PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Systematic Approximations to Susceptible-Infectious-Susceptible Dynamics on Networks

Matt J. Keeling[1,2,3]*, **Thomas House**[1,2,4], **Alison J. Cooper**[5], **Lorenzo Pellis**[1,2]

1 Zeeman Institute: SBIDER, University of Warwick, Coventry, United Kingdom, **2** Mathematics Institute, University of Warwick, Coventry, United Kingdom, **3** School of Life Sciences, University of Warwick, Coventry, United Kingdom, **4** School of Mathematics, University of Manchester, Manchester, United Kingdom, **5** School of Engineering, University of Warwick, Coventry, United Kingdom

* M.J.Keeling@warwick.ac.uk

Abstract

Network-based infectious disease models have been highly effective in elucidating the role of contact structure in the spread of infection. As such, pair- and neighbourhood-based approximation models have played a key role in linking findings from network simulations to standard (random-mixing) results. Recently, for SIR-type infections (that produce one epidemic in a closed population) on locally tree-like networks, these approximations have been shown to be exact. However, network models are ideally suited for Sexually Transmitted Infections (STIs) due to the greater level of detail available for sexual contact networks, and these diseases often possess SIS-type dynamics. Here, we consider the accuracy of three systematic approximations that can be applied to arbitrary disease dynamics, including SIS behaviour. We focus in particular on low degree networks, in which the small number of neighbours causes build-up of local correlations between the state of adjacent nodes that are challenging to capture. By examining how and when these approximation models converge to simulation results, we generate insights into the role of network structure in the infection dynamics of SIS-type infections.

Author Summary

Networks are now widely used to model infectious diseases, but have posed significant mathematical challenges. Recently analytic results have been obtained for 'one-off' network epidemics that follow the SIR paradigm, but these results do not carry over to other scenarios—most significantly to many sexually transmitted infections, where accounting for network structure is vital. Here, we show that it is possible to obtain the large-population dynamics of such diseases on networks through systematic approximations. We focus on a mathematically challenging case of SIS dynamics on networks with low degree.

Check for updates

OPEN ACCESS

Citation: Keeling MJ, House T, Cooper AJ, Pellis L (2016) Systematic Approximations to Susceptible-Infectious-Susceptible Dynamics on Networks. PLoS Comput Biol 12(12): e1005296. doi:10.1371/journal.pcbi.1005296

Editor: Katia Koelle, Duke University, UNITED STATES

Received: June 3, 2016

Accepted: December 9, 2016

Published: December 20, 2016

Copyright: © 2016 Keeling et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: TH is funded by an EPSRC fellowship (EP/J002437/1), MJK & LP were funded by EPSRC grant (EP/H016139/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

There is a strong and deep connection between networks and the spread of infectious diseases [1–9]. Virtually all infections can be thought of as propagating through a network of (epidemiologically-relevant) contacts between individuals in the population, with the structure of this underlying network determining much of the infection dynamics. Therefore an understanding of population-level transmission at the scale of individual hosts is closely linked to a study of the properties of the underlying transmission network. Recent advances in network science have highlighted how both local and global structure of the network are key in the dynamics of infection [2, 10–13].

While networks are being increasingly used for airborne and close-contact infections (such as influenza [14] and RSV [15]) which spread through social contacts, the epidemiological network literature was originally formulated for sexually transmitted infections (STIs) where the network is generally more clearly defined. Classic examples include homosexual contact networks from early HIV studies [1] and the Colorado-Springs study of sexual contacts in the high-risk heterosexual population [16]. While a focus on STIs has substantial advantages in terms of determining the network, it also places constraints on the epidemiological dynamics that need to be considered. The overwhelming majority of STIs (e.g. chlamydia or gonorrhoea, although not HIV) can be approximated using the Susceptible-Infected-Susceptible (SIS) paradigm, where infected individuals are treated and recover to the susceptible state, and hence are able to be re-infected multiple times. Although SIS models are inherently lower-dimensional than their SIR (Susceptible-Infected-Recovered) counterparts, potential reinfection of the same individual (multiple times) leads to more complex dynamical behaviour between neighbouring nodes on a network and makes it more difficult to generate tractable results [17].

When details of the complete network are available, and we are dealing with a particular applied problem, then the most straightforward approach is to simulate the dynamics of infection on the given network (e.g. [5, 16, 18]). However in anything but ideal circumstances simulation may be problematic. For example, using simulations alone: it may be difficult to understand sensitivity to elements in network structure or biases in the way the network connections were sampled; it is computationally challenging to infer epidemiological parameters; and it may be difficult to gain a robust understanding of the causal determinants of the observed dynamics. Approximations that maintain the analytic tractability of traditional ODE (Ordinary Differential Equation) models, but take account of elements of network structure provide a possible solution. These have been quite successful for 'one off' epidemics that obey the SIR paradigm, with notable advances including the 'effective degree' approach of Ball and Neal [19], the probability-generating function approach of Volz [20] (and its reduction to a single dynamical equation by Miller [21]) and the model of Lindquist et al. [22] (originally also called 'effective-degree' model, but to which we refer throughout as 'neighbourhood model' to avoid confusion with the earlier use of this terminology). Such models, together with pairwise or related approximations [23, 24] discussed extensively in this paper, have been shown to be exact methods of calculation of marginal probabilities for the stochastic SIR model for finite explicitly known networks [25–28]. Such models can also reproduce the expected course of the stochastic SIR model with constant infection and recovery rates on large configuration model networks with with several recent asymptotic proofs of convergence published [29–32]. Much work has also focussed on extending such methods to weighted [33] and dynamic [34] networks, as well as to models with arbitrary duration of the infectious period [35–37], with the common denominator that on clustered networks results from all approaches are only approximate.

Despite all these successes concerning SIR models (or related models such as SEIR, which includes an exposed period [17, 24]), the same is not generally true for infections without long-lasting immunity, with realistic demographic turnover or with significant viral mutation [38, 39]—i.e., the majority of pathogens of interest. Here we focus on STIs since the motivation for use of a network is strong [40]. These diseases are of major public health importance and the appropriate modelling framework (the SIS model on a network, also called the 'contact process' and frequently considered in theoretical studies) is the most challenging for approximation models to capture. To fully predict the dynamics and hence the impact of control on a range of sexually transmitted infections requires mathematical models that can account for both network structure of sexual partnership and the complications that arise from reinfection that is associated with SIS-type behaviour [40]. Here we consider three distinct approaches to capture the dynamic build-up of correlations between nearby individuals on the network—each approximation methodology has an associated integer that can be increased to achieve greater levels of accuracy. We stress, however, that our approach does not rely on special features of the SIS model but can be applied to the full spectrum of disease-dynamics models used to inform applied epidemiology and public health (including those with short-term immunity and hence SIRS-type dynamics).

