Journal of Theoretical Biology 500 (2020) 110328

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Improving pairwise approximations for network models with susceptible-infected-susceptible dynamics

Trystan Leng [a,*], Matt J. Keeling [b]

[a] EPSRC & MRC Centre for Doctoral Training in Mathematics for Real-World Systems, University of Warwick, United Kingdom
[b] Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, Mathematics Institute and School of Life Sciences, University of Warwick, United Kingdom

ARTICLE INFO

Article history:
Received 7 October 2019
Revised 2 April 2020
Accepted 8 May 2020
Available online 23 May 2020

Keywords:
Epidemics
SIS-dynamics
Moment closure
Networks

ABSTRACT

Network models of disease spread play an important role in elucidating the impact of long-lasting infectious contacts on the dynamics of epidemics. Moment-closure approximation is a common method of generating low-dimensional deterministic models of epidemics on networks, which has found particular success for diseases with susceptible-infected-recovered (SIR) dynamics. However, the effect of network structure is arguably more important for sexually transmitted infections, where epidemiologically relevant contacts are comparatively rare and longstanding, and which are in general modelled via the susceptible-infected-susceptible (SIS)-paradigm. In this paper, we introduce an improvement to the standard pairwise approximation for network models with SIS-dynamics for two different network structures: the isolated open triple (three connected individuals in a line) and the k-regular network. This improvement is achieved by tracking the rate of change of errors between triple values and their standard pairwise approximation. For the isolated open triple, this improved pairwise model is exact, while for k-regular networks a closure is made at the level of triples to obtain a closed set of equations. This improved pairwise approximation provides an insight into the errors introduced by the standard pairwise approximation, and more closely matches both higher-order moment-closure approximations and explicit stochastic simulations with only a modest increase in dimensionality to the standard pairwise approximation.

1. Introduction

The spread of any epidemic can be conceptualised as a process on a network, where individuals are represented as nodes and epidemiologically relevant contacts as edges between nodes. An abundance of different network-based approaches to disease spread have been developed over the years, varying in scope, application, and sophistication. These range from, at one extreme, Markovian state-based models, where the probability of a system being in a certain state is given exactly by its master equations (see Kiss et al., 2017 for an introduction to such methods), to explicit stochastic simulations of epidemics on networks (see Goodreau et al., 2017 and Whittles et al., 2019 for recent examples) at the other. Both approaches have limitations. The exponentially increasing state-space with network size for state-based models mean these exact descriptions are computationally unfeasible for most networks of real-world interest; and while stochastic simulations can deal with networks of these sizes, such methods offer little or no analytical tractability, making sensitivity to network structure hard to quantify and the causal determinants of the resulting dynamics hard to identify.

One network approach that aims to bridge this gap is *moment-closure approximation*. In a population, the rate of change of the number of infected individuals will depend upon how many susceptible-infected pairs there are. The rate of change of these pairs, in turn, depends upon the number of triples, and so on up to the full size of the population. Moment-closure approximation methods obtain a closed set of ordinary differential equations (ODEs) for the disease dynamics by approximating the dynamics of higher-order moments (e.g. triples) in terms of lower-order moments (e.g. singletons and pairs). By doing so, one obtains a relatively simple ODE model that retains much of the tractability of mean-field approximation models (the standard approach to modelling the spread of infectious diseases) but that also explicitly accounts for some aspects of network structure. Hence, there has been much interest and research into such methods, and into the errors such approximations introduce into a model (Keeling et al., 2016; Pellis et al., 2015; Sharkey, 2011; Taylor et al., 2012).

* Corresponding author.
E-mail address: T.Leng@Warwick.ac.uk (T. Leng).

https://doi.org/10.1016/j.jtbi.2020.110328
0022-5193/

There has been considerable progress in this moment-closure method for diseases that can be modelled via the susceptible-infected-recovered (SIR) paradigm: the determinants of errors in such methods are detailed by Sharkey (2011); the exactness of a closure at the level of triples for tree-like networks is proven by Sharkey et al. (2015); this framework is extended by Kiss et al. (2015) to more realistic network structures that include loops; (Trapman, 2007) defines a reproduction number for pairwise approximation; (House, 2015) provides an algebraic moment-closure for such diseases based on Lie algebraic methods; while (Pellis et al., 2015) explore the exactness of closures when infective periods are of a constant duration.

By comparison, progress has been modest for diseases with susceptible-infected-susceptible (SIS) dynamics, equivalent to the network-based contact process (Liggett, 2013), where recovery from infection does not lead to immunity. Despite its lower dimensionality than the SIR model, the possibility of reinfection can cause correlations between indirectly connected individuals to accrue over time. Consequently, moment-closure approximations on networks with SIS-dynamics are in general not exact, and their analytical tractability is limited. Of the progress that has been made: important formal results on their derivability from exact state-based models have been achieved by Taylor et al. (2012), Taylor and Kiss (2014) and Keeling et al. (2016) compare three systematic moment-closure approximations against stochastic simulations; (House et al., 2009) develop a motif-based approach that outperforms simpler methods for particular network topologies; while (Simon and Kiss, 2015) develop a compact pairwise approximation that agrees well with ODE models of a much higher dimensionality.

Capturing network structure is at its most important when edges between nodes are sparse but relatively long lasting. This, alongside the more well-defined nature of epidemiologically relevant contacts, means that moment-closure methods are potentially most valuable for understanding the spread of sexually transmitted infections (STIs). However, most STIs are modelled using the SIS-paradigm (though notably not HIV). Thus, both understanding the errors introduced by moment-closure approximations for diseases with SIS-dynamics, and improving upon these approximations, is vital for the successful application of such methods to public–health problems.

In this paper, we introduce improvements to the standard pairwise approximation for diseases with SIS-dynamics. In particular, we do this for the isolated open triple and for k-regular networks, by explicitly obtaining equations for the rates of change of the errors between triples and their standard pairwise approximation. By applying a closure to these equations, we obtain a closed set of equations that better approximate the true dynamics of infection, with only a modest increase in dimensionality. In the case of the isolated open triple, such a model is exact, while for k-regular networks, closures at the level of order-four structures have to be applied. Specifically, in Section 2 we discuss the isolated open triple, obtaining exact expressions for the appropriate errors and their rates of change, thus obtaining an exact set of equations describing the disease dynamics on this network topology. In Section 3, we use the results from the isolated open triple to inform our improved approximation on k-regular networks, i.e. networks with no loops and where each individual has k neighbours. We consider both higher-order moment-closure approximations and explicit stochastic simulations for this type of network, to act as benchmarks for our improved pairwise approximation. In Section 4, we compare this improved approximation to the standard pairwise approximation, to higher-order approximation models, and to stochastic simulations. In Section 5, we discuss some of the limitations to such an approach, and highlight some potential areas where we believe further research could be fruitful.

2. The isolated open triple

In this section, we consider the errors introduced by performing pairwise approximation on *isolated open triples* for a disease with SIS-dynamics. We define an isolated open triple as a central individual c connected to two neighbouring individuals x and y, where x and y remain unconnected, as illustrated in Fig. 1. By investigating this topology, the errors introduced by a pairwise approximation are not obfuscated by errors introduced from any external source, and exact results using the master equation approach (Kiss et al., 2017) can be generated.

We consider a diseases with SIS-dynamics, that is, upon recovery from infection (I) an individuals returns to the susceptible (S) class. We can described this process on the 3-network in terms of its states, of which there are eight - corresponding to whether each individual belongs to the S or I class - so a particular state $A \in \{S,I\}^3$. We denote the probability of being in a certain state $\mathbb{P}(x = X, c = C, y = Y)$ as $[X_xC_cY_y]$, where $X, C, Y \in \{S,I\}$. If we consider recovery from infection, γ, and transmission across partnerships, τ, to be Poisson processes, then the above situation is a continuous-time Markov process, and can be fully described by its Master equations (see Kiss et al., 2017, Chapter 2 for an introduction to this approach).

We set initial probabilities of each state by assuming random initial conditions, i.e. by taking $I_0 \sim U(0,1)$ and setting $[I_xI_cI_y]_0 = I_0 \times I_0 \times I_0$ and so on. Note, under this assumption, we have the symmetries $[S_xS_cI_y] = [I_xS_cS_y]$ and $[S_xI_cI_y] = [I_xI_cS_y]$. Thus, the dynamics of the isolated open triple are fully and exactly described by the following six ODEs:

Model 1 – The isolated open triple

$$\dot{[S_xS_cS_y]} = \gamma(2[S_xS_cI_y] + [S_xI_cS_y]) \tag{1}$$

$$\dot{[S_xS_cI_y]} = \dot{[I_xS_cS_y]} = \gamma([S_xI_cI_y] + [I_xS_cI_y] - [S_xS_cI_y]) - \tau[S_xS_cI_y] \tag{2}$$

$$\dot{[S_xI_cS_y]} = \gamma(2[S_xI_cI_y] - [S_xI_cS_y]) - 2\tau[S_xI_cS_y] \tag{3}$$

$$\dot{[S_xI_cI_y]} = \dot{[I_xI_cS_y]} = \gamma([I_xI_cI_y] - 2[S_xI_cI_y]) + \tau([S_xS_cI_y] + [S_xI_cS_y] - [S_xI_cI_y]) \tag{4}$$

$$\dot{[I_xS_cI_y]} = \gamma([I_xI_cI_y] - 2[I_xS_cI_y]) - 2\tau[I_xS_cI_y] \tag{5}$$

$$\dot{[I_xI_cI_y]} = -3\gamma[I_xI_cI_y] + 2\tau([S_xI_cI_y] + [I_xS_cI_y]) \tag{6}$$

Note that the disease-free state $[S_xS_cS_y]$ is absorbing, and so given long enough this system will always evolve to this state. Hence, without an external source of infection, a disease cannot persist indefinitely with an isolated open triple (or indeed, within any isolated graph of finite topology). If we wish to consider initial conditions that do not assume random mixing, e.g. pure initial conditions, eight equations are required. These are given in full in Appendix A.

2.1. The pairwise approximation for the isolated open triple

We now introduce the pairwise approximation for the open triple. It is important to note that we are considering a *local* moment-closure approximation, i.e. we are tracking the dynamics and errors introduced for a particular subgraph, as opposed to a *global* moment-closure approximation, where we apply closures at a population level.

We begin by considering equations for the probability of individuals (nodes of the open triple) being in a certain state $A \in \{S,I\}$, where we denote $\mathbb{P}(a = A)$ as $[A_a]$. ODEs describing the rate of change of these states can be obtained by summing the rates of change from the appropriate triples, e.g. $\dot{[S_x]} = \dot{[S_xS_cS_y]} + \dot{[S_xS_cI_y]} + \dot{[S_xI_cS_y]} + \dot{[S_xI_cI_y]}$. We observe that the state of an individual depends on the probability of pairs of individuals being in certain states: we denote $\mathbb{P}(a = A, b = B)$ as $[A_aB_b]$ and also

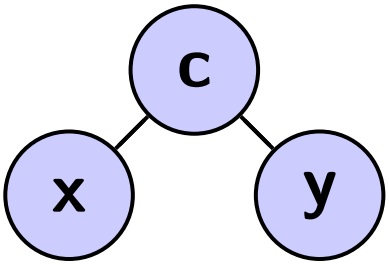


Fig. 1. Graphical representation of the isolated open triple. A central node c connected to two other nodes x and y. For the SIS model such triples have eight possible states.

obtain these by summing the appropriate triples. We arrive at the following equations:

$$[\dot{S_x}] = [\dot{S_y}] = \gamma[I_x] - \tau[S_xI_c] \tag{7}$$

$$[\dot{S_c}] = \gamma[I_c] - 2\tau[I_xS_c] \tag{8}$$

$$[S_x\dot{I}_c] = [I_c\dot{S}_y] = \gamma([I_xI_c] - [S_xI_c]) + \tau([S_xS_cI_y] - [S_xI_c]) \tag{9}$$

$$[I_x\dot{S}_c] = [S_c\dot{I}_y] = \gamma([I_xI_c] - [I_xS_c]) - \tau([I_xS_cI_y] + [I_xS_c]) \tag{10}$$

where $[I_a] = 1 - [S_a]$ and $[I_aI_b] = [I_b] - [S_aI_b] = [I_a] - [I_aS_b]$. Thus, we see that the rate of change of the probability of the infection status of individuals depends on the infection status of certain pairs, which themselves depend on the infection status of certain triples. This set of equations is unclosed, as we do not have expressions representing the time evolution of the disease status of these triples. Typically, studies have obtained a closed set of equations by assuming that the infection status of individuals x and y are conditionally independent given the infection status of individual c (Sharkey, 2008; Sharkey, 2011; Pellis et al., 2015). That is, we make the following assumption:

$$[S_xS_cI_y] \approx \frac{[S_xS_c][S_cI_y]}{[S_c]}\,[I_xS_cI_y] \approx \frac{[I_xS_c][S_cI_y]}{[S_c]} = \frac{[I_xS_c]^2}{[S_c]} \tag{11}$$

Observing that $[S_xS_c] = [S_x] - [S_xI_c]$, and that $[S_c] = [S_xS_c] - [I_xS_c]$, we obtain a closed set of three equations, which we refer to as the *pairwise approximation for the isolated open triple*, given in full below:

Model 2 – The pairwise approximation for the isolated open triple

$$[S_x\dot{S}_c] = [S_c\dot{S}_y] = \gamma([I_xS_c] + [S_xI_c]) - \tau\frac{[S_xS_c][S_cI_y]}{[S_c]} \tag{12}$$

$$[S_x\dot{I}_c] = [I_c\dot{S}_y] = \gamma([I_xI_c] - [S_xI_c]) + \tau\left(\frac{[S_xS_c][S_cI_y]}{[S_c]} - [S_xI_c]\right) \tag{13}$$

$$[I_x\dot{S}_c] = [S_c\dot{I}_y] = \gamma([I_xI_c] - [I_xS_c]) - \tau\left(\frac{[I_xS_c]^2}{[S_c]} + [I_xS_c]\right) \tag{14}$$

2.2. Quantifying errors

We can now compare the pairwise approximation model (Eqs. 12–14) to the exact model (Eqs. 1–6). The approximate model captures the dynamics of the system at low values of the transmission rate τ, but if τ is sufficiently high, the approximate model behaves qualitatively different to the exact model – there is no absorbing state, and we have a non-zero stationary probability of individuals being infected (Fig. 2). While in Model 1 $[S_xS_cS_y]$ never decreases, in Model 2 its approximation $[S_xS_c][S_cS_y]/[S_c]$ can decrease. This decrease occurs because of the rate of change of $[I_c]$ to $[S_c]$. In Model 1, this transition only affects $[S_xS_cS_y]$ from the state $[S_xI_cS_y]$, which only ever increases the probability of $[S_xS_cS_y]$. However, in Model 2 the decoupling of the two pairs and single means that this transition, with certain within pair correlations, can lead to a decrease in $[S_xS_c][S_cS_y]/[S_c]$.

