}^{i-},\hat{m})\right] \\
&-\tau|M_{k,j}|N_{in}^{SI}(\hat{m}) - \tau\sum_{i=1}^{|M_{k,j}|}\left[N_{ex}^{SI}(\hat{m}_{k,j}^{i})\right].
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\left[N_{\hat{m}}(S^{k-1})C^{k-1}\right]_j =& \sum_{i=1}^{c_{k-1}} N_{\hat{m}}(S_i^{k-1})C_{i,j}^{k-1} \\
=& \gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node } (N-k+1) \text{ recovering }) \\
& -(\text{number of } \hat{m} \text{ lost by node } (N-k+1) \text{ recovering })\Big] \\
& +\gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node } (N-k+2) \text{ recovering }) \\
& -(\text{number of } \hat{m} \text{ lost by node } (N-k+2) \text{ recovering })\Big] \\
& +\dots \\
& +\gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of } \hat{m} \text{ gained by node } (N) \text{ recovering }) \\
& -(\text{number of } \hat{m} \text{ lost by node } (N) \text{ recovering })\Big] \\
=& \gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^+ \text{ where node } (N-k+1) \text{ is infective} \\
& \qquad \text{and where node } (N-k+1)'s \text{ recovery would lead to a motif of type } \hat{m}) \\
& -(\text{number of elements of } M_{k,j} \text{ of which node } (N-k+1) \text{ belongs })\Big] \\
& +\gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^+ \text{ where node } (N-k+2) \text{ is infective} \\
& \qquad \text{and where node } (N-k+2)'s \text{ recovery lead to a motif of type } \hat{m}) \\
& -(\text{number of elements of } M_{k,j} \text{ of which node } (N-k+2) \text{ belongs })\Big] \\
& +\dots \\
& +\gamma\Big[N_{\hat{m}}(S_j^k) + (\text{number of elements of } M_{k,j}^+ \text{ where node } (N) \text{ is infective} \\
& \qquad \text{and where node } N's \text{ recovery would lead to a motif of type } \hat{m}) \\
& -(\text{number of elements of } M_{k,j} \text{ of which node } (N) \text{ belongs })\Big],
\end{aligned}
$$

grouping the terms we obtain

$$
\left[N_{\hat{m}}(S^{k-1})C^{k-1}\right]_j = \gamma k N_{\hat{m}}(S_j^k) + \gamma \sum_{i=1}^{|M_{k,j}^+|} N^I(\hat{m}_{k,j}^{i+}, \hat{m}) - \gamma |M_{k,j}| \left(N^I(\hat{m})\right).
$$

Defining

$$
\mathcal{A}_j^{k+1} = \tau \sum_{i=1}^{|M_{k,j}^-|} \left[N_{in}^{SI}(\hat{m}_{k,j}^{i-}, \hat{m}) + N_{ex}^{SI}(\hat{m}_{k,j}^{i-}, \hat{m})\right] - \tau |M_{k,j}| N_{in}^{SI}(\hat{m}) - \tau \sum_{i=1}^{|M_{k,j}|} \left[N_{ex}^{SI}(\hat{m}_{k,j}^{i})\right]
$$

$$
\mathcal{C}_j^{k-1} = \gamma \sum_{i=1}^{|M_{k,j}^+|} \left[N^I(\hat{m}_{k,j}^{i+}, \hat{m})\right] - \gamma |M_{k,j}| \left(N^I(\hat{m})\right)
$$

and setting $\mathcal{A}^{k+1} = [A_1^{k+1}, A_j^{k+1}, \ldots, A_{c_k}^{k+1}]$ and $\mathcal{C}^{k-1} = [C_1^{k-1}, C_j^{k-1}, \ldots, C_{c_{k-1}}^{k-1}]$ yields,