Although the long-term aim is to utilise such approximation techniques to gain a clear understanding of the dynamics of STIs (as well as other infections that confer short-duration immunity) on realistic networks, we focus this paper on understanding when simple modelling techniques fail. One occasion when simple approaches fail is in the case of extreme heterogeneity [41–44]. Although risk-structure (or heterogeneity in network structure) is a highly important aspect of modelling STI—especially in terms of defining individual risk—we argue that many studies have focussed on this aspect [45, 46] and that it is usually possible to capture epidemiological effects of population heterogeneity by modestly increasing the system's dimension through the introduction of multiple risk-groups (e.g. low- and high-risk behaviour).

In contrast to the heterogeneous case, the impact of a limited number of contacts and the build-up of dynamical correlations in the state of neighbouring nodes is a less studied issue that presents deep conceptual challenges, especially for SIS dynamics. We therefore focus on developing a better understanding of this problem by ignoring many realistic features of STIs (we briefly comment on them in the discussion) and considering the idealised case of a homogeneous degree or 'k-regular' networks with low connectivity and hence greater importance of the link with each contact (in particular $k = 3$ and $k = 2$). In these remarkably simple networks, the effects of local correlations are at their strongest and are not masked by the impact of degree heterogeneity. To illustrate this concept we compare simple risk-structured mean-field (random-mixing) models (which account for degree heterogeneity within the network but not correlations that develop due to contact structure) with results from stochastic network simulations of SIS infection dynamics (Fig 1). This example demonstrates that when either the mean degree or the variance in the degree distribution increases, so the standard risk-structured ODE model provides a better fit to the simulated dynamics. The agreement between these simple models and simulations is worst for a homogeneous degree 3 network and hence it is this test scenario we predominantly consider throughout this paper.

Methods

Discussion of models

In this work, simulation models and traditional mean-field approximation models (that ignore network structure) represent two extremes in terms of analytic tractability and computational efficiency. Two approximate models for SIS-dynamics (Materials and Methods) have been developed that lie between these extremes: *pairwise* and *neighbourhood* approximations.

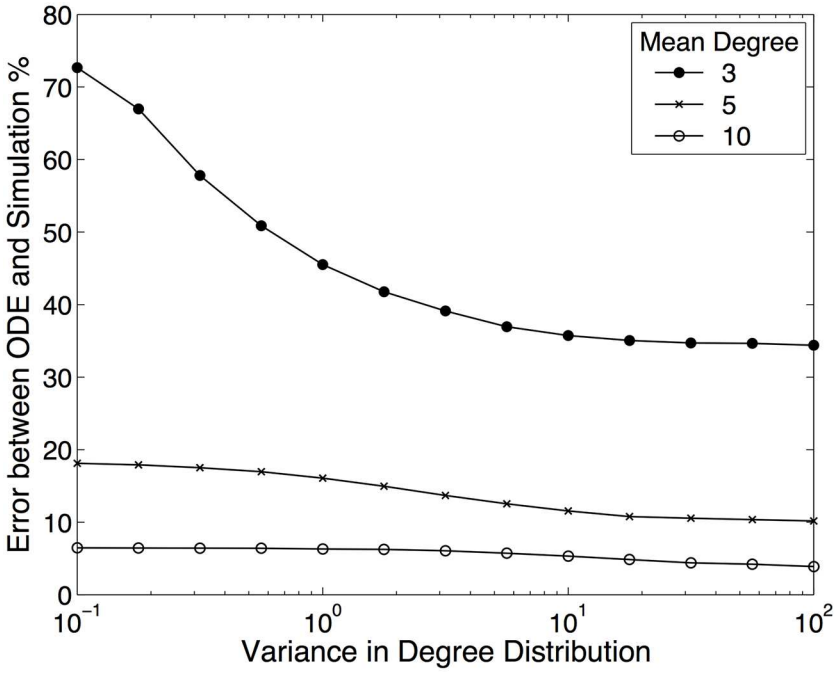

Fig 1. Comparison between risk-structured mean-field results and stochastic network simulations for SIS infection dynamics. The error is the relative percentage error between the risk-structured mean-field model (Supporting Information) and network simulation results for the mean prevalence of infection: $Error = 100 \times \|\bar{I}_{ODE} - \bar{I}_{network}\| / \bar{I}_{network}$. The degree distribution obeys $P(k) = \rho \exp(-\alpha(k - K)^2)$ for all $k \geq 2$, where α and K are determined to give a desired mean and variance, and ρ is a normalising constant. Other methods of generating degree distributions give similar results. We insist on $k \geq 2$ (for all nodes) as this generally ensures that the majority of network form a single giant component. The network is formulated using the Molly-Reed algorithm [47] with 100,000 nodes, but ensuring that there were no self-contacts and no multiple connections between individuals. We assume the recovery rate of individuals γ is scaled to be one, while the transmission rate across a contact scales with the mean number of contacts, $\tau = 2/\bar{k}$. As such, the mean prevalence in both simulation and ODE models lies between 30 and 50%.

doi:10.1371/journal.pcbi.1005296.g001

Pairwise approximations [45, 48–50] consider the dynamic states of pairs of individuals that are connected in the network and hence capture some of the build-up of local correlations within the network. Neighbourhood approximations [22] have appeared more recently and can be conceptualised as a more sophisticated, though higher-dimensional, extension to the pairwise approximation. Neighbourhood approximations model the number of connected individuals of each type around a central individual; for SIS dynamics this is simply the number of S and I connected to a central individual of a given state. This means that neighbourhood models capture higher-order correlations within the network, as they effectively model multiple chains of three connected individuals sharing the same central node.

We consider methods to extend the pairwise and neighbourhood approaches; either increasing the size of the subgraph considered (e.g. going from modelling pairs to modelling triple motifs) or increasing the number of node states by counting infection events.

Subgraph or motif expansions to the pairwise models track the dynamics of the possible states of increasingly larger motifs or subgraphs of m connected individuals within the network. Clearly as m becomes large, we precisely account for the full dynamics on larger sections of the network and hence expect our approximations to become more accurate. However with increasing m comes increasing number of motifs and also higher dimension dynamics. For degree $k = 3$ networks we consider motifs of size $m = 1$ (the standard mean-field model), $m = 2$

(the traditional pairwise model) as well as $m = 3$ and $m = 4$; for the special case of degree $k = 2$ networks, we are able to consider larger motifs up to $m = 16$ due to the linear structure of all $k = 2$ motifs (see Fig 2).