Comparing the exact value for triples with their approximation at any given time, we observe this approximation underestimates the probability of the state $[I_xS_cI_y]$, and overestimates the probability of the state $[S_xS_cI_y]$. Indeed, the underestimate of $[I_xS_cI_y]$ is exactly the overestimate of $[S_xS_cI_y]$ (Fig. 3).

To understand why, consider the quantities $\alpha_{[S_xS_cI_y]} := [S_xS_cI_y][S_c] - [S_xS_c][S_cI_y]$ and $\alpha_{[I_xS_cI_y]} := [I_xS_cI_y][S_c] - [I_xS_c]^2$, borrowing notation from Sharkey et al. (2015), which quantify the difference between triples and their approximations. By expanding $[S_c] = [S_xS_cS_y] + 2[S_xS_cI_y] + [I_xS_cI_y]$, $[S_xS_c] = [S_xS_cS_y] + [S_xS_cI_y]$, and $[S_cI_y] = [S_xS_cI_y] + [I_xS_cI_y]$ and cancelling the appropriate terms, we arrive at the fact that both quantities are equal but opposite in sign, and thus we now define α_S as:

$$\alpha_S = \alpha_{[I_xS_cI_y]} = -\alpha_{[S_xS_cI_y]} = [S_xS_cS_y][I_xS_cI_y] - [S_xS_cI_y]^2 \tag{15}$$

Noting further that $\alpha_{[S_xS_cS_y]} = \alpha_S$, while clearly $\alpha_{[I_xS_cS_y]} = \alpha_{[S_xS_cI_y]} = -\alpha_S$, we observe that the difference between true and approximate triple values for all triples with susceptible central individuals depends upon one quantity α_S. Similarly, the difference between true and

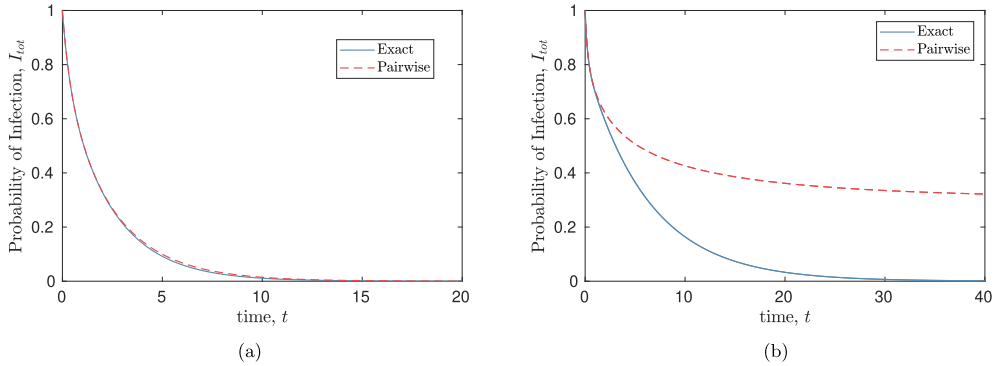


Fig. 2. Comparing exact and pairwise models for the isolated open triple. In (a), we see at low values of transmission between connected individuals ($\tau = 1$), the pairwise approximation (red) captures the probability of an individual being infected (given by $I_{tot} = ([I_x] + [I_c] + [I_y]/3)$ of the exact model (blue) reasonably well. In (b), we see that for higher values of τ (here $\tau = 3$), the pairwise model evolves to a non-zero stationary probability of individuals being infected, while the exact model always proceeds to the disease-free equilibrium. For all plots, we set $\gamma = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

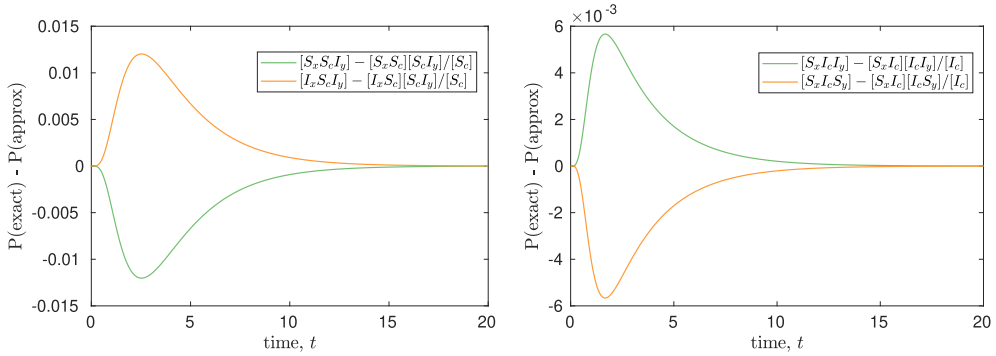


Fig. 3. Comparing exact and approximated probability of isolated open triples. In (a), we see that the approximation overestimates the probability of being in state $[S_xS_cI_y]$, and underestimates the probability of an isolated open triple being in state $[I_xS_cI_y]$. Similarly, (b) we see this approximation underestimates the probability of $[S_xI_cI_y]$ and overestimates the probability of $[S_xI_cS_y]$. In both cases, the overestimate of one is equal to the underestimate of the other. In both plots we set $\tau = 1, \gamma = 1$.

approximate triple values of all triples with infected central individuals depends only on one quantity, which we denote α_I:

$$\alpha_I = [I_xI_cI_y][S_xI_cS_y] - [S_xI_cI_y]^2 \tag{16}$$

2.3. Improving the pairwise approximation

If instead of using the approximations from Eq. (11), we let $[S_xS_cI_y] = ([S_xS_c][S_cI_y] - \alpha_S)/[S_c]$ in Eq. (9), and let $[I_xS_cI_y] = ([I_xS_c]^2 + \alpha_S)/]S_c]$ in Eq. (10), we obtain the rates of change of pairs in terms of singletons, pairs, and α_S. To obtain a closed set of equations, we must consider $\dot{\alpha_S}$, where the rates of change of triples can be obtained from the exact model.

$$\dot{\alpha_S} = [S_xS_c\dot{S_y}][I_xS_cI_y] + [I_xS_c\dot{I_y}][S_xS_cS_y] - 2[S_xS_c\dot{I_y}][S_xS_cI_y] \tag{17}$$
$$= \gamma(\phi_S - 2\alpha_S) - 2\tau\alpha_S \tag{18}$$
$$\text{where } \phi_S = [S_xS_cS_y][I_xI_cI_y] + [S_xI_cS_y][I_xS_cI_y] - 2[S_xS_cI_y][S_xI_cI_y] \tag{19}$$
$$= \frac{1}{[S_c][I_c]}(([S_xS_c][I_xI_c] - [I_xS_c][S_xI_c])^2 + \alpha_I[S_c]^2 + \alpha_S[I_c]^2) \tag{20}$$

Thus, the rate of change of α_S depends in turn on the rate of change of α_I, which is given by:

$$\dot{\alpha_I} = [I_xI_c\dot{I_y}][S_xI_cS_y] + [S_xI_c\dot{S_y}][I_xI_cI_y] - 2[S_xI_c\dot{I_y}][S_xI_cI_y] \tag{21}$$
$$= -4\gamma\alpha_I + 2\tau(\phi_I - \alpha_I) \tag{22}$$
$$\text{where } \phi_I = 2[S_xI_cS_y][I_xS_cI_y] - 2[S_xS_cI_y][S_xI_cI_y] \tag{23}$$
$$= \frac{2}{[S_c][I_c]}([S_xI_c]^2[I_xS_c]^2 - [S_xS_c][I_xS_c][S_xI_c][I_xI_c] + [I_xS_c][S_c]\alpha_I + [S_xI_c][I_c]\alpha_S) \tag{24}$$

We insist that ϕ_S and ϕ_I are 0 if either $[S_c] = 0$ or $[I_c] = 0$. Using the above equations, we arrive at a closed set of equations that describes exactly the disease dynamics of the open triple, without any reference to the particular states of triples themselves, by tracking the error terms α_S and α_I. Model 3 below describes in full this improved pairwise model, with ϕ_S and ϕ_I described as above:

Model 3 – Improved pairwise model of the isolated open triple

$$[S_x\dot{S_c}] = \gamma([I_xS_c] + [S_xI_c]) - \tau\frac{[S_xS_c][I_xS_c] - \alpha_S}{[S_c]} \tag{25}$$
$$[S_x\dot{I_c}] = \gamma([I_xI_c] - [S_xI_c]) + \tau\left(\frac{[S_xS_c][I_xS_c] - \alpha_S}{[S_c]} - [S_xI_c]\right) \tag{26}$$
$$[I_x\dot{S_c}] = \gamma([I_xI_c] - [I_xS_c]) - \tau\left(\frac{[I_xS_c]^2 + \alpha_S}{[S_c]} + [I_xS_c]\right) \tag{27}$$
$$\dot{\alpha_S} = \gamma(\phi_S - 2\alpha_S) - 2\tau\alpha_S, \text{ with } \phi_S \text{ as in Eq.(20)} \tag{28}$$
$$\dot{\alpha_I} = -4\gamma\alpha_I + 2\tau(\phi_I - \alpha_I), \text{ with } \phi_I \text{ as in Eq.(24)} \tag{29}$$

By including α_S and α_I and their time-evolution in Model 3, we obtain a system of ODEs that describes exactly the dynamics of the open triple. However, it is worth noting that this new model is of no lower dimensionality than Model 1. Despite this, we believe this is still a valuable model to have obtained explicitly. There are two principal reasons for this: firstly, by creating a system where errors α_S and α_I are tracked explicitly, we can obtain results and gain an understanding about the ways in which the standard pairwise approximation (which ignores the action of α_S and α_I) fails to capture the disease dynamics of the isolated open triple; and secondly, the derivation of this model informs our strategy of how to derive an improved pairwise approximation for k-regular networks, where there is a significant reduction in dimensionality.

Upon numerical evaluation, interesting results about the error terms α_S and α_I arise. When considering the whole state space, both error terms can be either negative or positive ($\alpha_S, \alpha_I \in [-1/4\ 1/4]$). However, this is not the case when starting from either random or pure initial conditions; in both scenarios, $\alpha_S \geqslant 0$ and $\alpha_I \leqslant 0$. This is numerically demonstrated in Appendix B. Consequently, assuming random or pure initial conditions, we arrive at the following bounds:

$$\frac{[S_xS_c]^2}{[S_c]} \leqslant [S_xS_cS_y] \qquad \frac{[I_xI_c]^2}{[I_c]} \geqslant [I_xI_cI_y]$$
$$\frac{[I_xS_c]^2}{[S_c]} \leqslant [I_xS_cI_y] \qquad \frac{[S_xI_c]^2}{[I_c]} \geqslant [S_xI_cS_y]$$
$$\frac{[S_xS_c][S_cI_y]}{[S_c]} \geqslant [S_xS_cI_y] \qquad \frac{[S_xI_c][I_cI_y]}{[I_c]} \leqslant [S_xI_cI_y]$$

Of these, the bound $[I_xI_c]^2/[I_c] \geqslant [I_xI_cI_y]$ is of particular interest. In previous moment-closure studies, it has been suggested heuristically that moment-closure models *underestimate* the probability of $[I_xI_cI_y]$ triples (Taylor et al., 2012). This does hold if the system is closed at the level of individuals, i.e. if we assume that the infection status of neighbours are independent. The above result demonstrates that the opposite is true if the system is closed at the level of pairs: $\mathbb{P}(x = I, y = I|c = I) \leqslant \mathbb{P}(x = I|c = I) \times \mathbb{P}(y = I|c = I)$.

For random initial conditions, α_S and α_I appear to be uniquely defined by the pairs $[S_xS_c]$ and $[S_cI_y]$, in other words α_S and α_I appear to be *functions* of $[S_xS_c]$ and $[S_cI_y]$. In theory, given values of $[S_xS_c]$ and $[S_cI_y]$, one could determine the values of α_S and α_I exactly, consequently reducing the dimensionality of Model 3, as equations for their time evolution would no longer be necessary. As α_S and α_I appear to be functions of two variables, they can be represented visually as surfaces, with $[S_xS_c]$ and $[S_cI_y]$ as x and y-axes, and with α_S or α_I as the z-axis. Included in supplementary material are animations of the evolution of the shape of these surfaces as we increase τ. These animations confirm the above bounds.

3. *k*-regular networks

In Section 2, we considered the accuracy of the standard pairwise approximation on the isolated open triple, and derived a closed exact set of equations describing the errors such an approximation makes. We could do so because we could compute exactly the probability of the states of the open triple (Model 1), and working backwards we could derive expressions for $\dot{\alpha}_S$ and $\dot{\alpha}_I$ solely in terms of $[S_xS_c]$, $[S_cI_y]$, α_S, and α_I - i.e. solely in terms of pairs and errors terms. Informed by these results, we move on to consider pairwise approximations for *k*-regular networks. *k*-regular networks are defined as networks in which each individual has *k* neighbours. Here, we consider *k*-regular networks which are infinite and contain no loops [1]. Being infinite, the disease dynamics on such a network cannot be described exactly by a closed set of ODEs, unless a closure at some level is exact, as in Sharkey et al. (2015) for diseases with SIR-dynamics. As stated previously, the possibility of reinfection induces correlations between distantly connected individuals, meaning the method used by Sharkey et al. (2015) is not successful for diseases with SIS-dynamics. However, one can close the system at a higher level than pairs and by doing so, we can obtain expressions for $\dot{\alpha}_S$ and $\dot{\alpha}_I$ solely in terms of pairs and error terms. While these are still approximations to the true disease dynamics on a *k*-regular network, doing so makes a considerable improvement on the standard pairwise approximation. This is the strategy we employ in this section.

While these *k*-regular networks are clearly idealisations far removed from any real-world sexual network, we believe that they are a useful example to study for a number of reasons. The impact of a small number of contacts, and the resulting dynamical correlations between non-adjacent individuals, is still relatively poorly understood (Keeling et al., 2016). In these idealised networks, the errors such correlations introduce into moment-closure approximations are at their most pronounced, and are not muddied by errors introduced from other sources, such as clustering or heterogeneity. While heterogeneity in the number of contacts individuals have is apparent in any real-world sexual network, and is important to capture when modelling STIs, the effect of heterogeneity has been studied extensively (Eames et al., 2002; Simon and Kiss, 2015), and can oftentimes be modelled by introducing multiple risk-groups into a mean-field approximation model (e.g. Edwards et al., 2010). Additionally, in the case of an infinite network, each individual has exactly the same properties, allowing us to bridge the gap from local to global moment-closure approximation.