$$
\begin{aligned}
|\dot{\mathcal{M}}| =& \sum_{k=1}^{N-1} \left[N_{\hat{m}}(S^{k+1})A^{k+1} - \tau \left(N_{\hat{m}}(S^K) * N_{SI}(S^k) \right) \right] X^k - \sum_{k=1}^{N} \left[\gamma k N_{\hat{m}}(S^k) - N_{\hat{m}}(S^{k-1})C^{k-1} \right] X^k \\
=& \sum_{k=1}^{N-1} \left[\tau \left(N_{\hat{m}}(S^K) * N_{SI}(S^k) \right) + \mathcal{A}^{k+1} - \tau \left(N_{\hat{m}}(S^K) * N_{SI}(S^k) \right) \right] X^k - \\
& \sum_{k=1}^{N} \left[\gamma k N_{\hat{m}}(S^k) - \left(k N_{\hat{m}}(S^k) + \mathcal{C}^{k-1} \right) \right] X^k \\
=& \sum_{k=1}^{N-1} \left[\mathcal{A}^{k+1} \right] X^k + \sum_{k=1}^{N} \left[\mathcal{C}^{k-1} \right] X^k \\
=& \sum_{k=1}^{N-1} \sum_{j=1}^{c_k} \mathcal{A}_j^{k+1} X_j^k + \sum_{k=1}^{N} \sum_{j=1}^{c_k} C_j^{k-1} X_j^k \\
=& \sum_{k=1}^{N-1} \sum_{j=1}^{c_k} \left\{ \tau \sum_{i=1}^{|M_{k,j}^-|} \left[N_{in}^{SI}(\hat{m}_{k,j}^{i-}, \hat{m}) + N_{ex}^{SI}(\hat{m}_{k,j}^{i-}, \hat{m}) \right] - \tau |M_{k,j}| N_{in}^{SI}(\hat{m}) - \tau \sum_{i=1}^{|M_{k,j}|} \left[N_{ex}^{SI}(\hat{m}_{k,j}^{i}) \right] \right\} X_j^k \\
& + \sum_{k=1}^{N} \sum_{j=1}^{c_k} \left\{ \gamma \sum_{i=1}^{|M_{k,j}^+|} \left[N^I(\hat{m}_{k,j}^{i+}, \hat{m}) \right] - \gamma |M_{k,j}| \left(N^I(\hat{m}) \right) \right\} X_j^k \\
=& \tau \mathcal{N}_{in}^{SI}(\hat{m}^-, \hat{m}) + \tau \mathcal{N}_{ex}^{SI}(\hat{m}^-, \hat{m}) - \tau |\mathcal{M}| N_{in}^{SI}(\hat{m}) - \tau \mathcal{N}_{ex}^{SI}(\hat{m}) \\
& + \gamma \mathcal{N}^I(\hat{m}^+, \hat{m}) - \gamma |\mathcal{M}| N^I(\hat{m}).
\end{aligned}
$$

Which matches equation 17 from Theorem 1. □

5.2 Proof that the conjectured exact effective degree model is derivable from the Kolmogorov equations

Letting $\hat{m}$ be an $S_{s,i}$-type motif from the effective degree model earlier and using Theorem 1, we find that the exact equations can be written as