Similarly, it is feasible to expand the neighbourhood model, which we index by parameter n. Again we consider $n = 1$ to be the standard mean-field model, while $n = 2$ accounts for the states of all neighbours of a central individual, and expansions to neighbours of neighbours ($n = 3$) is also possible although of very high dimensional for networks of $k = 3$ or above.

The reinfection counting extension explicitly tracks the number of times an individual has been infected, effectively increasing the number of states for each individual (i.e. disease state × number of times infected). To create a finite system, we track the infection times up to a maximum of L (which now incorporates all those individuals infected L times or more). The motivation for this extension derives from a failure in traditional SIS pairwise models to account for the correlation between infected and newly recovered individuals. This extension should therefore improve the performance of the approximation model during the early stages of invasion when infection is rare. However, the long-term equilibrium dynamics when all individuals have been infected L times or more (assuming the infection persists), will be identical to that of the pairwise model.

Although we take the simulation model as our gold-standard, deriving precise values for particular quantities is often computationally intensive and naive methods can be improved. For early epidemic growth rates, we generate a finite Cayley tree (thereby eliminating all clustering) and study the dynamics until infection hits an outer leaf. For endemic prevalence it is not possible to use a Cayley tree; any finite Cayley tree must have lower degree at the outer leaves which would influence the dynamics. Instead, we generate large networks using the Molloy-Reed (or configuration) algorithm [47], and ensure that self connections, multiple connections between nodes and short loops (of five or less connections) are removed by randomly shuffling connections. Moreover, far greater accuracy can be achieved when estimating quantities from simulation by realising that the expected rate of change of infection is determined by the state of the network and is given exactly by mechanistic models (such as Eq 3) where the variables are taken directly from the simulation. This allows us to remove some of the effects of stochasticity from the calculation. We therefore use this expected rate of change to directly calculate early growth rates, and use the long-term relationship between prevalence and expected rate of change to find the endemic equilibrium prevalence (see S1 Text). Fig 3 shows the advantage of this method, reducing the variance in our estimate of mean endemic prevalence and hence improving the accuracy of any fixed duration simulation.

Mathematical definition of models

We now layout in some detail the different approximation models used within this paper: mean-field; standard pairwise; reinfection counting; motif models; and neighbourhood models. The elements captured in each approximation are illustrated in Fig 4.

Full dynamics and notation. We consider individuals labelled with integers $i, j, \ldots \in \{1, \ldots, N\}$ connected on a network with adjacency matrix $\boldsymbol{A} = (A_{i,j})$; where N is the number of nodes in the network, and $A_{i,j}$ is one if nodes i and j are connected or zero otherwise. The state of individual i at time t is given by a Bernoulli random variable $X_i(t)$ taking the value S_i if i is susceptible and I_i if i is infectious. The events and rates of the full underlying dynamics are

$$\begin{aligned} S_i &\to I_i \text{ at rate } \tau \sum_j A_{i,j} \mathbb{1}_{\{X_j(t)=I_j\}} \,, \\ I_i &\to S_i \text{ at rate } \gamma \,, \end{aligned} \tag{1}$$