In this section, we define *global* moment-closures for *k*-regular networks. That is, we define a closure in terms of population-level quantities rather than for the probabilities of particular individuals being in certain states. Accordingly, we use the notation $[S]$ to represent the *proportion* of individuals who are susceptible, $[SI]$ to represent the *proportion* of pairs where one individual is susceptible and one individual is infected, and so on. While it is standard within the moment-closure literature to refer to *numbers* of these quantities, we find that dealing with proportions avoids much of the combinatorial rigmarole involved, and has a more obvious correspondence with the methods described in Section 2. The following results hold true whether referring to proportions or numbers - in Appendix C, we provide a conversion table to transform the results from this section to numbers, and provide the model derived in this section in terms of numbers.

While the derivation of this moment-closure is independent to that of the previous section, and can be treated as a separate modelling exercise, we will observe that there are clear analogies between the two. This correspondence occurs because *k*-regular networks are isotropic - number of partnerships, as well as transmission and recovery rates, are homogeneous across the population. An alternative conceptualisation is that if we were to randomly sample one individual (or a higher-order motif) from a *k*-regular network, the probability of it being in a given state is directly equal to the proportion of the population in that state. Conversely, if we consider a population of infinitely many isolated open triples from Section 2, then the proportion in a given state is equal to the probability of one triple being in that state. Therefore while Section 2 is formulated in terms of probabilities and Section 3 is formulated in terms of proportions, we are effectively modelling interchangeable quantities.

3.1. Mean-field and pairwise approximations for k-regular networks

The following equation describes the rate of change of $[S]$ for any network (Simon et al., 2011):

$$[\dot{S}] = \gamma[I] - \lambda[SI] = \gamma(1-[S]) - \lambda[SI] \tag{30}$$

In the case of *k*-regular networks, $\lambda = k\tau$. By assuming the disease status of constituent individuals in pairs are uncorrelated, i.e. $[SI] \approx [S][I]$, we arrive at the mean-field approximation for the *k*-regular network, which is equivalent to the standard SIS-model:

Model 4 – The mean-field approximation for *k*-regular networks

$$[\dot{S}] = \gamma[I] - k\tau[S][I] = \gamma(1-[S]) - k\tau[S](1-[S]) \tag{31}$$

If instead we want to close the system at a higher-order moment, we must consider the rate of change of $[SI]$:

$$[\dot{SI}] = \gamma([II]-[SI]) - \tau[SI] + (k-1)\tau[SSI] - (k-1)\tau[ISI] \tag{32}$$

To close this system of equations, we must approximate the proportion of triples $[SSI]$ and $[ISI]$. We use the standard pairwise approximation of Rand (1999) and Keeling, 1999, commonly attributed to Kirkwood (1935). Using straight line brackets to denote *numbers* of individuals, etc. this is expressed as:

$$|ABC| \approx \frac{(k-1)}{k}\frac{|AB||BC|}{|B|} \iff [ABC] \approx \frac{[AB][BC]}{[B]} \tag{33}$$

When terms are expressed in terms of numbers this must be scaled by the factor $(k-1)/k$; this scaling factor disappears for *k*-regular networks when expressed in terms of proportions. This can be shown by converting either formulation of the approximation to the other using the conversion table provided in Appendix C. Using this approximation, we obtain:

Model 5 – The pairwise approximation for *k*-regular networks

$$[\dot{SS}] = 2\gamma[SI] - 2(k-1)\tau\frac{[SS][SI]}{[S]} \tag{34}$$

$$[\dot{SI}] = \gamma([II]-[SI]) - \tau[SI] + (k-1)\tau\frac{[SS][SI]}{[S]} - (k-1)\tau\frac{[SI]^2}{[S]} \tag{35}$$

where $[S] = [SS] + [SI]$, $[I] = 1-[S]$, $[IS] = [SI]$, and $[II] = 1 - [SS] - 2[SI]$.

3.2. Improving pairwise approximations for k-regular networks

Once again, we can look to improve the pairwise approximation by considering the rate of change of triples. Reintroducing subscripts (the position of individuals is illustrated in Fig. 4), the state of $x-c-y$ triples depend upon topologies consisting of four connected individuals: line graphs of length 4 $[A_aX_xC_cY_y]$ and $[X_xC_cY_yB_b]$, capturing the external force of infection acting upon

[1] diseases with SIS-dynamics on *k*-regular networks have been studied before are referred to in the theoretical literature as the contact process on the homogeneous tree T_{k-1} (Liggett, 2013).

individuals on the periphery of the triple, and star graphs with three outer individuals, $[X_xC_cY_yZ_z]$, capturing the external force of infection upon the central individual.

The rates of change for the triples in an infinite k-regular network, derived from House et al. (2009), are given in Appendix D. As before, we define α values as the difference between triple values and their standard pairwise approximation. Once again, the following relations hold:

$$[S] = [SSS] + 2[SSI] + [ISI] \tag{36}$$
$$[SS] = [SSS] + [SSI] \tag{37}$$
$$[SI] = [SSI] + [ISI] \tag{38}$$

Thus, as for the isolated open triple, the difference between triple values and pairwise approximations depend only upon two quantities: α_S and α_I, which are as defined in Eqs. (15) and (16). We can use the triple equations from Appendix D to obtain expressions for $\dot{\alpha_S}$ and $\dot{\alpha_I}$ for this type of network:

$$\dot{\alpha_S} = [S\dot{S}S][ISI] + [I\dot{S}I][SSS] - 2[S\dot{S}I][SSI] \tag{39}$$
$$= \gamma(\Phi_S - 2\alpha_S) + \tau(\beta_S - 2\alpha_S) \tag{40}$$
$$\text{where } \beta_S = 2(k-1)([I_aS_xS_cI_y][SS] - [I_aS_xS_cS_y][SI]) \tag{41}$$
$$+ (k-2)(2[S_xS_cI_yI_z][SSI] - [S_xS_cS_yI_z][ISI] - [I_xS_cI_yI_z][SSS])$$

$$\dot{\alpha_I} = [I\dot{I}I][SIS] + [S\dot{I}S][III] - 2[S\dot{I}I][SII] \tag{42}$$
$$= -4\gamma\alpha_I + \tau(\beta_I + 2\Phi_I - 2\alpha_I) \tag{43}$$
$$\text{where } \beta_I = 2(k-1)([I_aS_xI_cI_y][SI] - [I_aS_xI_cS_y][II]) \tag{44}$$
$$- (k-2)(2[S_xS_cI_yI_z][SII] - [I_xS_cI_yI_z][SIS] - [S_xS_cS_yI_z][III])$$

Despite being calculated for triples within a k-regular network, we find that $\Phi_S = \phi_S$ and $\Phi_I = \phi_I$ as previously defined for the isolated open triple in Eqs. (20) and (24), and so use the ϕ_S and ϕ_I terms henceforth. We therefore obtain a closed set of equations by once again setting

$$[ABA] \approx \frac{[AB]^2 + \alpha_B}{[B]}[ABC] \approx \frac{[AB][BC] - \alpha_B}{[B]} \tag{45}$$

But now we must also make some approximation for order four terms. We do this by making the following closures:

$$[A_xS_cB_yI_z] \approx \frac{[ASB][BSI][ASI][S]}{[AS][BS][IS]} \tag{46}$$
$$[I_aS_xA_cB_y] \approx \frac{[ISA][SAB]}{[SA]} \tag{47}$$

Thus, we can again express $\dot{\alpha_S}$ and $\dot{\alpha_I}$ as (complicated) functions of $[SS], [SI], \alpha_S$ and α_I. Using this, we arrive at a system of four ODEs, which we call the improved pairwise approximation for k-regular networks:

Model 6 – The improved pairwise approximation for k-regular networks

$$[S\dot{S}] = 2\gamma[SI] - 2(k-1)\tau\frac{[SS][SI] - \alpha_S}{[S]} \tag{48}$$

$$[S\dot{I}] = \gamma([II] - [SI]) - \tau[SI] + (k-1)\tau\frac{[SS][SI] - \alpha_S}{[S]} \tag{49}$$
$$- (k-1)\tau\frac{[SI]^2 + \alpha_S}{[S]} \tag{50}$$
$$\dot{\alpha_S} = \gamma(\phi_S - 2\alpha_S) + \tau(\beta_S - 2\alpha_S) \tag{51}$$
$$\dot{\alpha_I} = -4\gamma\alpha_I + \tau(\beta_I + 2\phi_I - 2\alpha_I) \tag{52}$$

where ϕ_S and ϕ_I are defined as in Section 2.

3.3. Higher-order moment-closure approximations

To assess the accuracy gained by modelling the error terms α_S and α_I, we compare our model to higher-order moment-closures. The first of these we refer to as a *neighbourhood* closure, previously described by Lindquist et al. (2011) and Keeling et al. (2016), where we model a central individual and their number of infected neighbours explicitly. This system is described by $2 \times (k+1)$ ODEs. The second of these we refer to as an *extended triple closure*, where we explicitly model a central triple and every neighbour of this triple. This system is described by 2^{3k-1} equations (though its dimensionality can be reduced by accounting for symmetries). In both cases, we approximate the external force of infection on outer individuals by exploiting the symmetry of the topology of the k-regular network. While each model is still an approximation towards the true dynamics of a k-regular network, in virtue of closing the system at a higher order, these models are expected to have a greater accuracy. From these higher-order models, we can also obtain estimates of the terms α_S and α_I, with which we can compare the α terms obtained from the improved pairwise model for the k-regular network (Model 6).

3.3.1. The neighbourhood closure

For the neighbourhood closure, we model a central individual and their number of infected neighbours explicitly. Visually then, we are modelling a star topology. The rate of change of state of the 'star' will depend upon both the internal configurations and the immediate neighbours of the star. We show this visually in Fig. 5. To close this system of equations, we make the assumption that the configuration of two overlapping 'stars' are conditionally independent given the infection status of the two shared individuals of the combined configuration. As we only need to consider the effect of an external force of infection if the relevant neighbour is susceptible (S), there are only two quantities relevant to the external force of infection on that individual, depending on the infection status of the original central individual (S or I), which denote λ_S and λ_I accordingly. These terms are constructed by summing all configurations of the external neighbours including an infected individuals, multiplied by the number of infected external neighbours in that configuration, divided by the sum of all possible configurations of external neighbours. Denoting a central individual in state $A \in \{S, I\}$ with $i \in \{0, 1, \ldots, k\}$ infected neighbours as $[A_i]$, the neighbourhood model can thus be described by the following set of equations:

Model 7 – The neighbourhood approximation model for k-regular networks

$$[\dot{S_i}] = \begin{cases} 0, & i < 0 \\ \gamma([I_i] + (i+1)[S_{i+1}]) + \tau((k-i+1)[S_{i-1}] - i[S_i]) + \lambda_S((k-i+1)[S_{i-1}] - (k-i)[S_i]), & 0 \leqslant i \leqslant k \\ 0, & i > k \end{cases} \tag{53}$$

$$[\dot{I_i}] = \begin{cases} 0, & i < 0 \\ \gamma(i+1)([I_{i+1}] - [[I_i]) + \tau((k-i+1)[I_{i-1}] + i[S_i]) + \lambda_I((k-i+1)[I_i] - (k-i)[I_i]), & 0 \leqslant i \leqslant k \\ 0, & i > k \end{cases} \tag{54}$$

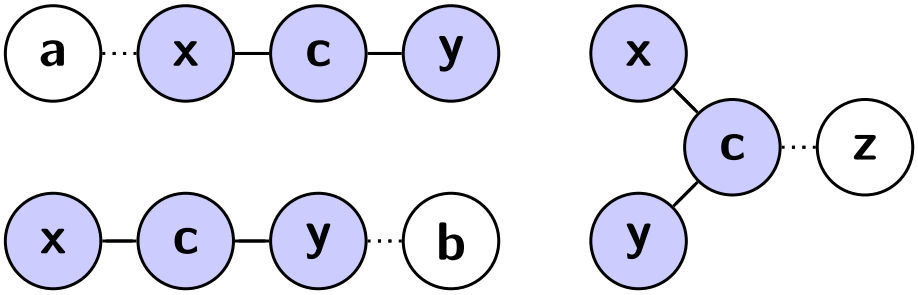


Fig. 4. Dependence on order-four structures in a k-regular network. The state $[X_xC_cY_y]$ of our triple of interest (shaded in blue) depends on the state of two order-four network structures – length four line-graphs $[A_aX_xC_cY_y]$ and $[X_xC_cY_yB_b]$ and the 'star' graph with three outer individuals $[X_xC_cY_yZ_z]$. The positions of a, b, and z relative to the triple of interest are shown visually here. N.B. that $[X_xC_cY_yB_b] \equiv [B_aY_xC_cX_y]$, meaning only one length-four line graph term is necessary in the equations below.