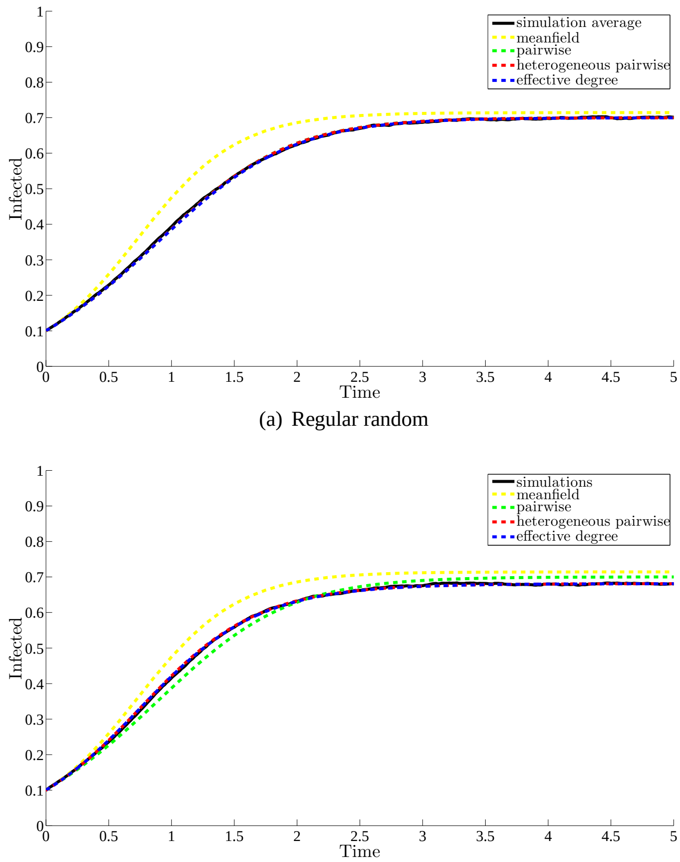
$$
\begin{aligned}
\frac{dS_{s,i}}{dt} =& \tau\mathcal{N}^{SI}_{in}(\hat{m}^-,\hat{m}) + \tau\mathcal{N}^{SI}_{ex}(\hat{m}^-,\hat{m}) - \tau|\mathcal{M}|N^{SI}_{in}(\hat{m}) - \tau\mathcal{N}^{SI}_{ex}(\hat{m}) \\
&+ \gamma\mathcal{N}^{I}(\hat{m}^+,\hat{m}) - \gamma|\mathcal{M}|N^{I}(\hat{m}) \\
=& \tau \times (\text{the total expected number of SI connections within } S_{s+1,i-1}\text{-type motifs} \\
&\quad \text{where if infection occurs we obtain a } S_{s,i}\text{-type motif }) \\
&+ \tau \times (\text{the total expected number of SI connections where S lies within} \\
&\quad S_{s+1,i-1}\text{-type motifs and the I is external to the given motif} \\
&\quad \text{and where, were an infection to occur, we obtain a } S_{s,i}\text{-type motif }) \\
&- \tau S_{s,i} \times (\text{number of SI connections within an individual } S_{s,i}\text{-type motifs}) \\
&- \tau \times (\text{the total expected number of SI connections where S belongs to} \\
&\quad S_{s,i}\text{-type motifs and the I is external to the given motif }) \\
&+ \gamma \times (\text{the total expected number I's within } S_{s-1,i+1}\text{-type and } I_{s,i}\text{-type motifs} \\
&\quad \text{where there recovery would give a } S_{s,i}\text{-type motif }) \\
&- \gamma S_{s,i} \times (\text{number of I within an individual } S_{s,i}\text{-type motif}) \\
=& \tau\,[ISS_{s+1,i-1}] - \tau i S_{s,i} - \tau\,[ISS_{s,i}] + \gamma I_{si} + \gamma(i+1)S_{s-1,i+1} - \gamma i S_{s,i}
\end{aligned}
$$

which is indeed the conjectured exact equation for $S_{s,i}$ (similar derivation holds for $I_{s,i}$). To clarify the above derivation we note that a term such as $\tau\mathcal{N}^{SI}_{in}(\hat{m}^-,\hat{m})$ will make no contribution to the resultant equation as there are no internal *SI* connections within $S_{s-1,i+1}$-type motifs along which an infection would lead to an $S_{s,i}$-type motif. However other terms, such as $\tau\mathcal{N}^{SI}_{ex}(\hat{m}^-,\hat{m})$, have a direct correspondence with the resultant output (in this case the $\tau\,[ISS_{s+1,i-1}]$ term).

6 Comparison of the closed models

In comparing the models the obvious question to ask is when does one model perform better than another, i.e. which model approximates better or more accurately the simulation results or the solution of the Kolmogorov/master equations where solvable. As discussed earlier, the pairwise model is known to perform well on networks that are well characterised by the average degree (i.e. regular random and Erdős-Rényi graphs). What is less known is under what circumstances do the heterogenous pairwise and effective degree models outperform one another.

To assess the performance of the three closed models we compared individual simulations to the solutions of the ODE's on four different types of undirected network. Firstly we use regular random networks where all nodes have the same number of randomly chosen neighbours. Secondly, on an Erdős-Rényi random network where the distribution of degrees converges to a Poisson distribution. Figure 3 plots simulation results against the different solutions of the ODEs for these two networks. On the regular network, whilst the two different pairwise models and the effective degree offer an improvement in performance over the standard meanfield equations, there is little to distinguish between the improved approaches. On the Erdős-Rényi random networks, the pairwise model improves on the meanfield model and, in turn, the effective degree and heterogeneous pairwise models improve even further on this. Again, however, there is little to distinguish between effective degree and the heterogeneous pairwise models.



(a) Regular random

(b) Erdos-Renyi

Fig. 3 ODE performance on different networks Each network is of size $N = 500$ and with disease parameters given by $\gamma = 1$ and $\tau = 0.5$. Average prevalence was calculated from individual simulations on 100 different networks. (a) Regular random network, each node having degree 7. (b) Erdős-Rényi random network with average degree 7.