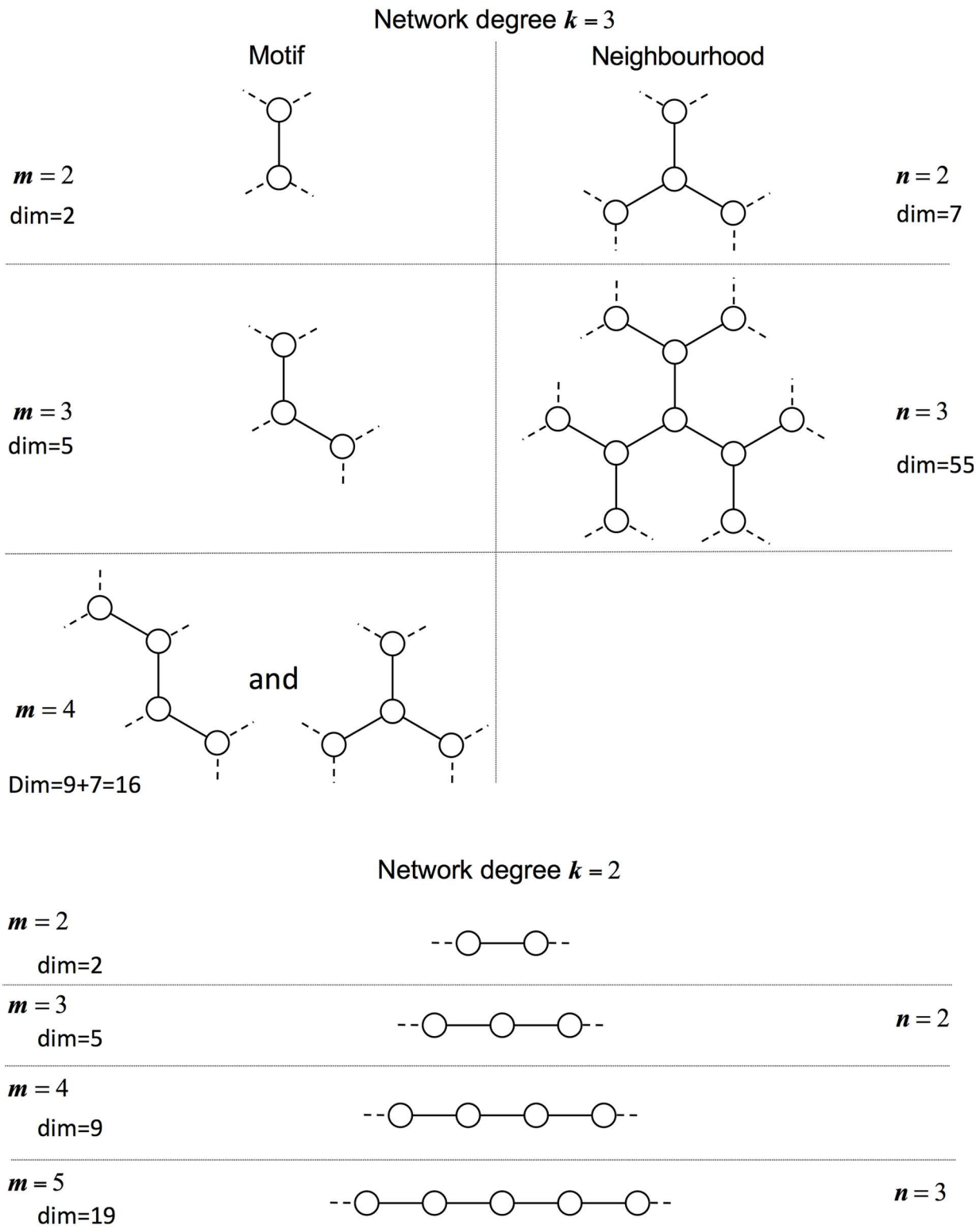

Fig 2. Subgraphs modelled by each approximation scheme. *k* denotes the (uniform) node degree and *m* or *n* indexing the size of the subgraph expansions for motif and neighbourhood schemes, respectively. Nodes within the subgraphs are shown and joined with solid lines; connections to other nodes (and hence where approximations are needed) are shown with dashed lines. The dimension of the associated ODEs for approximating the *SIS* dynamics is also given, accounting for symmetry and conservation. For degree $k = 2$ the dimension of the system can be calculated for a general *m* by considering the number of possible states with

symmetries dim = $2^{m-1} + 2^{M-1} - 1$ where $M = \lfloor (m+1)/2 \rfloor$. Note, in this case,the natural relationship between the motif and neighbourhood models when $m = 2n - 1$.

doi:10.1371/journal.pcbi.1005296.g002

where $\mathbb{1}$ is the indicator function. We use the pair-wise methodology and nomenclature developed by Keeling [23], where

$$\begin{aligned} [A] &:= \mathbb{E}\left[\sum_i \mathbb{1}_{\{X_i(t)=A_i\}}\right] , \\ [AB] &:= \mathbb{E}\left[\sum_{i,j} \mathbb{1}_{\{X_i(t)=A_i \& X_j(t)=B_j \& A_{i,j}=1\}}\right] , \end{aligned} \tag{2}$$

and similarly for larger structures.

Mean-field approximation. The following equations (and (Eq 5) below) can be shown to follow from (Eq 1) for any network [51]:

$$\frac{\mathrm{d}[S]}{\mathrm{d}t} = -\tau[SI] + \gamma[I] \ , \quad \frac{\mathrm{d}[I]}{\mathrm{d}t} = \tau[SI] - \gamma[I] \ . \tag{3}$$

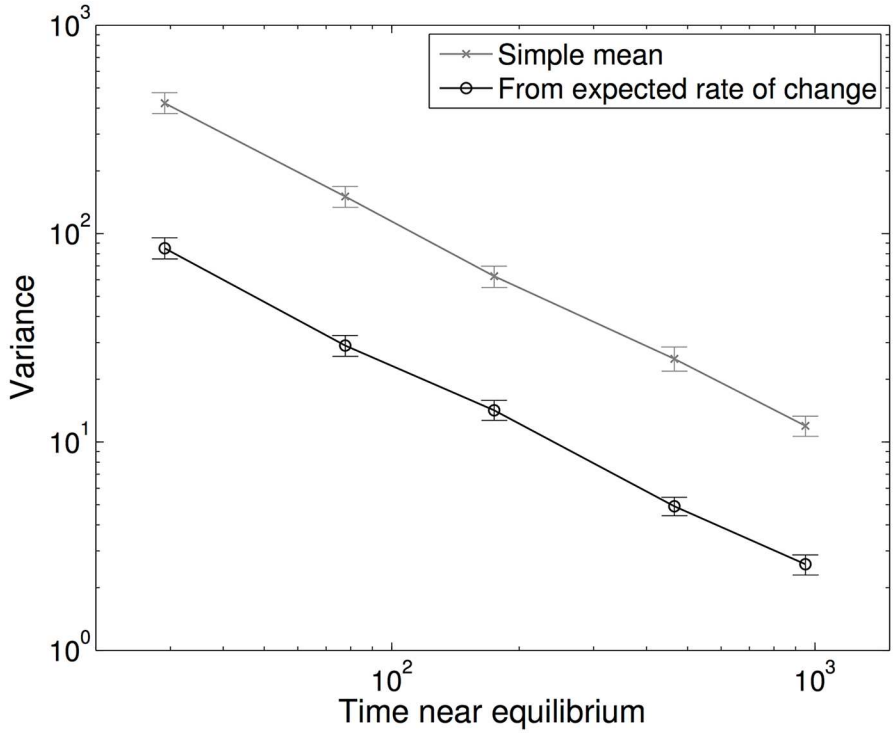

Fig 3. Variance about the true expected endemic level of prevalence. As calculated from very large simulations for two methods of determining the endemic prevalence for an SIS model on a network. Grey cross refer to taking the simple mean of the prevalence from time-series data, black circles are generated by fitting to the expected rates of change. Both are calculated once the simulation is close to its endemic state. Simulations are performed on a small network of 10,000 nodes for a relatively short time to highlight the differences; $k = 3$, $\tau = 1$, $\gamma = 1$.