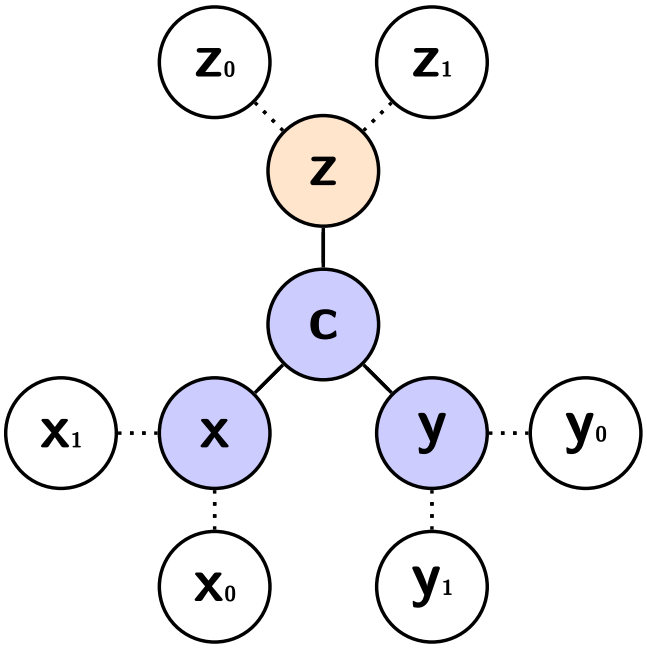


Fig. 5. The external force of infection on a neighbourhood. Here we illustrate the external force of infection on a neighbourhood in the neighbourhood approximation for k-regular networks, for the example of $k = 3$. Shaded blue is our triple of interest, shaded in orange are any additional individuals that are modelled explicitly, while shaded in white are individuals not explicitly modelled who exert a force of infection on the explicitly modelled neighbourhood. In this approximation, we model a central individual c, and the number of infected neighbours c as (here shown by x, y, and z). The external force of infection on the explicitly modelled neighbourhood will depend upon order-six structures: $[X_xC_cY_yZ_zX_{0_{x_0}}X_{1_{x_1}}]$, $[X_xC_cY_yZ_zY_{0_{y_0}}Y_{1_{y_1}}]$, and $[X_xC_cY_yZ_zZ_{0_{z_0}}Z_{1_{z_1}}]$. To close the system, we make the approximation that, e.g. $[X_xC_cY_yZ_zX_{0_{x_0}}X_{1_{x_1}}] \approx ([X_xC_cY_yZ_z] \times [X_xC_cX_{0_{x_0}}X_{1_{x_1}}])/[X_xC_c]$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where λ_S and λ_I are given by:

$$\lambda_S = \tau \frac{\sum_{i=0}^{k-1} i(k-i)[S_i]}{\sum_{i=0}^{k-1} (k-i)[S_i]} \tag{55}$$

$$\lambda_I = \tau \frac{\sum_{i=0}^{k-1} i(i+1)[S_{i+1}]}{\sum_{i=0}^{k-1} (i+1)[S_{i+1}]} \tag{56}$$

To obtain estimates for α_S (and α_I) from this model, we must derive the proportion of triples implied by the assumptions of the neighbourhood model. This can be calculated as follows. For a given triple $[XCY]$, we let l indicate whether X and Y are infected. (If $X = Y = S$, $l = 0$. If $X = S$ and $Y = I$, or $X = I$ and $Y = S$, $l = 1$. If $X = Y = I$, $l = 2$.) In the neighbourhood model we explicitly model a central individual and the number of its k immediate neighbours who are infected. $[XCY]$ will occur as subgraphs of configurations that comprise C_l to C_{k+l-2}. Assuming there are i additional infected individuals surrounding c (in addition to those specified by X and Y), there are $\binom{k}{i+l}$ different configurations such that a central individual C has $i + l$ infected neighbours. Of these, there are $\binom{k-2}{i}$ configurations once the position of the [XCY] subgraph is determined, as there are $k - 2$ positions left to fill with i infected individuals. Hence $\binom{k-2}{i} / \binom{k}{i+l}$ of $[C_{i+l}]$ contain $[XCY]$, and so we arrive at the formula:

$$[XCY] = \sum_{i=0}^{k-2} \frac{\binom{k-2}{i}}{\binom{k}{i+l}} [C_{i+l}] \tag{57}$$

3.3.2. The extended triple closure

For the extended triple closure, we model a triple and each of its neighbours explicitly, requiring 2^{3k-1} equations. The state of this system will depend on the infection status of the neighbours of these neighbours, i.e. the rate of change of states in the extended triple depend upon order $4k - 2$ configurations (illustrated in Appendix E). We approximate these external forces on the extended triple by assuming that the state of these higher-order structures amount to overlapping extended triple topologies conditionally independent given the state of their shared individuals, of which there are $2k$. A detailed explanation of the extended triple closure model is provided in Appendix E.

3.4. Stochastic simulations

We use explicit stochastic simulations as our final benchmark for the accuracy of our approximate models. It is not computationally possible to construct infinite loopless networks for simulations. Instead, large random graphs where each individual has k neighbours can be constructed using the Molloy-Reed algorithm (Molloy et al., 1995), which should behave similarly for very large network sizes. We use the methods outlined by Keeling et al., 2016 to remove short loops and to efficiently calculate the quasi-equilibrium prevalence of infection.

4. Comparing models

In this section, we compare the previous described k-regular network models; in order of dimensionality, these are: the mean-field approximation model (Model 4), the pairwise approximation model (Model 5), the improved pairwise approximation model (Model 6), the neighbourhood approximation model (Model 7), and the extended triple approximation model. As we are considering a disease with SIS-dynamics, the models evolve to an endemic prevalence of infection (given a sufficiently high transmission rate) - we use this as the primary metric for model comparison. All of these models are approximations of the true system, where there are infinitely many individuals, but we expect as we increase the dimensionality of approximation we also increase the accuracy of the model. We compare all approximate models to explicit stochastic simulations on networks of 10,000 individuals.

In Fig. 6, we compare the endemic prevalence generated by the four models that do not explicitly model α to stochastic simulations – the improved pairwise approximation (which utilises the dynamics of α) is considered in Figs. 7 and 8. While we notice large differences between mean-field and pairwise models, the difference in prevalence between models decreases as we increase the dimensionality of the model. For $k = 3$, there is little difference between the neighbourhood and extended triple approximation models, and there is excellent agreement between the extended triple model and stochastic simulation. For $k = 4$ and $k = 10$, the extended triple model is omitted, as the neighbourhood approximation models match closely to stochastic simulations. This indicates that including further complexity into a model may be unnecessary, or may not be worth the increasing complexity or computational expense. For $k = 2$, there is still a significant difference between simulation and the extended triple model. However, this is unsurprising, as previous research (Keeling et al., 2016) has shown that errors persist when much larger neighbourhoods are modelled explicitly. Fig. 6 also illustrates that as we increase k, models tend towards the mean-field approximation (which can be considered the $k \to \infty$ limit). In Appendix F, we provide a proof of this for the pairwise model, and outline how this would be proved in the general case. We also see that as we increase k, the difference between pairwise and neighbourhood approximation models decreases, although the pairwise model consistently predicts higher endemic prevalences.

Now, we turn our attention to the improved pairwise approximation (Model 6), which tracks the errors α_S and α_I explicitly. Here we focus on the examples $k = 2$ and $k = 3$, though comparable results are found for all higher values of k. The error in our pairwise model depends on only one term: α_S. This term captures the error between the 'true' value of triples and the standard pairwise approximation of their values. We can obtain estimates for α_S from each of our higher-order models, noting that the improved pairwise approximation (Model 6) is based on consideration of four connected nodes. Comparing α_S between models allows us to assess the extent to which the improved pairwise approximation is successful in capturing the errors introduced to the pairwise approximation induced by dynamics of higher-order structures. By plotting α_S as a function of the pairs $[SS]$ and $[SI]$ we obtain

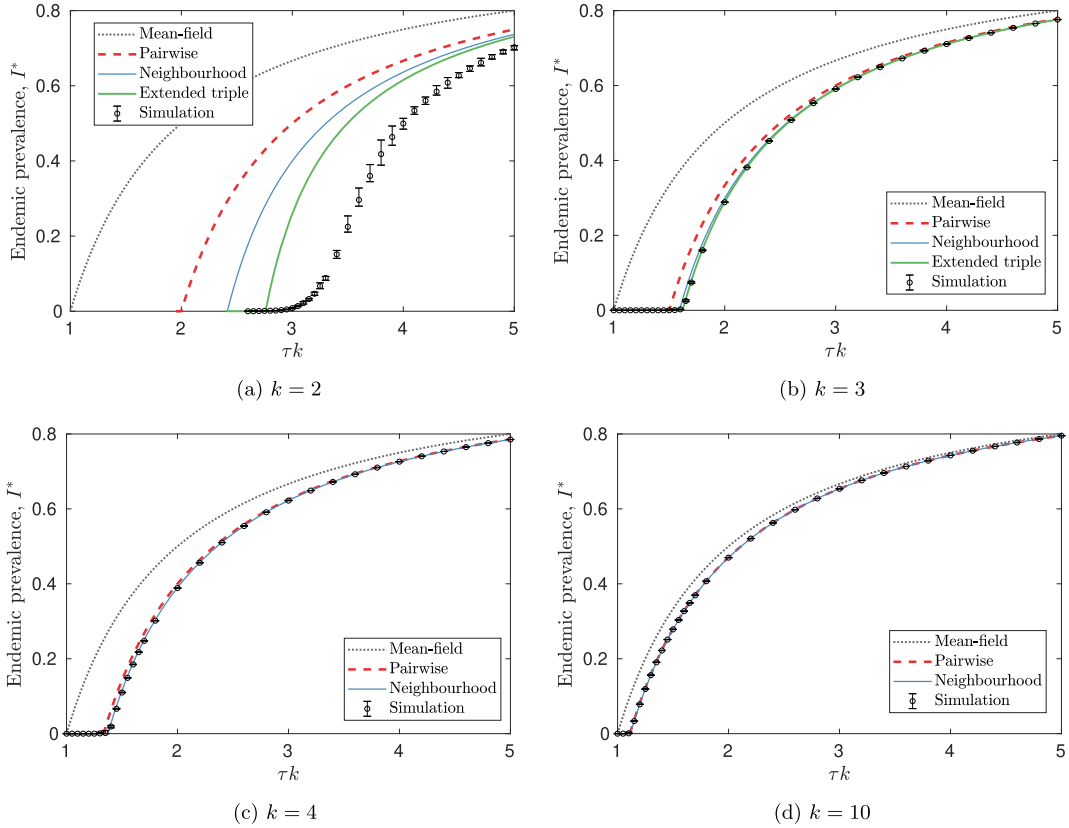


Fig. 6. Comparing approximate models for k-regular networks. Here we compare endemic prevalence (I^*) against $\lambda = \tau k$ for mean-field (grey, dotted), pairwise (red), neighbourhood (blue), extended triple (green) approximations for k-regular networks against explicit stochastic simulations (points) for a) $k = 2$, (b) $k = 3$ (c) $k = 4$, and (d) $k = 10$. For $k = 3$ simulations are matched well by the extended triple model, while for $k > 3$ simulations are matched well by the neighbourhood model. As k increases, all models move closer to the mean-field approximation, and the difference in I^* for a given λ between approximate models decreases. For stochastic simulations, each I^* point is calculated as the average of 150 runs, and error bars indicate 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

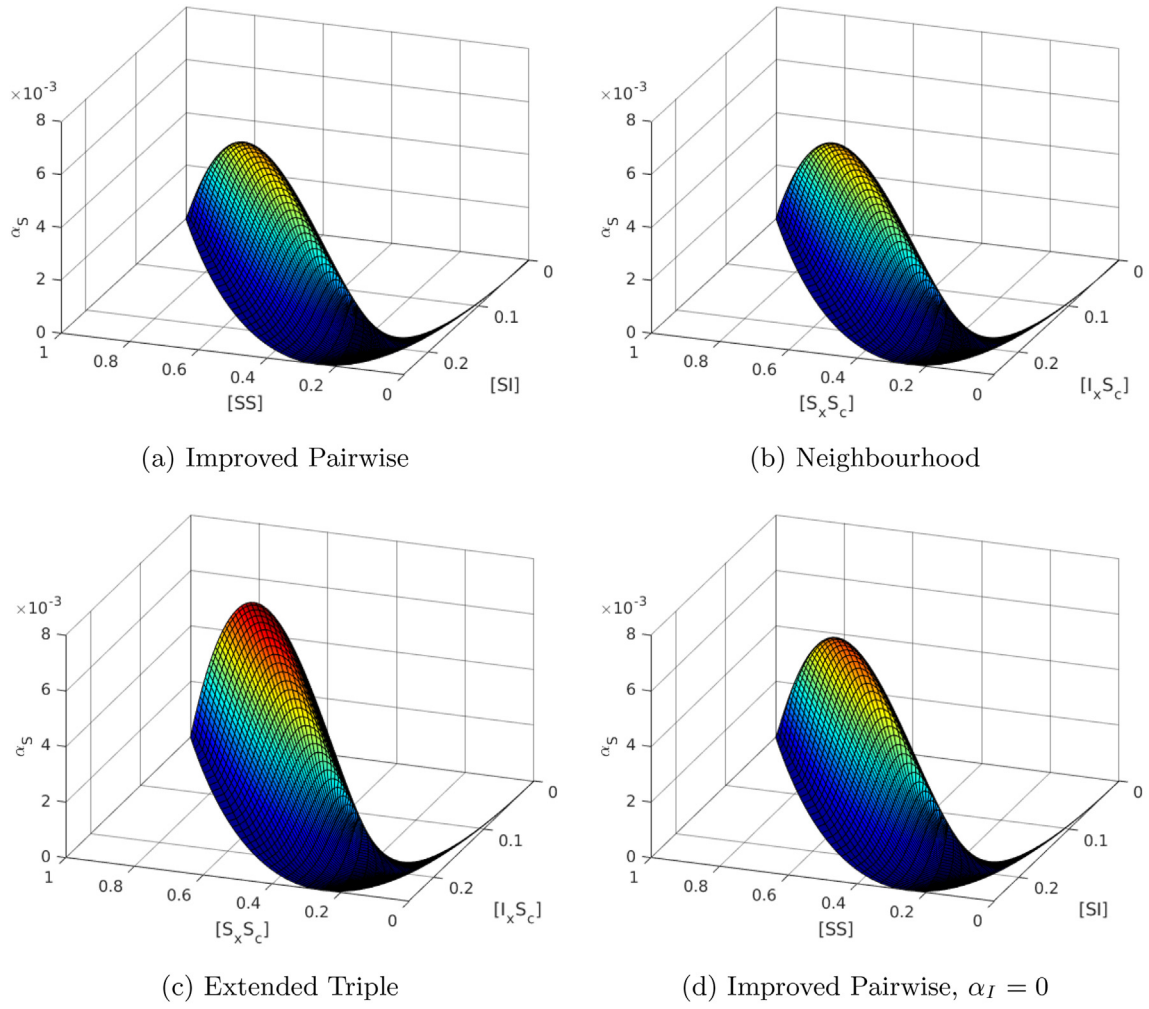
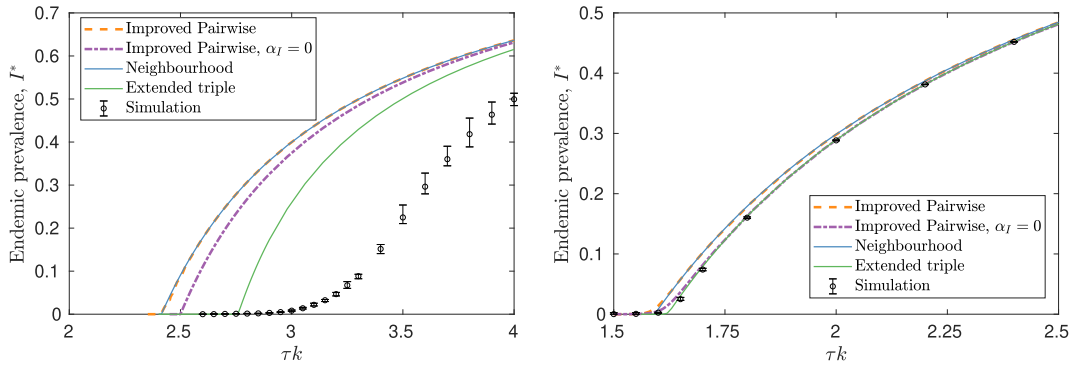


Fig. 7. Exploring the shape of α_S for different approximate models. Here we compare the shape of the error term α_S as a function of $[SS]$ and $[SI]$ for improved pairwise models ((a) and (d)), and as a function of $[S_xS_c]$ and $[I_xS_c]$ for neighbourhood and extended triple approximations ((b) and (c)), for the example $k = 3$. We observe that α_S in the improved pairwise (a) and the neighbourhood (b) approximation models are extremely similar, but that the improved pairwise approximation model underestimates this error compared to the extended triple approximation (c). By assuming $\alpha_I = 0$ and $\dot{\alpha}_I = 0$ (d), the resulting α_S surface more closely resembles that of the extended triple model. In all plots, we set $\tau = 1, \gamma = 1$.

surfaces; their shape informing our intuition of the behaviour of α_S as we move through $([SS], [SI])$-space. Firstly, we observe that the numerical result $\alpha_S \geqslant 0$that was true for the isolated open triple also holds true for each of these models (numerically demonstrated in Appendix G). Hence, the bounds obtained for triples $[S_xS_cS_y]$, $[I_xS_cI_y]$, and $[S_xS_cI_y]$ in Section 2 for the isolated open triple also hold for k-regular networks. Secondly, we observe the similarity between α_S surfaces obtained from the improved pairwise and neighbourhood approximation models. We do, however, see these are smaller than α_S from the extended triple. In other words, models that include higher-order correlations, such as the extended triple, have higher values of α_S than are obtained from the improved pairwise model.