To investigate further we ran simulations on networks exhibiting greater heterogeneity. Firstly on a bimodal network generated by the configuration model algorithm [18] and secondly on assortative bimodal networks generated by rewiring the aforementioned bimodal networks according to the algorithm of Newman, [17]. The results are illustrated in Figure 4. Whilst on random bimodal network there is little difference between heterogeneous pairwise and effective degree when assortativity is added, there is a clear improvement in the performance of the heterogenous pairwise model over the effective degree. This performance benefit must, however, be considered in terms of the model complexity given in table 3 (note in this table M is the maximum possible degree in the network and we given the minimum number of equations needed to implement the ODEs).

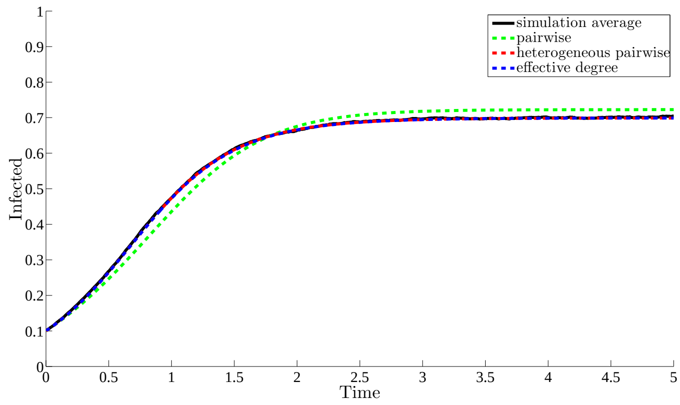
Table 3 Complexity of closed ODEs

Model	# equations	complexity
meanfield	1	$\mathcal{O}(1)$
pairwise	3	$\mathcal{O}(1)$
effective degree	$M(M+3)-1$	$\mathcal{O}(M^2)$
heterogeneous pairwise	$2M(M+1)-1$	$\mathcal{O}(M^2)$
Kolmogorov equations	2^N	$\mathcal{O}(2^N)$

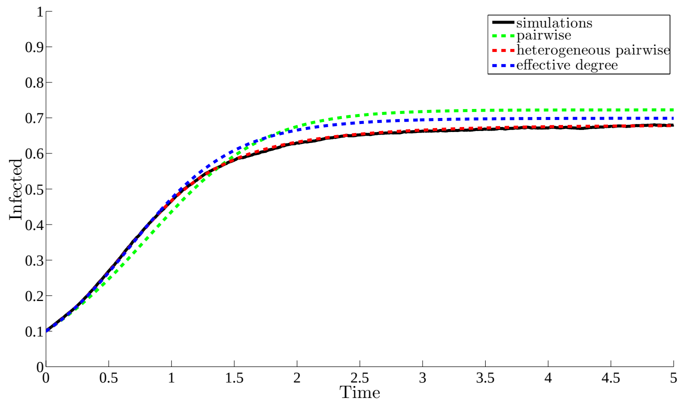
A final comparison between the performance of the different closed models is to look at their rate of convergence to the solution of the Kolmogorov equations on a complete (fully connected) network. On a complete network it is possible (see [11]) to reduce the full system of 2^N equations to just $N+1$ equations. This allows us to compare the true solution to the approximate solution of the meanfield, pairwise (equivalent to heterogenous pairwise on a complete graph) and effective degree models. Interestingly we find that all three exhibit $O(1/N)$ convergence, where although both pairwise and effective degree bring an improvement on meanfield, the difference between the convergence of the two is neglible and almost indecernible (see figure 5).

7 Discussion

In this paper we set out to achieve a greater understanding of the relation between some of the more common approaches to modelling disease dynamics. In doing so we conjectured an exact version of the effective degree model [15] and showed how this model could be used to recover the pairwise model [13]. We then extended this model to incorporate greater network structure and illustrated how, from this extension, we could then recover the heterogeneous pairwise model [6]. We then proved that the conjectured exact effective degree model was indeed exact by proving that a heuristic derivation of an ODE model for an arbitrary motif was derivable directly from the Kolmogorov equations and noting that the exact effective degree model was just a particular case of this heuristic model. Finally we considered the performance



(a) Bimodal



(b) Assortative bimodal

Fig. 4 ODE performance on different networks Each network is of size $N = 500$ and with disease parameters given by $\gamma = 1$ and $\tau = 0.5$. Average prevalence was calculated from individual simulations on 100 different networks. (a) Bimodal configuration model with 207 nodes of degree 4 and 293 nodes of degree 10. (b) Assortative bimodal network, with same degree distribution as (c) but rewired to have assortativity coefficient $r \approx 0.56$.

of the different models on four different type of networks and have analysed numerically the rate of convergence to the lumped Kolmogorov equations on a complete network. These comparisons suggest a performance hierarchy of models as illustrated in Figure 6 and it is worth noting that the performance benefit of the heterogenous pairwise model on networks exhibiting susceptible $\rightarrow$ infectious $\rightarrow$ removed (SIR) disease dynamics was also touched upon in [4].