doi:10.1371/journal.pcbi.1005296.g003

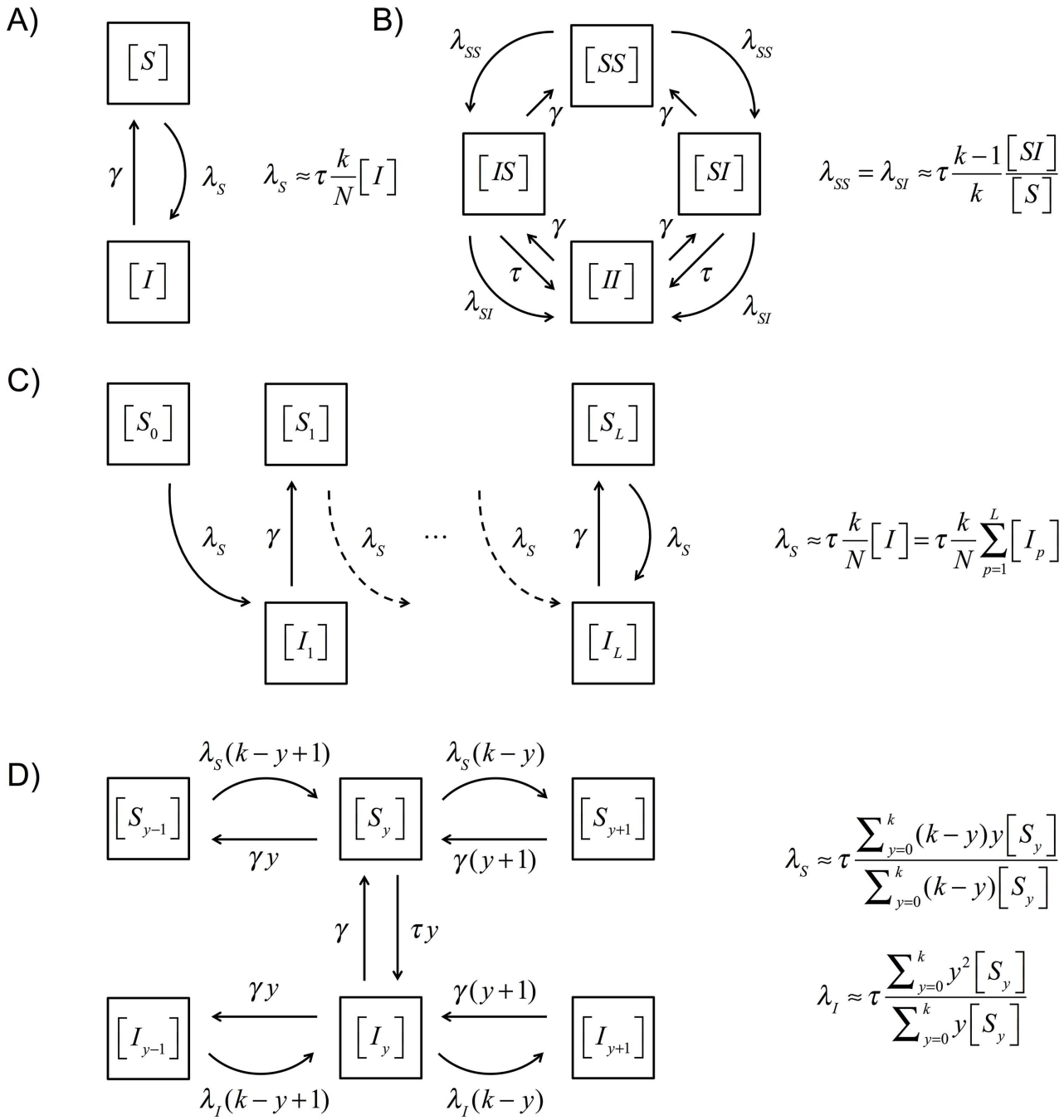

Fig 4. Models' states and transition rates for the systems obtained from the: A) Mean-field approximation (Eq 3) B) Standard pairwise approximation (Eq 6); C) Mean-field approximation with reinfection counting (Eq 7) and D) Neighbourhood model with $n = 2$ (Eq 9). Curved arrows represent transitions due to a force of infection coming from outside the single node or pair. Dashed arrows represent flows to and from compartments the dynamics of which are tracked but that are not drawn explicitly.

doi:10.1371/journal.pcbi.1005296.g004

The mean-field approximation for a k-regular network takes these together with the assumption

$$[SI] \approx (k/N)[S][I] \ . \tag{4}$$

Standard pairwise approximation; $m = 2$. Here we take (Eq 3) together with

$$\begin{aligned} \frac{\mathrm{d}[SS]}{\mathrm{d}t} &= 2\gamma[SI] - 2\tau[SSI] \ , \\ \frac{\mathrm{d}[SI]}{\mathrm{d}t} &= \gamma[II] + \tau[SSI] - \tau[SI] - \gamma[SI] - \tau[ISI] \ , \end{aligned} \tag{5}$$

noting that $[IS] = [SI]$ and $[II] = kN - [SS] - 2[SI]$. The standard pairwise approximation, typically attributed to Kirkwood [52], is

$$[ABC] \approx \frac{k-1}{k}\frac{[AB][BC]}{[B]} \ . \tag{6}$$

Following the work of [25, 30, 31] and [53], the approximation in (Eq 6) is exact for tree-like SIR epidemics when $B = S$, which is sufficient to guarantee the exactness of (Eq 5).

Systematic approximation 1: Reinfection counting. Here we count the number of times p that an individual has been infected up to a maximum of L. Using a straightforward notation (defined fully in S1 Text) the dynamics (Eq 3) become

$$\begin{aligned} \frac{\mathrm{d}[S_p]}{\mathrm{d}t} &= -\tau[S_p I] + \gamma[I_p] \ , \quad \forall p, \\ \frac{\mathrm{d}[I_p]}{\mathrm{d}t} &= \begin{cases} \tau[S_{p-1}I] - \gamma[I_p] \ , & p < L \ , \\ \tau[S_{p-1}I] + \tau[S_p I] - \gamma[I_p] \ , & p = L \ , \end{cases} \end{aligned} \tag{7}$$

where the subscript refers to the number of times an individual has been infected. Similarly, the dynamics (Eq 5) become

$$\begin{aligned} \frac{\mathrm{d}[S_p S_q]}{\mathrm{d}t} &= \gamma([S_p I_q] + [I_p S_q]) - \tau([S_p S_q I] + [IS_p S_q]) \ , \ \forall p, q \ , \\ \frac{\mathrm{d}[S_p I_q]}{\mathrm{d}t} &= \gamma[I_p I_q] + \tau[S_p S_{q-1} I] - \tau[S_p I_q] \\ &\quad - \gamma[S_p I_q] - \tau[IS_p I_q] \ , \quad q < L, \forall p \ , \\ \frac{\mathrm{d}[S_p I_L]}{\mathrm{d}t} &= \gamma[I_p I_L] + \tau[S_p S_{L-1} I] + \tau[S_p S_L I] - \tau[S_p I_L] \\ &\quad - \gamma[S_p I_L] - \tau[IS_p I_L] \ , \quad \forall p \ . \end{aligned} \tag{8}$$

We close these using the triple approximation (Eq 6), with the added notation that lack of a subscript refers to a sum over all possible infection counts (e.g. $[S_p\, I] = \Sigma_q[S_p\, I_q]$ or $[S_p\, S_q\, I] = \Sigma_r[S_p\, S_q\, I_r]$).

Systematic approximation 2: Motif models. To carry out a motif-based expansion at order m we write down the dynamics for the complete subgraphs of size m, whose rates of change will be functions of the complete subgraphs of sizes m and $m + 1$; these $m + 1$ subgraphs are then approximated using the general form of the Kirkwood closure, where the size-m motifs in the size-$(m + 1)$ motif are multiplied, and divided through by the over-counted size-$(m - 1)$ motifs, then divided through by the over-counted $m - 2$ motifs and so on until the

size-1 motifs are reached. This involves a large amount of notational development that is given in full for $m = 3$ in [54], and for the special case of $k = 2$ in S1 Text.