Comparing the prevalence of infection obtained from these models, we observe only a minor difference between improved pairwise and neighbourhood approximations. By including just two more equations (for α_S and α_I), we arrive at a model with an endemic prevalence much closer to results obtained from stochastic simulation, with only a marginal increase in dimensionality. Unlike the isolated open triple, α_I can be positive when $k > 2$ in each of the approximate models. However, this only occurs at very high transmission rates - typically when endemic prevalence $I^* > 0.8$ (Appendix G).

In an attempt to further improve the accuracy, and to reduce the dimensionality, of the model, we consider the effect of ignoring α_I on the shape of α_S in the improved approximation; noting that the values of α_S from the extended triple approximation are consistently larger than from the other lower-order approximations (Fig. 7). We do this by setting $\alpha_I = 0$, which is equivalent to using the standard pairwise approximation for triples with infected central individuals. This is in part justified by the fact that values of α_I are typically much smaller in magnitude than α_S (Appendix H). This assumption further reduces the dimensionality of the system, as we have one less variable. Moreover, as α_I is typically $\leqslant 0$, ignoring it will increase $\dot{\alpha}_S$, meaning we will generate higher values of α_S. (Positive values of α_I can only occur at very high values of τ; at such values, the disease dynamics on the k-regular network are already well approximated by the standard pairwise approximation). Indeed, comparing shapes of α_S (Fig. 7), we see this assumption provides a closer match to the values from the extended triple. In Fig. 8, we compare the endemic prevalence obtained using this $\alpha_I = 0$ assumption against the extended triple approximation, as well as against the improved pairwise approximation where α_I is a dynamic variable. Ignoring α_I provides an estimate closer to the extended triple approximation than accounting for α_I explicitly,



(a) $k = 2$ (b) $k = 3$

Fig. 8. Comparing improved pairwise approximations against higher-order approximations for $k = 2$ and $k = 3$-regular networks. We compare endemic prevalence I^* obtained from improved pairwise model (full in orange; $\alpha_I = 0$ in purple) against neighbourhood (blue) and extended triple (green) approximations, as well as against explicit stochastic simulations (points), as we vary $\lambda = \tau k$, for (a) $k = 2$ and (b) $k = 3$-regular networks. In both (a) and b) I^* obtained from the improved pairwise approximation is very similar to I^* obtained from the neighbourhood approximation. By assuming $\alpha_I = 0$ and $\dot{\alpha}_I = 0$, the dynamics of the improved pairwise approximation are closer to those of the extended triple approximation, and match I^* from stochastic simulations well for $k = 3$. For all models we set $\gamma = 1$. For stochastic simulations, each I^* point is calculated as the average of 150 runs, and error bars indicate 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which in the case of $k = 3$ matches stochastic simulations closely. In Appendix I we consider the time evolution of these models.

5. Discussion

Whenever detailed information on underlying network structure is available, detailed stochastic simulation of an epidemic on a network is always the 'gold standard' for any real-world application. In the absence of such information, moment-closure approximation methods for the spread of infections promise relatively simple models that allow us to understand the effect of network structure on the dynamics of an epidemic. The success of such a method, however, depends upon understanding the errors introduced by moment-closure approximations, and upon refinements that minimise such errors. While this approach has been successfully applied to diseases with SIR-dynamics, the dynamic build-up of correlations between distant individuals for diseases with SIS-dynamics means success for infections with this natural history has been more limited. However, as the dynamics of most STIs can be well approximated by the SIS-paradigm, and given the importance of network structure in this case, further research into this area is paramount. Indeed, there is already a considerable body of literature concerning moment-closure approximations for SIS-dynamics (Taylor et al., 2012; Taylor and Kiss, 2014; Keeling et al., 2016; House et al., 2009; Simon and Kiss, 2015), as well as other network approaches to diseases of this type (Floyd et al., 2012; Lee et al., 2013; Wilkinson and Sharkey, 2013), demonstrating this as an active research area.

This study improves upon the standard pairwise approximation by explicitly tracking the errors between the 'true' value of triples and their estimate from this approximation. We show that these errors are fully described by the quantity α_S for triples with susceptible central individuals, and by the quantity α_I for triples with infected central individuals. By tracking the time-evolution of these error terms, we improve upon the standard pairwise approximation by incorporating these terms into the modelling framework. For the isolated open triple (just three individuals connected in a line), both $\dot{\alpha}_S$ and $\dot{\alpha}_I$ are exactly described as functions of $[S_x S_c], [I_x S_c], \alpha_S$ and α_I; hence, in this case, the improved pairwise model is itself exact. For k-regular networks, $\dot{\alpha}_S$ and $\dot{\alpha}_I$ depend upon order-four structures. However, by approximating the prevalence of these structures via higher order moment-closures, we obtain expressions for $\dot{\alpha}_S$ and $\dot{\alpha}_I$ solely in terms of pairs, α_S and α_I. While such a model is not exact, explicitly modelling the time-evolution of these errors markedly improves upon the standard pairwise approximation for k-regular networks, obtaining prevalence estimates comparable both to models closed at even higher orders and to explicit stochastic simulations.

The findings of this paper contribute towards understanding the shape and direction of errors introduced by pairwise approximations. We show that the errors between triples and their standard approximation are quantified by just two values: α_S and α_I. Interestingly, we find numerically that $\alpha_S \geqslant 0$ and $\alpha_I \leqslant 0$, which inform us as to whether the standard pairwise approximation underestimates or overestimates the proportion of certain triples. While both bounds hold for the isolated open triple, only $\alpha_S \geqslant 0$ holds in general for k-regular networks. This result also appears to hold for the constituent triples of all other investigated topologies (line graphs up to length 10, star graphs with up to 10 neighbouring individuals, the extended triple with no external force of infection), while the result $\alpha_I \leqslant 0$ only appears to apply when central individuals in a triple have no other connections outside of the triple. We hence believe that an analytical exploration of such bounds could be fruitful, and would make an important contribution to this research area if such bounds could be proven generally. A deeper understanding of the shape, direction, and magnitude of such error terms is not only of interest to those concerned with using the improved pairwise approximation model described in this paper, but to any researcher interested in applying the standard pairwise approximation to a network model of a disease where recovery from infection does not lead to immunity.

In this paper, we compare approximations to the dynamics of k-regular networks closed at increasingly higher levels of complexity – from individual, to pair, to neighbourhood, to an extended neighbourhood. As we increase the dimensionality of a model, we expect to obtain more accurate results. On the other hand, models of high dimensionality are difficult to understand intuitively and are much more computationally expensive. Whether including such complexity is worthwhile depends on the task at hand. We believe that our improved pairwise approximation provides a reasonable compromise between intuition and complexity - this model is still described by a small number of ODEs, and has dynamics closely resembling those from the model closed at the level of neighbourhoods, more closely matching prevalence estimates obtained from

stochastic simulations. An unexpected result is that by ignoring α_I, i.e. using the standard pairwise approximation for triples with infected central individuals, one appears to obtain a better approximation to the true dynamics. It is important to establish if such a result holds generally, and if so why, or whether this result is a spurious convenience for k-regular networks.

The results here consider the two most ideal networks: the isolated open triple is the simplest possible network topology including three individuals, while in k-regular networks each individual has exactly k neighbours and there are no closed loops within the network. We consider these idealisations as it is in these networks that network structure is most dominant and the errors introduced by moment-closure approximations are most pronounced. But this means there is fertile ground for further exploration on both local and global scales. On a local scale, a taxonomy of the errors that occur for a variety of different small topologies, as has been done by Pellis et al. (2015) for diseases with SIR-dynamics, would be useful contribution to understanding the impact of local moment-closures for diseases with SIS-dynamics. On a global scale, understanding whether tracking the dynamics of error terms explicitly would be worthwhile in heterogeneous networks (building upon the work of Simon and Kiss (2015)), and assessing whether the same techniques can be applied in the presence of clustering, are important next steps.

This paper makes three assumptions common to the literature on the mathematics on epidemics on networks: first, that epidemiologically relevant contacts (the edges between nodes) are fixed throughout the epidemic and not dynamic; second, that these contacts are identical in kind, such that probability of infection for an individual from any partner of theirs is equal to any other partner; third, that individuals have exponentially distributed periods of infection (the Markovian assumption). Each of these are in some senses unrealistic: people's sexual partnerships change over time (it is a question of theoretical importance the extent to which the dynamics of epidemics on dynamics networks can by approximated by the dynamics of epidemics on static networks, which has begun to be explored (Volz and Meyers, 2007; Bansal et al., 2010)); for individuals in more than one partnership, the frequency of sexual contact will be different for each partnership, hence the probability of transmission across partnerships will also be different; whilst periods of infection may be better modelled as having a constant duration. For SIR-dynamics, a variety of dynamic network models incorporating moment-closure approximations, or other low-dimensional ODE models have been developed (Ball and Neal, 2008; Volz, 2008). So too are there a variety of dynamic network models for SIS-dynamics (e.g. Bauch and Rand, 2000; Leng and Keeling, 2018). Incorporating improved moment-closure approximations into such models, and exploring how the introduction of partnership formation and dissolution effects the errors introduced, are important next steps. While studies into the contribution of steady and casual partnerships to the spread of STIs has been explored (Xiridou et al., 2003; Hansson et al., 2019), heterogeneity in edge type is an underexplored topic for moment-closure approximations, even for diseases with SIR-dynamics. Assuming constant periods of infection, instead of making a Markovian assumption, can make closures exact for different network topologies in the case of SIR-dynamics (Pellis et al., 2015). Exploring this alternative assumption and its effect on errors α_S and α_I may prove interesting avenues of research.

With regards to modelling the spread of STIs, it is clear that research should continue to develop more realistic and more sophisticated stochastic simulations. However, we believe that approximate methods have an important role to play, in both developing an intuitive understanding of the effect of network structure on the fate of the spread of STIs, and as a benchmark to compare such simulations against. It is in this context that improving the accuracy of such approximate methods is paramount, and it is in this context that we believe we make a valuable contribution to the literature.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the Engineering and Physical Sciences Research Council and the Medical Research Council for their funding through the MathSys CDT (grant EP/L015374/1). We also thank the anonymous reviewers of this paper for their careful and considered comments that have led to a much improved manuscript.

Appendix A. Appendices

A.1. Appendix A – An exact model for the open triple

Not assuming random mixing at initial conditions, eight ODEs are required to describe the dynamics of the open triple exactly - these are given below.

$$\dot{[S_xS_cS_y]} = \gamma([S_xS_cI_y] + [I_xS_cS_y] + [S_xI_cS_y]) \tag{58}$$

$$\dot{[S_xS_cI_y]} = \gamma([S_xI_cI_y] + [I_xS_cI_y] - [S_xS_cI_y]) - \tau[S_xS_cI_y] \tag{59}$$

$$\dot{[I_xS_cS_y]} = \gamma([I_xI_cS_y] + [I_xS_cI_y] - [I_xS_cS_y] - \tau[I_xS_cS_y] \tag{60}$$

$$\dot{[S_xI_cS_y]} = \gamma([S_xI_cI_y] + [I_xI_cS_y] - [S_xI_cS_y]) - 2\tau[S_xI_cS_y] \tag{61}$$

$$\dot{[S_xI_cI_y]} = \gamma([I_xI_cI_y] - 2[S_xI_cI_y]) + \tau([S_xS_cI_y] + [S_xI_cS_y] - [S_xI_cI_y]) \tag{62}$$

$$\dot{[I_xI_cS_y]} = \gamma([I_xI_cI_y] - 2[I_xI_cS_y]) + \tau([I_xS_cS_y] + [S_xI_cS_y] - [I_xI_cS_y]) \tag{63}$$

$$\dot{[I_xS_cI_y]} = \gamma([I_xI_cI_y] - 2[I_xS_cI_y]) - 2\tau[I_xS_cI_y] \tag{64}$$

$$\dot{[I_xI_cI_y]} = -3\gamma[I_xI_cI_y] + \tau([S_xI_cI_y] + [I_xI_cS_y] + 2[I_xS_cI_y]) \tag{65}$$

A.2. Appendix B – Demonstrating $\alpha_S \geqslant 0$ and $\alpha_I \leqslant 0$ for different initial conditions on the isolated open triple

See Fig. 9.

A.3. Appendix C – Converting the improved pairwise model from proportions to numbers

While we find that considering proportions is a more convenient way to express the results from Section 3, we appreciate that others may prefer to use our results under the convention of terms referring to numbers of motifs. In this appendix we provide a conversion table to transform the terms from this section from proportions to numbers, and derive the improved pairwise model in terms of numbers. We stress that the improved pairwise model presented here is *equivalent* to the model presented in Section 3.