Whilst we have shown how current models can be extended in a way that can capture more network topology, these extensions have a more theoretical rather than practical motivation as their added complexity makes them not only less tractable but also more resource intensive in their solving, thus making the use of simulations more of an attractive proposition. As the links between these models are better understood,

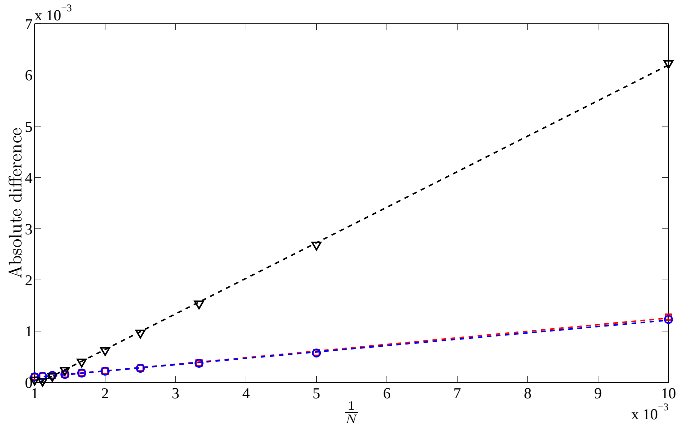


Fig. 5 Convergence to exact solution on a complete graph. Absolute difference between the exact steady state solution of the percentage of infected individuals and those calculated from three different ODE models for 10 different network sizes and initial prevalence of 40 percent. Black triangles represent meanfield, blue circles effective degree and red squares the pairwise equations. Linear lines of best fit are also shown. This shows that the error(N) appears to be of O(1/N) as N tends to infinity.

future work will likely focus on the following three areas. Firstly, a more realistic network will have a more clique-like structure. For example an individual is likely a member of a household in which he has regular contacts within and less regular contacts outside. Being able to incorporate this household structure within epidemic models is thus important in understanding the outbreak and necessary curtailment of an infectious disease (see [3,9,22]). Secondly, a network of individuals is not well represented by a static network. An individual may have regular contact with few individuals but may create or break contacts with others in ways that a static network representation cannot capture. For this reason it is important to take into consideration not only the dynamics of the disease but also the dynamics of the network and how the two impact on one another (see [7,12]). Thirdly, assuming we can write down exact differential equations we have to close them in some way. Understanding the performance of current, and also the derivation of new closures, is arguably the most important task ahead as it is the closures that limit the performance of any system of ODEs.

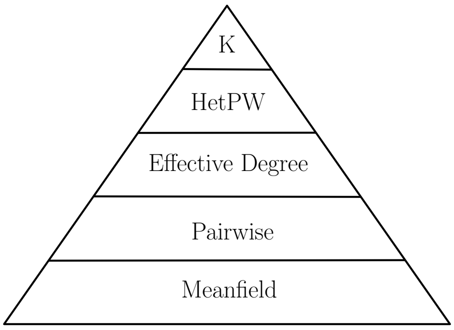


Fig. 6 Model performance hierarchy. Model performance hierarchy based on our observations. Here K represents the Kolmogorov equations and HetPW the heterogeneous pairwise equations.

Appendix 1

Derivation of the pairwise equation from the exact effective degree model for singles and pairs are as follows,

$$\frac{d}{dt}[S] = \sum_{s,i} \dot{S_{s,i}} = \gamma[I] - \tau[SI],$$
$$\frac{d}{dt}[I] = \sum_{s,i} \dot{I_{s,i}} = -\gamma[I] + \tau[SI],$$

where most terms from the original effective degree equations cancel and we have used that $\sum_{s,i} iS_{s,i} = [SI]$ and $\sum_{s,i} I_{s,i} = [I]$. For the pairs the effective degree model yields,