Systematic approximation 3: Neighbourhood model, $n = 2$. Suppose we write $[A_y]$ for the expected number of nodes in state A with y infectious neighbours, then the dynamics that follow from (Eq 1) are

$$\begin{aligned}\frac{\mathrm{d}[S_y]}{\mathrm{d}t} &= \gamma[I_y] + \lambda_S(k+1-y)[S_{y-1}] + \gamma(y+1)[S_{y+1}] \\ &\quad - \tau y[S_y] - \lambda_S(k-y)[S_y] - \gamma y[S_y] \ , \\ \frac{\mathrm{d}[I_y]}{\mathrm{d}t} &= \tau y[S_y] + \lambda_I(k+1-y)[I_{y-1}] + \gamma(y+1)[I_{y+1}] \\ &\quad - \gamma[I_y] - \lambda_I(k-y)[I_y] - \gamma y[I_y] \ .\end{aligned} \tag{9}$$

Where the forces of infection (λ_S and λ_I) refer to the rate of infection acting on a (susceptible) neighbour of the central node (in state S or I respectively). We then approximate the forces of infection by making the susceptible neighbour the centre of a new neighbourhood and looking for the distribution of consistent neighbourhoods:

$$\lambda_S \approx \tau \frac{\sum_{y=0}^{k}(k-y)y[S_y]}{\sum_{y=0}^{k}(k-y)[S_y]} \ , \quad \lambda_I \approx \tau \frac{\sum_{y=0}^{k} y^2[S_y]}{\sum_{y=0}^{k} y[S_y]} \tag{10}$$

This approach can be extended to order $n > 2$ by considering the dynamics in two parts: firstly the dynamics internal to each extended neighbourhood; secondly the force of infection on any susceptible individual at the edge of the extended neighbourhood. Again these forces of infection are found by considering consistent overlaps of the neighbourhoods centred on the susceptible neighbour under consideration.

Results

We begin by comparing the growth rates from four approximation models (mean-field, pairwise (motif $m = 2$), pairwise with reinfection counting ($L = 50$) and neighbourhood ($n = 2$), see S1 Text) with those from stochastic simulations on a Cayley tree, for different values of the transmission rate τ substantially above the critical value that permits successful invasion (Fig 5A, 5C and 5E). Unsurprisingly, the standard ODE model that ignores all elements of network structure (and hence ignores the negative *S-I* correlations that build-up and reduce transmission within a network) vastly over-estimates the early growth rate. Including some element of local structure, such as that captured by the pairwise (motif $m = 2$) model substantially improves the prediction of the growth rate but still overestimates compared to the simulated value. Finally, adding additional structure, either in terms of the reinfection counting or neighbourhood expansion enhances the accuracy. On closer inspection (Fig 5C and 5E) we observe that away from the critical invasion point, the reinfection counting model provides a highly accurate prediction of the early growth rate, outperforming all other approximation methods. In addition, as indicated by Fig 1, we find that for higher degree networks ($k = 6$, Fig 5E) all models, even the standard mean-field ODE model, provide a more accurate estimate of the true behaviour.

Turning our attention to the prevalence of infection (Fig 5B, 5D and 5F), it is clear that all approximation models (even the standard mean-field model) perform reasonably well when comparing their equilibrium values with the numerical estimates of the expected prevalence. As mentioned before, we also note that the standard pairwise model ($m = 2$) and reinfection counting pairwise model (for any L) have the same equilibrium prevalence—in the reinfection