First, to express the quantities in Section 3 in terms of numbers, we must be able to count the number of motifs relative to every individual. In a k-regular network, for every individual there are k pairs, for every pair there are $(k-1)$ triples, and for every triple there are $(k-1)$ line graphs of length 4 and $(k-2)$ 4-stars. Using straight line brackets $|X|$ to denote the *number* of individuals in state X etc. Table 1 below outlines equivalent terms:

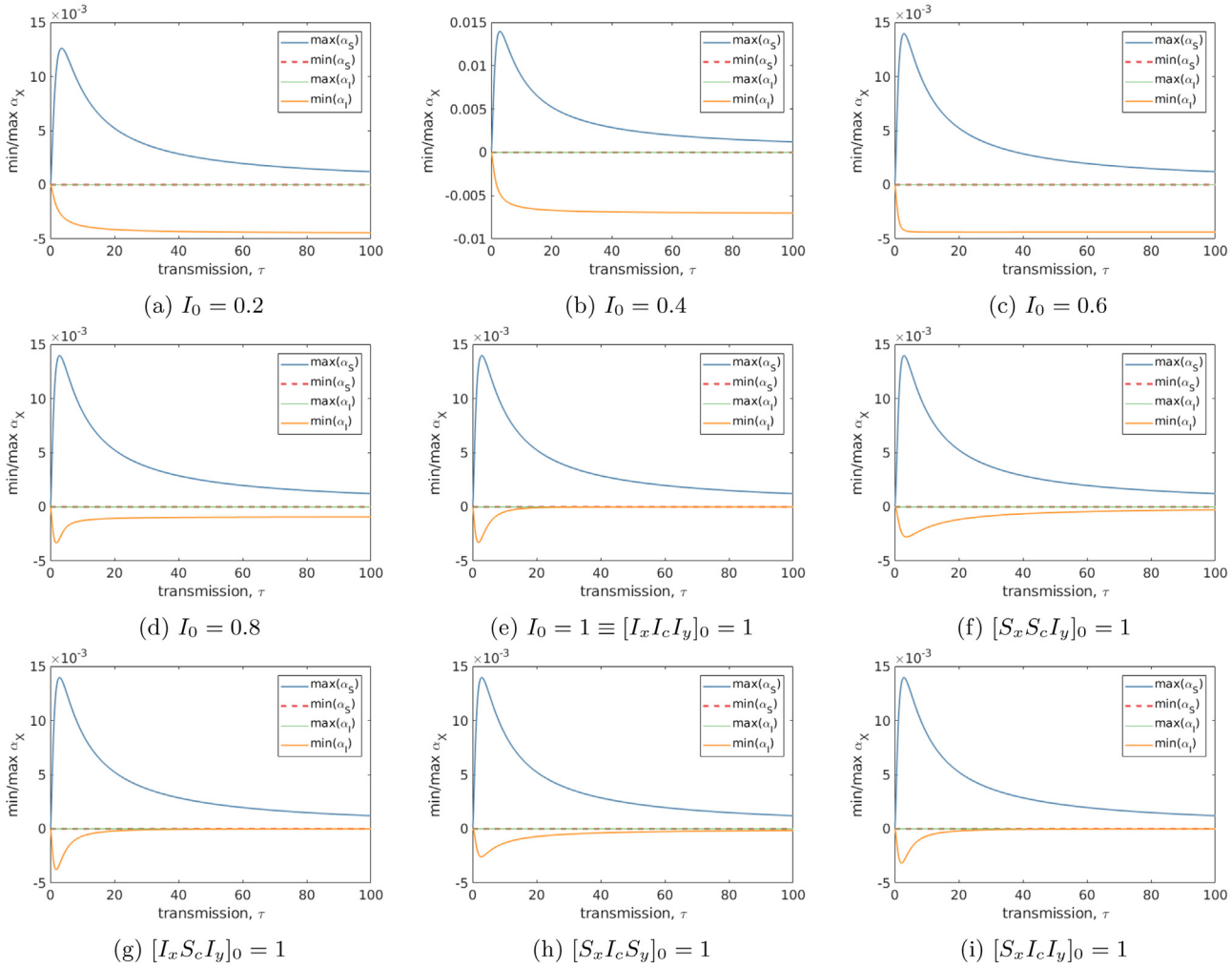


Fig. 9. Numerical demonstration of the bounds $\alpha_S \geqslant 0, \alpha_I \leqslant 0$ for the isolated open triple. We consider how $min(\alpha_X)$ and $max(\alpha_X)$, $X \in \{S, I\}$ vary with the transmission rate τ for the isolated open triple, for a range of different initial conditions – both random (a–e) and pure (e–i). These plots demonstrate the bounds $\alpha_S \geqslant 0$ and $\alpha_I \leqslant 0$ hold in general for the isolated open triple. In all plots we set $\gamma = 1$.

Using these conversions for example on Eqs. (30) and (32), we obtain the formally derived equations obtained by Taylor et al. (2012) (Theorem 1). We can also use these to convert our closures from proportions to numbers. Applying these, we obtain the unintuitive result that the closure $[XCY] \approx [XC][CY]/[C]$ in terms of proportions is equivalent to the closure $|XYZ| \approx (k-1)/k \times |XC||CY|/|C|$ in terms of numbers. Applying these conversions to this and to Eqns. (46) and (47) we obtain:

$$|XCY| \approx \frac{k-1}{k}\frac{|XC||CY|}{|C|} \tag{66}$$

$$|A_x S_c B_y I_z| \approx \frac{k(k-2)}{(k-1)^2}\frac{|ASB||BSI||ASI||S|}{|AS||BS||IS|} \tag{67}$$

$$|I_a S_x A_c B_y| \approx \frac{|ISA||SAB|}{|SA|} \tag{68}$$

To obtain the improved pairwise approximation in terms of numbers, we again consider the term α between a triple and its approximation. Below we consider $\alpha_{|ISI|}$:

$$\alpha_{|ISI|} = k|ISI||S| - (k-1)|SI|^2 \tag{69}$$

$$= k|ISI|\frac{|SSS| + 2|SSI| + |ISI|}{k(k-1)} - (k-1)\frac{(|SSS| + |SSI|)^2}{(k-1)^2} \tag{70}$$

$$\frac{\alpha_{|S|}}{k-1} \text{where} \alpha_{|S|} = |SSS||ISI| - |SSI|^2 \tag{71}$$

As before, we find $\alpha_{|ISI|} = \alpha_{|SSS|} = -\alpha_{|SSI|} = \alpha_{|S|}/(k-1)$. Defining $\alpha_{|I|}$ as $|III||SIS| - |SII|^2$, we similarly find $\alpha_{|SIS|} = \alpha_{|III|} = -\alpha_{|SII|} = \alpha_{|I|}/(k-1)$. By applying the conversions from the table to the equations from Appendix D, we can obtain expressions for the rate of change of $\alpha_{|S|}$ and $\alpha_{|I|}$:

$$\dot{\alpha_{|S|}} = \gamma(\phi_{|S|} - 2\alpha_{|S|}) + \tau(\beta_{|S|} - 2\alpha_{|S|}) \tag{72}$$

$$\text{where } \phi_{|S|} = |SSS||III| + |SIS|ISI| - 2|SSI||SII| \tag{73}$$

$$\text{and where } \beta_{|S|} = 2(k-1)(|I_a S_x S_c I_y||SS| - |I_a S_x S_c S_y||SI|)$$
$$+ 2|S_x S_c I_y I_z||SSI| - |S_x S_c S_y I_z||ISI| - |I_x S_c I_y I_z||SSS| \tag{74}$$

$$\dot{\alpha_{|I|}} = -4\gamma\alpha_{|I|} + \tau(\beta_{|I|} + 2\phi_{|I|} - 2\alpha_{|I|}) \tag{75}$$

$$\text{where } \phi_{|I|} = 2|SIS||ISI| - 2|SSI||SII| \tag{76}$$

$$\text{and where } \beta_{|I|} = 2(k-1)(|I_a S_x I_c I_y||SI| - |I_a S_x I_c S_y||II|)$$
$$- 2|S_x S_c I_y I_z||SII| + |I_x S_c I_y I_z||SIS| + |S_x S_c S_y I_z||III| \tag{77}$$

Finally, rearranging Eq. (68) and its analogues, substituting in $\alpha_{|X|}$ we can obtain the closure for triples in the improved pairwise approximation:

$$|ABA| \approx \frac{(k-1)^2|AB|^2 + \alpha_{|B|}}{k(k-1)|B|} \quad |ABC| \approx \frac{(k-1)^2|AB||BC| - \alpha_{|B|}}{k(k-1)|B|} \tag{78}$$

Thus we arrive at the improved pairwise approximation for k-regular networks, expressed in terms of numbers rather than proportions:

Model 6 in terms of numbers – The improved pairwise approximation for k-regular networks

$$|S\dot{S}| = 2\gamma|SI| - 2\tau\frac{(k-1)^2|SS||SI| - \alpha_{|S|}}{k(k-1)|S|} \tag{79}$$

$$|S\dot{I}| = \gamma(|II| - |SI|) - \tau|SI| + \tau\frac{(k-1)^2|SS||SI| - \alpha_{|S|}}{k(k-1)|S|} - \tau\frac{(k-1)^2|SI|^2 + \alpha_{|S|}}{k(k-1)|S|} \tag{80}$$

$$\dot{\alpha_{|S|}} = \gamma(\phi_{|S|} - 2\alpha_{|S|}) + \tau(\beta_{|S|} - 2\alpha_{|S|}) \tag{81}$$

$$\dot{\alpha_{|I|}} = -4\gamma\alpha_{|I|} + \tau(\beta_{|I|} + 2\phi_{|I|} - 2\alpha_{|I|}) \tag{82}$$

A.4. Appendix D – The rate of change of triples in a k-regular network

The state of triples in a k-regular network depend upon the state of order-four network structures: line-graphs of length four ($[A_aX_xC_cY_y]$, $[X_xC_cY_yB_b]$) and star graphs with three outer individuals ($[X_xC_cY_yZ_z]$). Assuming random initial conditions, by the symmetry of the system $[X_xC_cY_yB_b] = [B_aY_xC_cX_y]$, meaning only one length four line-graph term is needed in the equations below. Given that $[SSI] = [ISS]$ and $[SII] = [IIS]$, the rates of change of these triples are described by six ODEs, which can be derived from the system of Eqs. (12) described by House et al. (2009) by omitting terms that include closed loops and by converting the equations from numbers to proportions via the table in Appendix C.

$$[S\dot{S}S] = \gamma(2[SSI] + [SIS]) - \tau(k-2)[S_xS_cS_yI_z] - 2\tau(k-1)[I_aS_xS_cS_y] \tag{83}$$

$$[S\dot{S}I] = \gamma([SII] + [ISI] - [SSI]) - \tau[SSI] - \tau(k-2)[S_xS_cI_yI_z] + \tau(k-1)([I_aS_xS_cS_y] - [I_aS_xS_cI_y]) \tag{84}$$

$$[I\dot{S}I] = \gamma([III] - 2[ISI]) - 2\tau[ISI] - \tau(k-2)[I_xS_cI_yI_z] + 2\tau(k-1)[I_aS_xS_cI_y] \tag{85}$$

$$[SI\dot{S}] = \gamma(2[SII] - [SIS]) - 2\tau[SIS] + \tau(k-2)[S_xS_cS_yI_z] - 2\tau(k-1)[I_aS_xI_cS_y] \tag{86}$$

$$[S\dot{I}I] = \gamma([III] - 2[SII]) + \tau([SIS] + [SSI] - [SII]) + \tau(k-2)[S_xS_cI_yI_z]) + \tau(k-1)([I_aS_xI_cS_y] - [I_aS_xI_cI_y]) \tag{87}$$

$$[I\dot{I}I] = -3\gamma[III] + \tau(2[SII] + 2[ISI]) + \tau(k-2)[I_xS_cI_yI_z] + 2\tau(k-1)[I_aS_xI_cI_y] \tag{88}$$

A.5. Appendix E – The extended triple model

For this model, we model a triple and each of its neighbours explicitly. Thus, for a k-regular network, $3k-1$ individuals are modelled explicitly, meaning 2^{3k-1} equations are required to describe this model. By accounting for symmetries in the extended triple topology, one could reduce the dimensionality of this system. However, the method constructing the set of ODEs algorithmically described below models each state explicitly. Writing an algorithm that accounts for such symmetries, while possible, would be somewhat cumbersome, and as such we did not decide to pursue this. We approximate the external forces on this topology by assuming that the higher-order structures that the rate of change of states depend on can be approximated by conjoined extended triple topologies conditionally independent on the state of shared individuals.

We construct the extended triple model in two steps. Firstly, we construct a model with SIS-dynamics on the finite topology of the extended triple. To construct this model, we provide an algorithm for constructing SIS-models on graphs with any arbitrary finite topology. Secondly, we add on external force of infection to this model, which we achieve via relabelling.

E.1. An algorithm for constructing SIS-models on graphs with arbitrary topology

In this section we outline an algorithm for constructing a model with SIS-dynamics on networks of arbitrary topology. We can rewrite the full equations for the open triple in matrix form as follows: if we let $\underline{x} = \{[SSS], [SSI], [SIS], [SII], [ISS], [ISI], [IIS], [III]\}^T$. Then

$$\frac{d\underline{x}}{dt} = \gamma\underline{\underline{R}}\underline{x} + \tau\underline{\underline{N}}\underline{x} \tag{89}$$

States are ordered in this way so that they are interpreted as a binary string (e.g. $[SSS]$ as 000). For the open triple, these are given by:

$$\underline{\underline{R}} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 \end{bmatrix}$$

$$\underline{\underline{N}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 2 & 1 & 0 \end{bmatrix}$$

Thus, we need an algorithm to construct matrices $\underline{\underline{R}}$ and $\underline{\underline{N}}$ for an arbitrary graph topology, defined by its adjacency matrix $\underline{\underline{A}}$. Such an algorithm is detailed below:

1. Start with empty matrices $\underline{\underline{R}}$ and $\underline{\underline{N}}$ of size $2^a \times 2^a$, where a is the length of $\underline{\underline{A}}$.
2. Convert the decimal numbers d representing each state into binary vectors $\underline{b}$ of length l.
3. For each vector b, go through each entry i. If $b(i) = 1$, then $\underline{\underline{R}}(d,d) = \underline{\underline{R}}(d,d) - 1$. Let e be the decimal number obtained by changing $b(i)$ from 1 to 0, and $\underline{\underline{R}}(e,d) = \underline{\underline{R}}(e,d) + 1$
4. For each vector b, go through each entry j. If $b(j) = 0$, go through each entry k of b. If $b(k) = 1$ and $\underline{\underline{A}}(j,k) = 1$, then $\underline{\underline{N}}(d,d) = \underline{\underline{N}}(d,d) - 1$. Let e be the decimal number obtained by changing $b(j)$ from 0 to 1, and $\underline{\underline{N}}(e,d) = \underline{\underline{N}}(e,d) + 1$

Using this algorithm, we can construct a model with SIS-dynamics on the finite topology of the extended triple.*E.2. Relabelling – an example*

To consider the external force of infecting acting upon a particular state of the external triple, we must consider the external force of infection on the susceptible neighbours of that particular configuration. To evaluate this, we must consider the states in which this neighbour has no susceptible external partners, up to the state in which this neighbour has all susceptible external partners. We can achieve this by *relabelling* the system to give us equations describing the probability of being in said states.