$$\begin{aligned}
\frac{d}{dt}[SS] = &\sum_{s,i} s\dot{S_{si}} \\
=&-\tau\sum siS_{s,i} + \gamma\sum sI_{s,i} + \gamma\sum s(i+1)S_{s-1,i+1} - \gamma\sum isS_{s,i} \\
&+\tau\sum s[ISS_{s+1,i-1}] - \tau\sum s[ISS_{s,i}] \\
=&-\tau[ISS] + \gamma[IS] + \gamma\sum(s-1)(i+1)S_{s-1,i+1} + \gamma\sum(i+1)S_{s-1,i+1} \\
&-\gamma[ISS] + \tau\sum(s+1)[ISS_{s+1,i-1}] - \tau\sum[ISS_{s+1,i-1}] - \tau\sum s[ISS_{s,i}] \\
=&-\tau[ISS] + \gamma[IS] + \gamma[ISS] + \gamma[IS] - \gamma[ISS] - \tau[ISS] \\
=&-2\tau[ISS] + 2\gamma[IS],
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}[SI] = &\sum_{s,i} s\dot{I_{si}} \\
=&\tau\sum siS_{s,i} - \gamma\sum sI_{s,i} + \gamma\sum s(i+1)I_{s-1,i+1} - \gamma\sum siI_{s,i} \\
&+\tau\sum s[ISI_{s+1,i-1}] - \tau\sum s[ISI_{s,i}] \\
=&\tau[ISS] - \gamma[IS] + \gamma\sum(s-1)(i+1)I_{s-1,i+1} + \gamma\sum(i+1)I_{s-1,i+1} \\
&-\gamma[IIS] + \tau\sum(s+1)[ISI_{s+1,i-1}] - \tau\sum[ISI_{s+1,i-1}] - \tau\sum s[ISI_{s,i}] \\
=&\tau[ISS] - \gamma[IS] + \gamma[IIS] + \gamma[II] - \gamma[IIS] - \tau([ISI] + [IS]) \\
=&\tau([ISS] - [ISI] - [IS]) + \gamma([II] - [IS]),
\end{aligned}$$

$$
\begin{aligned}
\frac{d}{dt}[II] =& \sum_{s,i} i\dot{I_{si}} \\
=& \tau \sum i^2 S_{s,i} - \gamma \sum i I_{s,i} + \gamma \sum i(i+1) I_{s-1,i+1} - \gamma \sum i^2 I_{s,i} \\
&+ \tau \sum i[ISI_{s+1,i-1}] - \tau \sum i[ISI_{s,i}] \\
=& \tau \sum i(i-1) S_{s,i} + \tau \sum i S_{s,i} - \gamma[II] \\
&+ \gamma[III] - \gamma \sum i(i-1) I_{s,i} - \gamma \sum i I_{s,i} \\
&+ \tau \sum (i-1)[ISI_{s+1,i-1}] + \tau \sum [ISI_{s+1,i-1}] - \tau \sum i[ISI_{s,i}] \\
=& \tau[ISI] + \tau[IS] - \gamma[II] + \gamma[III] - \gamma[III] - \gamma[II] + \tau[ISI] + \tau[IS] \\
=& 2\tau([ISI] + [IS]) - 2\gamma[II].
\end{aligned}
$$

Appendix 2

Derivation of the heterogeneous pairwise equations from the effective degree with neighbourhood composition model with the neighbourhood composition model. For singles and pairs the following identities hold,