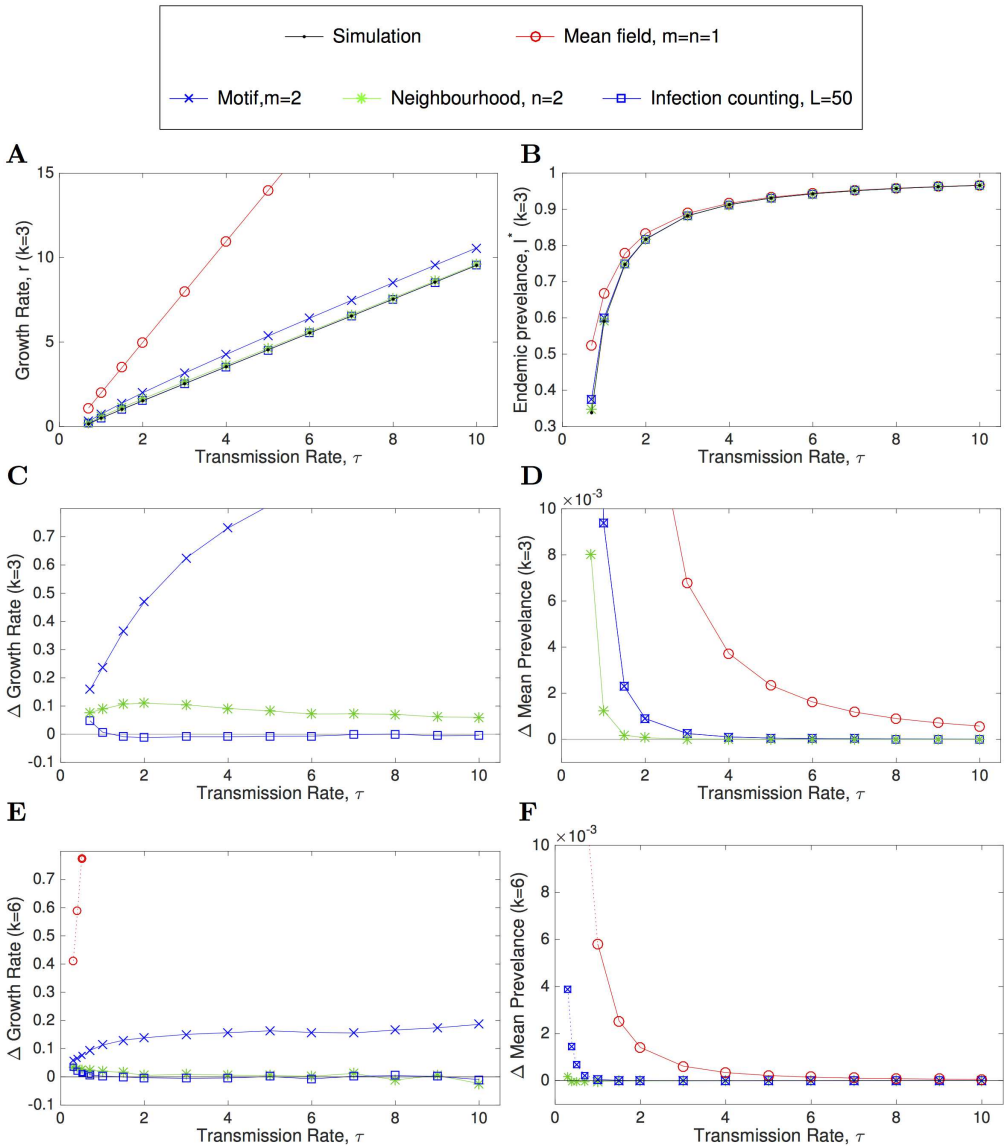

Fig 5. Growth rate (left column) and prevalence (right column) from a range of approximation models. Standard SIS model (Eq 3), $m = n = L = 1$); pairwise (motif) model (Eq 5, $m = 2$); neighbourhood model (Eq 9, $n = 2$); and reinfection counting (Eqs 7 and 8, $L = 50$). These are compared to findings from direct numerical simulation (Eq 1, black dots) for an SIS infection on a simple k-regular network. The upper panels show the absolute value of (A) the growth rate and (B) the prevalence as the transmission rate across a link (τ) is varied, for $k = 3$. The lower four panels show the difference between the approximations and the results from stochastic simulation (C and D: $k = 3$; E and F: $k = 6$). For $k = 6$ (panels E and F) additional smaller τ values (joined with dotted lines) are considered as the critical value that allows persistence is reduced compared to $k = 3$. Numerical simulations are performed on a 100,000 node network and run for sufficiently long that confidence intervals are negligible. A recovery rate $\gamma = 1$ is assumed throughout.

doi:10.1371/journal.pcbi.1005296.g005

counting model all individuals will eventually be infected more than L times, thereby reaching the upper limit. However, even taking L very large, the same quasi-equilibrium prevalence is reached even before a significant fraction of individuals hit the upper reinfection counting limit L. This is related to the loss of local correlation structure as the network becomes saturated with infection and paths of infection meet through medium and long loops within the network.

Comparing more closely results from the approximation models against simulated prevalence (Fig 5D and 5F) shows that the neighbourhood model ($n = 2$) outperforms the pairwise models ($m = 2$). This is to be expected as the neighbourhood model captures higher-order spatial structure within the network, effectively capturing the status of $k + 1$ connected individuals. However, all approximation models perform worse as the expected prevalence drops and the critical transmission rate is approached.

This comparison raises the question of how the motif and neighbourhood approximations perform as m and n are increased, incorporating more of the local network. We consider two cases. Firstly, $k = 3$ where only limited extensions of the models are feasible ($m = 3$, $m = 4$ and $n = 3$) as the dimension of the systems rapidly becomes large and the mathematical formulations are unwieldy. Secondly $k = 2$ (which we note is a special case [55, 56]) where neighbourhood and motif models are equivalent for $m = 2n - 1$, and where we can readily extend the approximation methods to extremely high orders (S1 Text). Fig 6 demonstrates the impact of taking these higher order approximations. Considering the $k = 2$ case (when the network is a linear system), increasing the order ($m = 1$ to $m = 16$) leads to a drop in endemic prevalence and convergence of the critical transmission rate to the estimated value (vertical line). At the

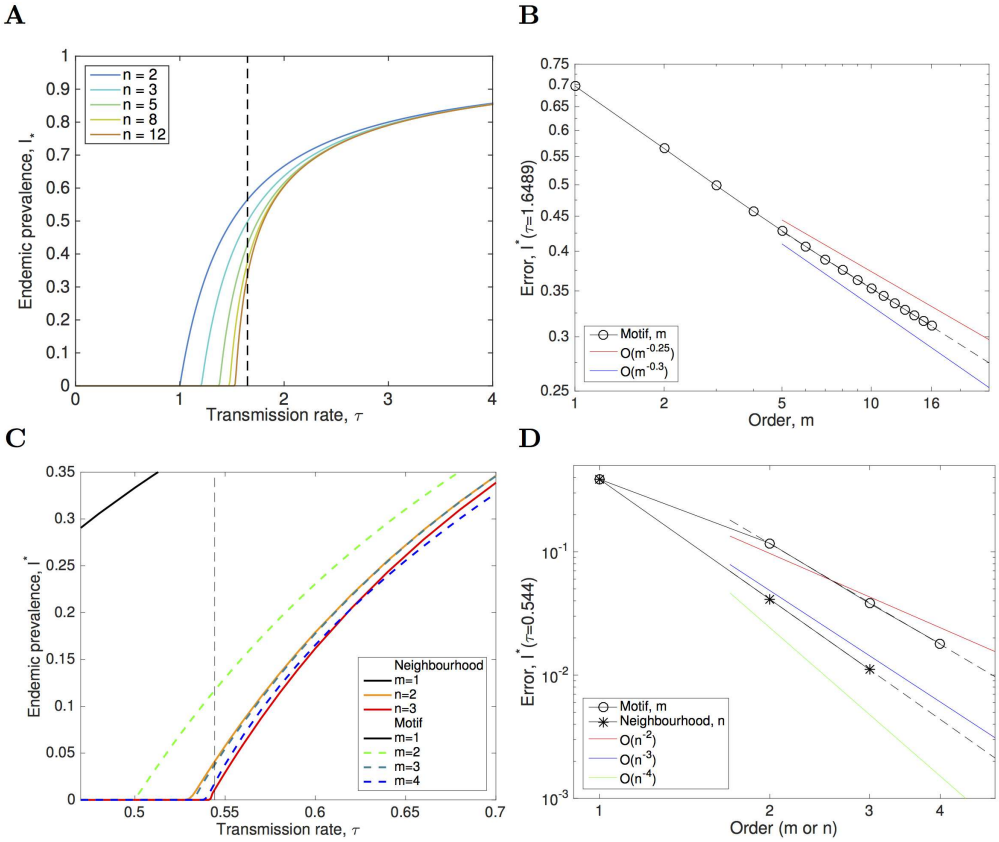

Fig 6. Endemic prevalence of infection for an SIS model for successively higher-order approximations. The focus is on the the value of the transmission rate near the critical point (shown as vertical dashed line), when $k = 2$ (A and B) and $k = 3$ (C, D). The errors at the critical point for the $k = 2$ degree network (B) show very slow convergence in the order of the approximation m; neighbourhood and motif model results coincide when $m = 2n - 1$ and are not reported. For $k = 3$ degree network, the errors at the estimated critical point (D) show rapid convergence as the models are extended to higher order (increasing m and n). The error in the motif model scales like $O(m^{-2.71})$ while for the much higher dimensional neighbourhood model the error scale like $O(n^{-3.24})$.

doi:10.1371/journal.pcbi.1005296.g006

critical transmission value (estimated as $\tau_C \approx 1.6489$ [57]) the error scales extremely slowly with the order of the model (approximately $O(m^{-0.271})$), Fig 6B). When returning to the case $k = 3$ that has been the main focus of this work (where we estimate $\tau_C = 0.544$ from numerous large scale simulations) both the approximation methods behave far better (motif error $\sim O(m^{-2.7})$; neighbourhood error $\sim O(n^{-3.2})$), and offer reliable predictions of endemic prevalence even quite close to the critical point as the order of the approximations increases.