Let us consider an example for a 3-regular network. Suppose we want to consider the external force of infection on the state $A = [S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}]$, with subscripts designating the positions described in Fig. D1. We include the semicolon to distinguish between the central triple and its neighbours. The only external force acting on this topology will be upon x_0, who is susceptible, by any external infected neighbour of x_0. Thus, the rate of change

of $[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}]$ will depend upon some order 10 terms: $[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}; I_{x_{00}}S_{x_{01}}]$, $[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}; S_{x_{00}}I_{x_{01}}]$, and $[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}; I_{x_{00}}I_{x_{01}}]$. We make a closure at this level by assuming, to take the first of these as an example:

$$[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}; I_{x_{00}}S_{x_{01}}] \approx \frac{[S_xS_cS_y; S_{x_0}I_{x_1}I_{c_0}I_{y_0}I_{y_1}] \times [S_{x_0}S_xS_c; I_{x_{00}}S_{x_{01}}I_{x_1}I_{c_0}S_y]}{[S_xS_cS_y; S_{x_0}I_{c_0}]} \tag{90}$$

However, as we have not modelled x_{00} and x_{01} explicitly, the probability of state $[S_{x_0}S_xS_c; I_{x_{00}}S_{x_{01}}I_{x_1}I_{c_0}S_y]$ remains undefined. However, as we start from random initial conditions, and given that a k-regular is isotropic, all extended triples within a k-regular network are equivalent. Because of this, we have:

$$[S_{x_0}S_xS_c; I_{x_{00}}S_{x_{01}}I_{x_1}I_{c_0}S_y] = [S_xS_cS_y; I_{x_0}S_{x_1}I_{c_0}I_{y_0}S_{y_1}] \tag{91}$$

Thus, we obtain an expression for this state by taking into account the symmetry of a k-regular network, and by relabelling individuals so that states containing individuals not explicitly modelled are defined in terms of explicitly modelled individuals exclusively.

We can now arrive at an expression for the external force of infection acting upon state A (λ_A), which is given by:

$$\lambda_A = \frac{\sum_{P,Q\in\{S,I\}} (1_{P=I} + 1_{Q=I}) \times [S_xS_cS_y; P_{x_0}Q_{x_1}I_{c_0}I_{y_0}S_{y_1}]}{\sum_{P,Q\in\{S,I\}} [S_xS_cS_y; P_{x_0}Q_{x_1}I_{c_0}I_{y_0}S_{y_1}]} \tag{92}$$

where $1_{P=I}$ and $1_{Q=I}$ are indicator functions.*E.3. Relabelling generally*

The particular relabelling depends upon the particular state of the external triple, and upon the particular neighbouring individuals whose external force of infection you are considering. The requires labellings for the $k = 3$ case are given in Table 2, and the required relabellings for a general k is given in Table 3. The header row gives the neighbouring individual whose external force of infection we are considering, while the leftmost column gives the new positions of states in a given column now occupy. External nodes that contribute to the external force of infection always occupy the relabelled x_i positions.*E.4. Constructing the extended triple model*

To make the extended triple model, we begin by constructing the model for the relevant finite topology with SIS-dynamics, as outlined previously in this section. To construct a model approximating a k-regular network, we must add an external force of infection to individuals neighbouring the central triple. The procedure is as follows:

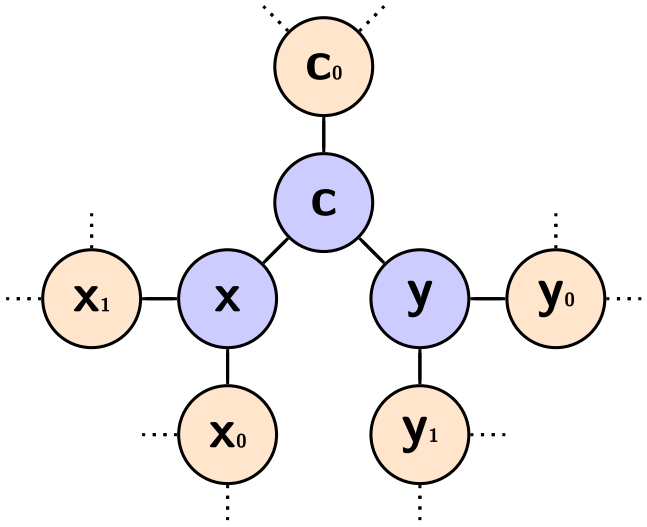


Fig. 10. A graphical representation of the extended triple approximation. Here we visualise the extended open triple model for $k = 3$. Shaded blue is our triple of interest, shaded in orange are any additional individuals that are modelled explicitly, while dotted lines show connections to individuals not explicitly modelled that exert an external force of infection upon the topology. The state of this topology will depend upon order 10 structures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Conversion table between proportions and numbers.

Motif	Ratio to # of individuals	Term (proportion)	Equivalent term (numbers)
Individual	1	$[X]$	$\|X\|/N$
Pair	k	$[XY]$	$\|XY\|/kN$
Triple	$k(k-1)$	$[XCY]$	$\|XCY\|/k(k-1)N$
Line graph (length 4)	$k(k-1)^2$	$[A_aX_xC_cY_y]$	$\|A_aX_xC_cY_y\|/k(k-1)^2N$
4-star	$k(k-1)(k-2)$	$[X_xC_cY_yZ_z]$	$\|X_xC_cY_yZ_z\|/k(k-1)(k-2)N$

Table 2
Relabelling for $k = 3$.

Individual	$\mapsto$	$x_0 = S$	$x_1 = S$	$c_0 = S$	$y_0 = S$	$y_1 = S$
x	$\mapsto$	X_0	X_1	C_0	Y_0	Y_1
c	$\mapsto$	X	X	C	Y	Y
y	$\mapsto$	C	C	Y	C	C
x_0	$\mapsto$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$
x_1	$\mapsto$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$
c_0	$\mapsto$	X_1	X_0	X	Y_1	Y_0
y_0	$\mapsto$	C_0	C_0	Y_0	C_0	C_0
y_1	$\mapsto$	Y	Y	Y_1	X	X

Table 3
Relabelling for general k.

Individual	$\mapsto$	$x_i = S$	$c_0 = S$	$c_i = S$	$y_0 = S$	$y_i = S$
x	$\mapsto$	X_i	C_0	C_i	Y_0	Y_i
c	$\mapsto$	X	C	C	Y	Y
y	$\mapsto$	C	Y	Y	C	C
x_i	$\mapsto$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$	$\{S,I\}$
c_0	$\mapsto$	X_{i+1}	X	X	Y_1	Y_{i+1}
c_j	$\mapsto$	X_{i+j+1}	C_j	C_{j+1}	C_{j+1}	C_{i+j+1}
y_0	$\mapsto$	Y	Y_0	Y_0	X	X
y_j	$\mapsto$	C_{j-1}	Y_j	Y_j	C_{j-1}	C_{j-1}

1. Construct ODEs for the SIS-dynamics for a graph of the relevant topology, with the central triple as the first three rows of the adjacency matrix.
2. Express each state N as a binary $\underline{b}$ (of length $l = 3k - 1$).
3. For each vector b, loop through entries $i \in \{4, \dots, l\}$. If $b(i) = 1$ calculate the external force of infection on this node, I_{ext}, by relabelling.
4. Subtract this, multiplied by τ and the state itself (i.e. $\tau N I_{ext}$), to that state's ODE (i.e. $\dot{N} = \dot{N} - \tau N I_{ext}$).
5. Let e be the decimal number obtained by changing $b(i)$ from 1 to 0, and let E be the state corresponding to this number. Add on the $\tau N I_{ext}$ to this ODE (i.e. $\dot{E} = \dot{E} + \tau N I_{ext}$).

A.6. Appendix F – Convergence to the mean-field approximation as $k \to \infty$

We believe that as $k \to \infty$, all models converge to the mean-field approximation. In this section, we show this is true for both the pairwise and improved pairwise approximation models, and outline how this would be approached in the general case.

For all models Eqs. (30) ($[\dot{S}]$) and (32) ($[\dot{SI}]$) hold exactly – only beginning to differ at the level of triples. Our contention is that as $k \to \infty$, $[SI] \to [S][I]$. First, we note that because $[SI] = [S] - [SS]$, $[SI] = [S][I] \iff [SS] = [S]^2$. We consider $[\dot{SS}]$,

$$[\dot{SS}] = 2\gamma[SI] - 2(k-1)\tau[SSI] \tag{93}$$

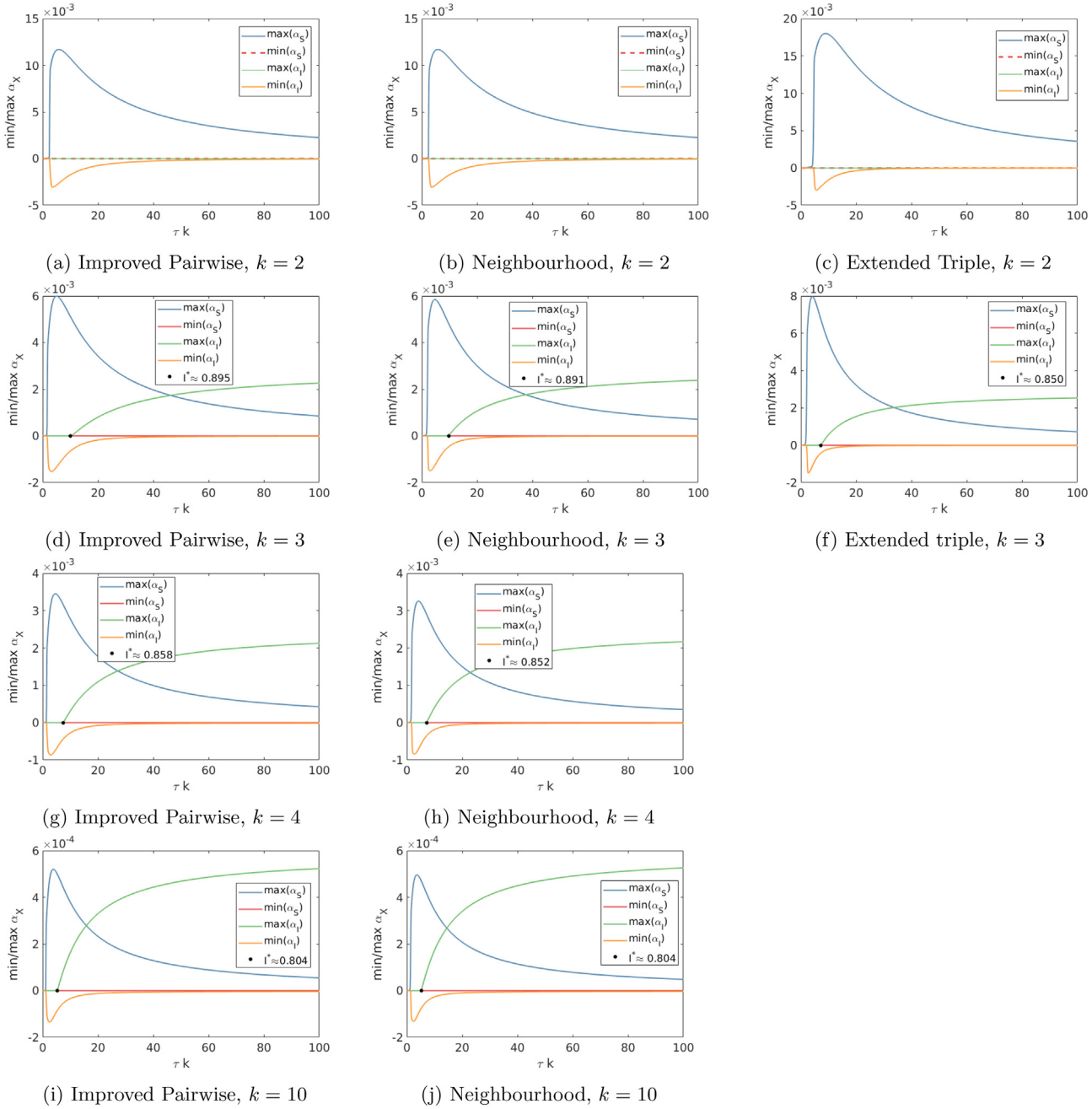


Fig. 11. Numerical exploration of α_S and α_I for different approximate models of the k-regular network. We consider how $min(\alpha_X)$ and $max(\alpha_X)$, $X \in \{S, I\}$ vary with τk for different approximate models of the k-regular network: Improved pairwise (left column), neighbourhood (centre column), extended triple (right column). These plots demonstrate the bound $\alpha_S \geqslant 0$ holds for all approximations of the k-regular network, but that $\alpha_I \leqslant 0$ only holds for the case $k = 2$. For $k > 2$, $max(\alpha_I) > 0$ given τ is sufficiently high. These transmission rates correspond to high endemic prevalences - in all cases $I^* > 0.8$. In all plots we set $\gamma = 1$.