$$
\begin{aligned}
\frac{d}{dt}[S^n] &= \sum_{|s'|+|i'|=N} \dot{S_{s',i'}} = \gamma[I^n] - \tau[S^n I], \\
\frac{d}{dt}[I^n] &= \sum_{|s'|+|i'|=N} \dot{I_{s',i'}} = -\gamma[I^n] + \tau[S^n I],
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\left[S^l S^n\right] =& \sum_{|s'|+|i'|=n} s'_l \dot{S_{s',i'}} \\
=& -\tau \sum s'_l |i'| S_{s',i'} + \gamma \sum s'_l I_{s',i'} + \gamma \sum s'_l \sum_{k=1}^{M} (i'_k + 1) S_{s'_{k-},i'_{k+}} \\
&- \gamma \sum s'_l |i'| S_{s',i'} + \tau \sum s'_l \sum_{k=1}^{M} \left[IS^k S_{s'_{k+},i'_{k-}}\right] - \tau \sum s'_l \left[ISS_{s',i'}\right] \\
=& -\tau\left[IS^n S^l\right] + \gamma\left[S^l I^n\right] + \gamma \sum s'_l \sum_{k \neq l} (i'_k + 1) S_{s'_{k-},i'_{k+}} \\
&\gamma \sum (s'_l - 1)(i'_l + 1) S_{s'_{l-},i'_{l+}} + \gamma \sum (i'_l + 1) S_{s'_{l-},i'_{l+1}} - \gamma\left[IS^n S^l\right] \\
&+ \tau \sum s'_l \sum_{k \neq l} \left[IS^k S_{s'_{k+},i'_{k-}}\right] + \tau \sum (s'_l + 1)\left[IS^l S_{s'_{l+},i'_{l-}}\right] \\
&- \tau \sum \left[IS^l S_{s'_{l+},i'_{l-}}\right] - \tau \sum s'_l \left[ISS_{s',i'}\right] \\
=& -\tau\left[IS^n S^l\right] + \gamma\left[S^l I^n\right] + \gamma\left[S^l S^n I\right] + \gamma\left[I^l S^n\right] \\
&- \gamma\left[IS^n S^l\right] - \tau\left[IS^l S^n\right] \\
=& -\tau\left[IS^n S^l\right] - \tau\left[IS^l S^n\right] + \gamma\left[S^l I^n\right] + \gamma\left[I^l S^n\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\left[S^l I^n\right] = &\sum_{|s'|+|i'|=n} s'_l \dot{I_{s',i'}} \\
=&\tau \sum s'_l |i'| S_{s',i'} - \gamma \sum s'_l I_{s',i'} + \gamma \sum s'_l \sum_{k=1}^{M} (i'_k+1) I_{s'_{k-},i'_{k+}} \\
&- \gamma \sum s'_l |i'| I_{s',i'} + \tau \sum s'_l \sum_{k=1}^{M} \left[IS^k I_{s'_{k+},i'_{k-}}\right] - \tau \sum s'_l \left[ISI_{s',i'}\right] \\
=&\tau \left[IS^n S^l\right] - \gamma \left[S^l I^n\right] + \gamma \sum s'_l \sum_{k\neq l} (i'_k+1) I_{s'_{k-},i'_{k+}} \\
&\gamma \sum (s'_l-1)(i'_l+1) I_{s'_{l-},i'_{l+}} + \gamma \sum (i'_l+1) I_{s'_{l-},i'_{l+1}} - \gamma \left[II^n S^l\right] \\
&+ \tau \sum s'_l \sum_{k\neq l} \left[IS^k I_{s'_{k+},i'_{k-}}\right] + \tau \sum (s'_l+1) \left[IS^l I_{s'_{l+},i'_{l-}}\right] \\
&- \tau \sum \left[IS^l I_{s'_{l+},i'_{l-}}\right] - \tau \sum s'_l \left[ISI_{s',i'}\right] \\
=&\tau \left[IS^n S^l\right] - \gamma \left[S^l I^n\right] + \gamma \left[S^l I^n I\right] + \gamma \left[I^l I^n\right] \\
&- \gamma \left[II^n S^l\right] - \tau \left[IS^l I^n\right] - \tau \left[S^l I^n\right] \\
=&\tau \left[IS^n S^l\right] - \tau \left[IS^l I^n\right] - \tau \left[S^l I^n\right] + \gamma \left[I^l I^n\right] - \gamma \left[S^l I^n\right],
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}\left[I^l I^n\right] = &\sum_{|s'|+|i'|=n} i'_l \dot{I_{s',i'}} \\
=&\tau \sum i'_l |i'| S_{s',i'} - \gamma \sum i'_l I_{s',i'} + \gamma \sum i'_l \sum_{k=1}^{M} (i'_k+1) I_{s'_{k-},i'_{k+}} \\
&- \gamma \sum i'_l |i'| I_{s',i'} + \tau \sum i'_l \sum_{k=1}^{M} \left[IS^k I_{s'_{k+},i'_{k-}}\right] - \tau \sum i'_l \left[ISI_{s',i'}\right] \\
=&\tau \sum i'_l \left(|i'|-1\right) S_{s',i'} + \tau \sum i'_l S_{s',i'} - \gamma \left[I^l I^n\right] \\
&+ \gamma \left[I^l I^n I\right] - \gamma \sum i'_l \left(|i'|-1\right) I_{s',i'} - \gamma \sum i'_l I_{s',i'} \\
&+ \tau \sum i'_l \sum_{k\neq l} \left[IS^k I_{s'_{k+},i'_{k-}}\right] + \tau \sum \left(i'_l-1\right) \left[IS^l I_{s'_{l+},i'_{l-}}\right] \\
&+ \tau \sum \left[IS^l I_{s'_{l+},i'_{l-}}\right] - \tau \sum i'_l \left[ISI_{s',i'}\right] \\
=&\tau \left[I^l S^n I\right] + \tau \left[I^l S^n\right] - 2\gamma \left[I^l I^n\right] \\
&+ \gamma \left[I^l I^n I\right] - \gamma \left[I^l I^n I\right] + \tau \left[IS^l I^n\right] + \tau \left[S^l I^n\right] \\
=&\tau \left[I^l S^n I\right] + \tau \left[I^l S^n\right] - 2\gamma \left[I^l I^n\right] + \tau \left[IS^l I^n\right] + \tau \left[S^l I^n\right].
\end{aligned}
$$