Now, we introduce $\lambda = \tau k$, which remains constant as k increases. We make the assumption that $[SS] = [S]^2$ initially and consider their time evolution:

$$\dot{([S]^2)} = 2[S][\dot{S}] = 2\gamma[S][I] - 2\lambda[S]^2[I] \tag{94}$$

$$[S\dot{S}] = 2\gamma[S][I] - 2\frac{(k-1)}{k}\lambda[SSI] \to 2\gamma[S][I] - 2\lambda[SSI] \text{as} k \to \infty \tag{95}$$

These equations are equal, and therefore the relationship $[SS] = [S]^2$ continues to hold, conditional on $[SSI] = [S]^2[I]$. In general we need to show that the relationship $[SSI] = [S]^2[I]$ continues, given that it holds initially.*F.1 – Convergence for the pairwise approximation model*

Under the standard pairwise model, $[SSI] = [SS][SI]/[S]$. Assuming that $[SS] = [S]^2$, it is clear that $[SSI] = [S]^2[I]$. Given that at $t = 0$, $[SS] = [S]^2$, and that $[SS] = [S]^2 \Rightarrow [S\dot{S}]^2$ the convergence of the standard pairwise model is proved by induction.*F.2 – Convergence for the improved pairwise approximation model*

Under this model $[SSI] = ([SS][SI] - \alpha_S)/[S]$, i.e. $[SSI] = [S]^2[I] \iff ([SS] = [S]^2, \alpha_S = 0)$. Let us assume that $[SS] = [S]^2$, $\alpha_S = 0$, and $\alpha_I = 0$. Then $[SSI] = [S]^2[I]$ and by examining Eqns. (40) and (43), we find that $\dot{\alpha_S} = 0$ and $\dot{\alpha_I} = 0$. Given that at $t = 0, [SS] = [S]^2$, $\alpha_S = 0, \alpha_I = 0$, and that $[SS] = [S]^2, \alpha_S = 0, \alpha_I = 0 \Rightarrow [S\dot{S}] = [\dot{S}]$, $\dot{\alpha_S} = 0, \dot{\alpha_I} = 0$, the convergence of the improved pairwise model is proved by induction.*F.3 – Convergence in the general case*

More generally, we believe that as $k \to \infty$, spatial correlation at a particular level is only introduced by spatial correlations at a higher level. For example, correlations only enter the pairwise model if there are correlations at the level of pairs, correlations only enter the improved pairwise model if there are correlations at the level of pairs and triples (α terms), etc. Given that by assumption we start with no spatial correlation at any level, it follows that correlations are never introduced. However, we believe that the proof of this more general claim is beyond the remit of this paper.

A.7. Appendix G – Exploring α_S and α_I for different approximate models of the k-regular network

Fig. 10,11

A.8. Appendix H – Exploring the shape of α_I for different approximate models

Fig. 12

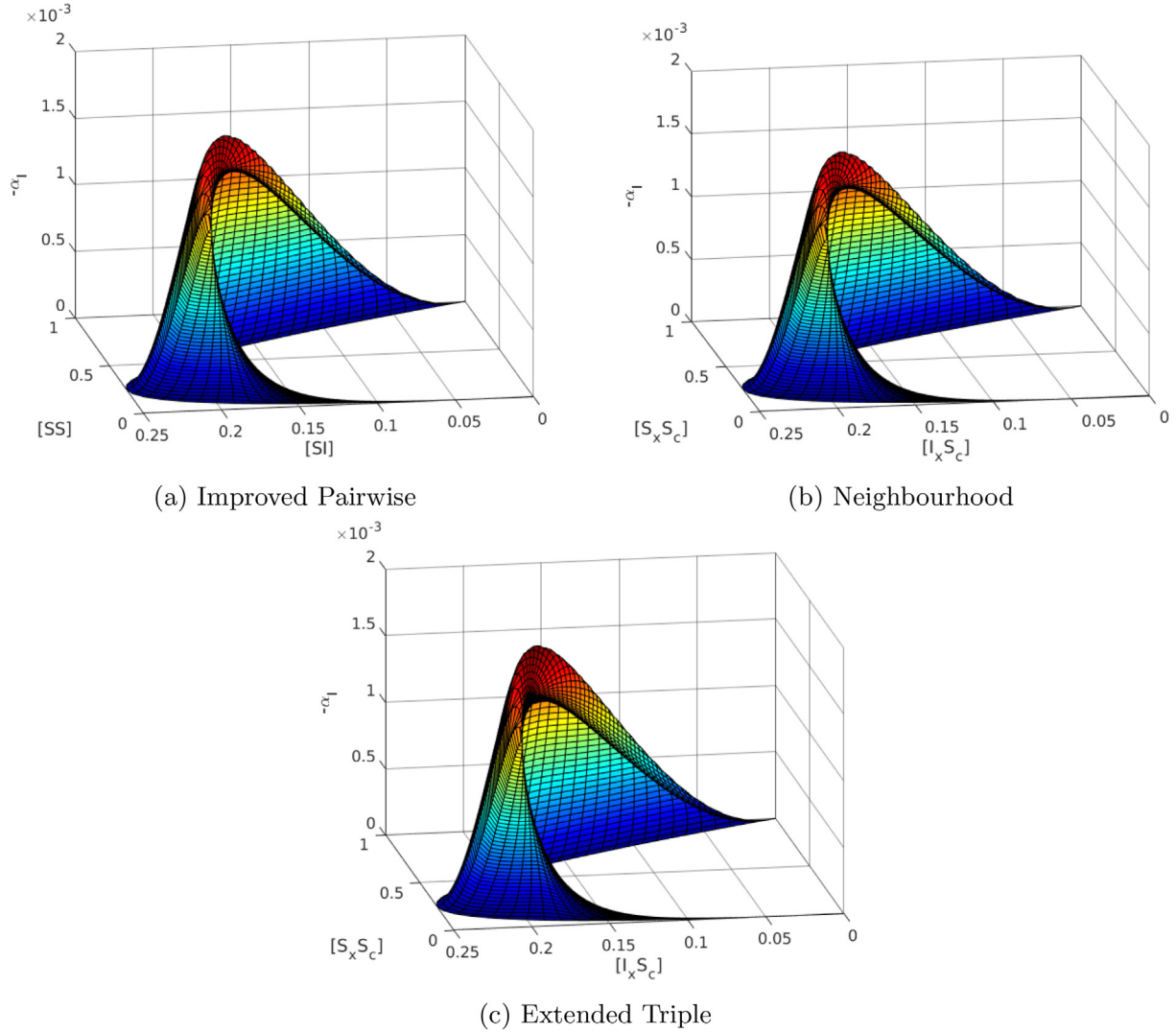
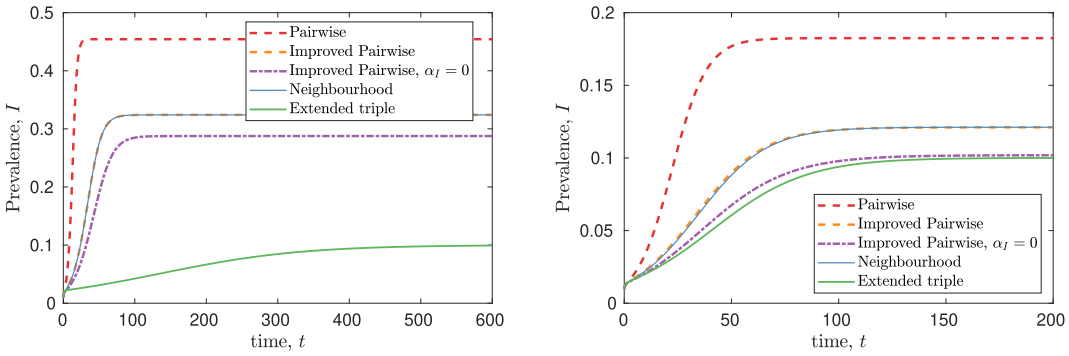


Fig. 12. Exploring the shape of α_I for different approximate models. Here we compare the shape of the error term $-\alpha_I$ as a function of $[SS]$ and $[SI]$ for the improved pairwise model (a), and as a function of $[S_xS_c]$ and $[I_xS_c]$ for neighbourhood and extended triple approximations ((b) and (c)), for the example $k = 3$. We observe that α_I surfaces in all three models are very similar, and that their magnitude is much smaller than their corresponding α_S surfaces (Fig. 8). In all plots, we set $\tau = 1, \gamma = 1$.



(a) $k = 2$ (b) $k = 3$

Fig. 13. Comparing the time-evolution of improved pairwise approximations against higher-order approximations for $k = 2$ and $k = 3$-regular networks. (a) and (b) illustrate the performance of improved pairwise approximations compared to the higher-order neighbourhood (blue) and extended triple (green) approximations. We choose values τ s.t. $I^* = 0.1$ in the extended triple model ((a) $\tau = 1.4163$, (b) $\tau = 0.5744$). In both (a) and (b) there is little difference between the time-evolution of improved pairwise (orange) and neighbourhood model. In (b), while the improved pairwise model with $\alpha_I = 0$ (purple) matches the endemic prevalence of the extended triple approximation closely, its evolution to this equilibrium state differs. In both plots we set $\gamma = 1$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

A.9. Appendix I – Exploring the time-evolution of approximate models for $k = 2$ and $k = 3$

Here, we present the time-evolution of improved pairwise models, as well as for the neighbourhood and extended triple approximation models, for $k = 2$ and $k = 3$. We note that the improved pairwise approximation matches closely to that of the neighbourhood approximation models. We also see that while for $k = 3$ the improved pairwise model with $\dot{\alpha}_I = 0$ matches the endemic prevalence of the extended triple approximation closely, the same cannot be said about their time-evolution. Fig 13.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.jtbi.2020.110328.

References

Ball, F., Neal, P., 2008. Network epidemic models with two levels of mixing. Math. Biosci. 212 (1), 69–87.

Bansal, S., Read, J., Pourbohloul, B., Meyers, L.A., 2010. The dynamic nature of contact networks in infectious disease epidemiology. J. Biol. Dyn. 4 (5), 478–489.

Bauch, C., Rand, D.A., 2000. A moment closure model for sexually transmitted disease transmission through a concurrent partnership network. Proceedings of the Royal Society of London. Series B: Biological Sciences 267 (1456), 2019–2027.

Eames, K.T., Keeling, M.J., 2002. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proc. Nat. Acad. Sci. 99 (20), 13330–13335.

Edwards, R., Kim, S., van den Driessche, P., 2010. A multigroup model for a heterosexually transmitted disease. Math. Biosci. 224 (2), 87–94.

Floyd, W., Kay, L., Shapiro, M., 2012. A covering-graph approach to epidemics on SIS and SIS-Like networks. Bull. Math. Biol. 74 (1), 175–189.

Goodreau, S.M., Rosenberg, E.S., Jenness, S.M., Luisi, N., Stansfield, S.E., Millett, G.A., Sullivan, P.S., 2017. Sources of racial disparities in HIV prevalence in men who have sex with men in Atlanta, GA, USA: a modelling study. Lancet HIV 4 (7), e311–e320.

Hansson, D., Leung, K., Britton, T., Strömdahl, S., 2019. A dynamic network model to disentangle the roles of steady and casual partners for HIV transmission among MSM. Epidemics 27, 66–76.

House, T., 2015. Algebraic moment closure for population dynamics on discrete structures. Bull. Math. Biol. 77 (4), 646–659.

House, T., Davies, G., Danon, L., Keeling, M.J., 2009. A motif-based approach to network epidemics. Bull. Math. Biol. 71 (7), 1693–1706.

Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. Roy. Soc. Lond. Ser. B Biol. Sci. 266(1421), 859–867..

Keeling, M.J., House, T., Cooper, A.J., Pellis, L., 2016. Systematic approximations to susceptible-infectious-susceptible dynamics on networks. PLoS Comput. Biol. 12, (12) e1005296.

Kirkwood, J.G., 1935. Statistical mechanics of fluid mixtures. J. Chem. Phys. 3 (5), 300–313.

Kiss, I.Z., Miller, J.C., Simon, P.L., et al., 2017. Mathematics of Epidemics on Networks. Springer, Cham.

Kiss, I.Z., Morris, C.G., Sélley, F., Simon, P.L., Wilkinson, R.R., 2015. Exact deterministic representation of Markovian epidemics on networks with and without loops. J. Math. Biol. 70 (3), 437–464.

Lee, H.K., Shim, P.-S., Noh, J.D., 2013. Epidemic threshold of the susceptible-infected-susceptible model on complex networks. Phys. Rev. E 87, (6) 062812.

Leng, T., Keeling, M.J., 2018. Concurrency of partnerships, consistency with data, and control of sexually transmitted infections. Epidemics 25, 35–46.

Liggett, T.M., 2013. In: Stochastic interacting systems: contact, voter and exclusion processes, Vol. 324. springer science & Business Media.

Lindquist, J., Ma, J., Van den Driessche, P., Willeboordse, F.H., 2011. Effective degree network disease models. J. Math. Biol. 62 (2), 143–164.

Molloy, M., Reed, B., 1995. A critical point for random graphs with a given degree sequence. Random Struct. Algorithms 6 (2–3), 161–180.

Pellis, L., House, T., Keeling, M.J., 2015. Exact and approximate moment closures for non-Markovian network epidemics. J. Theor. Biol. 382, 160–177.

Rand, D., 1999. Correlation equations and pair approximations for spatial ecologies. Adv. Ecol. Theory Principles Appl. pp. 100–142..

Sharkey, K.J., 2008. Deterministic epidemiological models at the individual level. J. Math. Biol. 57 (3), 311–331.

Sharkey, K.J., 2011. Deterministic epidemic models on contact networks: correlations and unbiological terms. Theor. Popul. Biol. 79 (4), 115–129.

Sharkey, K.J., Kiss, I.Z., Wilkinson, R.R., Simon, P.L., 2015. Exact equations for SIR epidemics on tree graphs. Bull. Math. Biol. 77 (4), 614–645.

Simon, P.L., Kiss, I.Z., 2015. Super compact pairwise model for SIS epidemic on heterogeneous networks. J. Complex Networks 4 (2), 187–200.

Simon, P.L., Taylor, M., Kiss, I.Z., 2011. Exact epidemic models on graphs using graph-automorphism driven lumping. J. Math. Biol. 62 (4), 479–508.

Taylor, M., Simon, P.L., Green, D.M., House, T., Kiss, I.Z., 2012. From Markovian to pairwise epidemic models and the performance of moment closure approximations. J. Math. Biol. 64 (6), 1021–1042.

Taylor, T.J., Kiss, I.Z., 2014. Interdependency and hierarchy of exact and approximate epidemic models on networks. J. Math. Biol. 69 (1), 183–211.

Trapman, P., 2007. Reproduction numbers for epidemics on networks using pair approximation. Math. Biosci. 210 (2), 464–489.

Volz, E., 2008. SIR dynamics in random networks with heterogeneous connectivity. J. Math. Biol. 56 (3), 293–310.

Volz, E., Meyers, L.A., 2007. Susceptible–infected–recovered epidemics in dynamic contact networks. Proc. Roy. Soc. B Biol. Sci. 274 (1628), 2925–2934.

Whittles, L.K., White, P.J., Didelot, X., 2019. A dynamic power-law sexual network model of gonorrhoea outbreaks. PLoS Comput. Biol. 15, (3) e1006748.

Wilkinson, R.R., Sharkey, K.J., 2013. An exact relationship between invasion probability and endemic prevalence for Markovian SIS dynamics on networks. PloS One 8, (7) e69028.

Xiridou, M., Geskus, R., de Wit, J., Coutinho, R., Kretzschmar, M., 2003. The contribution of steady and casual partnerships to the incidence of HIV infection among homosexual men in amsterdam. AIDS 17 (7), 1029–1038.