References

1. L.J. Allen. Introduction to stochastic epidemic models. In *Mathematical Epidemiology* [Lecture Notes in Mathematics, vol 1945], Springer, Berlin. 81-130 (2008).
2. R.M. Anderson & R.M. May. Infectious diseases of humans: dynamics and control. Oxford University Press (1999).
3. F. Ball, D. Sirl & P. Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. Math. Biosci. 224, 2, 53-73 (2010).
4. L. Danon, A.P. Ford, T. House, et al. Networks and the Epidemiology of Infectious Disease. Interdisciplinary Perspectives on Infectious Diseases, vol. 2011 (2011)
5. O. Diekmann & J.A.P. Heesterbeek. Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Wiley, Chichester.
6. K.T.D. Eames &M.J. Keeling. Modelling dynamic and network heterogeneneities in the spread of sexually transmitted diseases. Proc. Natl Acad. Sci. USA 99, 1333013335 (2002).
7. T. Gross, C.D. D'Lima & B. Blasius. Epidemic dynamics on an adaptive network. Phys. Rev. Letters 96, 208701-4 (2006)
8. C. Hadjichrysanthou, M. Broom & I.Z. Kiss. Approximating evolutionary dynamics on networks using a neighbourhood configuration model. J. Theor. Biol. 312, 13-21 (2012).
9. T. House, M.J. Keeling. Deterministic epidemic models with explicit household structure. Math. Biosci. 213, 29-39 (2008).
10. T. House, G. Davies, L. Danon & M.J. Keeling. A motif-based approach to network epidemics. Bulletin of Mathematical Biology, 71, 1693-1706 (2009).
11. P.L. Simon, M. Taylor & I.Z. Kiss. Exact epidemic models on graphs using graph-automorphism driven lumping. J. Math. Biol. 62, 479-508 (2011).
12. I.Z. Kiss, L. Berthouze, T.J. Taylor & P.L. Simon. Modelling approaches for simple dynamic networks and applications to disease transmission models. Proc. R. Soc. A. 468, 1332 (2012).
13. M.J. Keeling. The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. Lond. B 266, 859-867 (1999).
14. E. Kenah & J.C. Miller. Epidemic percolation networks, epidemic outcomes, and interventions. Interdisciplinary Perspectives on Infectious Diseases, 2011, (2011).
15. J. Lindquist, J. Ma, P. van den Driessche & F.H. Willeboordse. Effective degree network disease models. J. Math. Biol. 62, 143-164 (2011).
16. V. Marceau, P. Noel, L. Hbert-Dufresne, A. Allard & L.J. Dub. Adaptive networks: Coevolution of disease and topology. Phys. Rev. E 82, 036116 (2010).
17. M.E.J. Newman. Mixing patterns in networks. Phys. Rev. E 67, 026126 (2003).
18. M.E.J. Newman. The structure and function of complex networks. SIAM Rev. 45, 167-256 (2003).
19. D.A. Rand. Correlation equations and pair approximations for spatial ecologies in Advanced Ecological Theory: principles and applications, Edited by Jacqueline McGlade. Blackwell Science (1999).
20. K. Sharkey. Deterministic epidemiological models at the individual level. J. Math. Biol. 57, 311-331 (2008).
21. M. Taylor, P.L. Simon, D.M. Green, T. House & I.Z. Kiss. From Markovian to pairwise epidemic models and the performance of moment closure approximations. J. Math. Biol. 64, 1021-1042 (2012).
22. E. Volz, J.C. Miller, A. Galvani, L.A. Meyers. Effects of Heterogeneous and Clustered Contact Patterns on Infectious Disease Dynamics PLoS Comput. Biol. 7(6), e1002042 (2011)