Discussion

Moment closure approximations for the spread of infections on networks can be highly informative, especially when uncertainty in the underlying network structure precludes detailed simulation of a specific case. By generating relatively simple, tractable models (in the form of ODEs), an intuitive understanding can be developed for the spread of infection that does not rely on precise measurement of network structure. This approach has been highly successful for infections with SIR-type dynamics [23–25], where recovery leads to lifelong protection; however, for infections that obey the SIS paradigm and can therefore be contracted multiple times this closure approach does not have the same level of precision [40, 45, 48, 49]. Most sexually transmitted infections are well approximated by SIS-type dynamics, and modelling sexually transmitted infections requires an appreciation of the dynamic implications of the sexual contact network, due to the relatively low numbers of sexual contacts at any time. Therefore, although closure approximations for SIS-type infections on networks is highly challenging, it is nevertheless an area of considerable applied importance. Several other recent studies have considered the behaviour of SIS models on networks [42, 45, 49–51, 54, 58–61] showing that this is a field of active research where there are substantial challenges in establishing rigorous analytical results and in matching approximations, simulations and real data.

In this paper we have mainly focussed on homogeneous random networks where each individual has exactly $k = 3$ contacts, and all contacts are considered bi-directional. This restricts our attention to the highly challenging case of small and homogeneous degree, as higher mean degree or greater heterogeneity leads to infection prevalences that are closer to mean-field predictions that ignore the local correlations that arise from network structure. For homogeneous random networks of this type we show that, as expected, pairwise approximations (that consider the state of two connected individuals) outperform standard models (that ignore any correlations within the network), while neighbourhood-based models (originally called 'effective degree models' [22]—that consider the state of all neighbours around a central individual) outperform pairwise models (Fig 5A and 5B), and in turn extended neighbourhood models (that consider neighbours of neighbourhoods) are even more accurate (Fig 6C). This is unsurprising since closing the approximation at higher orders, and therefore essentially modelling more of the underlying local behaviour, is always likely to provide a more accurate description of the population-scale dynamics.

We also investigated extensions to the standard closure models, including a count of the number of times an individual has been infected. This removes some of the inaccuracies that pairwise (and other) approximation models suffer from when trying to capture the early stages of infection in a largely susceptible population. The results of this improvement to the pairwise model generates far better predictions for the early growth-rate of infection, offering a substantial improvement over both standard pairwise models and neighbourhood models (Fig 5). It would therefore seem prudent, although dimensionally-challenging, to combine reinfection counting with closure models that operate at the whole neighbourhood scale (or even larger) thereby enabling an approximation to both the early and endemic dynamics. However, it may

be far simpler to use the model most appropriate to the setting, depending on whether it is early growth or endemic prevalence that is required.

With moment closure models there is always the temptation to include higher order terms in the approximation and close at one order higher. We considered higher-order approximations to SIS dynamics on both $k = 3$ and $k = 2$ networks, noting that the SIS model on $k = 2$ is identical to classic 1-dimensional contact process [55, 56]. For the 1-dimensional $k = 2$ case, we are able to extend the modelling approach to much higher orders, but find that these closures still overestimate the prevalence near to the critical point. For the $k = 3$ network, we are only able to extend the neighbourhood model one additional step (considering neighbours of neighbours). However, as we capture the status of more contacts around the central individual, this extended neighbourhood model provides a very accurate approximation to the SIS behaviour, although it is complex to construct and relatively high-dimensional.

Throughout we have focused on expanding our approximations to ever higher orders, but the upper bounds to what can be achieved differ between methods. For the reinfection counting model (where dimension of the system is $2L^2 - 1$ and hence grows relatively slowly with L) taking L past 50 had very little effect, so further expansion was irrelevant. For degree $k = 2$ networks, where motif and neighbourhood expansions are equivalent, the limiting factor was the dimension of the system. The ODEs were simulated up to $m = 16$ at which point the dimension is $2^{m-1} + 2^{M-1} - 1 = 32{,}895$ (where $M = \lceil (m + 1)/2 \rceil = 8$). For degree $k = 3$ the main limitation is not the dimension of the system but the complexity of closure approximations which utilise the probabilities of overlapping subgraphs; considering approximation higher than $m = 4$ or $n = 3$ is possible, but the gains in accuracy may not be worth the considerable effort.

Finally, we consider the fact that sexual networks are highly heterogeneous (with some individuals having many more life-time partners than others [3, 62]). This risk heterogeneity is important for understanding who becomes infected, but the action of this heterogeneity is readily captured by traditional risk-structured mean-field models [63, 64] that ignore network correlations (Fig 1). Extending all the models discussed in this paper to capture degree heterogeneity is possible although one needs to specify, in addition to the degree distribution, a degree correlation matrix—the choice of which can have a dramatic impact on the disease dynamics [9]. Furthermore, the dimensionality of the system can rapidly exceed currently available computational resources due to the combinatorial number of possible configurations, especially for the neighbourhood model. For example, considering neighbourhood models $n = 2$: 5 equations are needed for $k = 2$ networks and 7 equations for $k = 3$ networks; however when the degree is heterogeneous far more equations are required, in a network where all nodes are degree 1 or 2 then 27 equations are needed but if nodes can be degree 1, 2 or 3 then the number of equations rises to 165. For neighbourhood models approximated at the next order ($n = 3$) the effect is even more dramatic; a heterogeneous networks where nodes can be degree 1, 2 or 3 requires 65,015 equations. We note, however, realistic sexual networks may have a large proportion of population with relatively few connections and where infections spread poorly; this leads us to believe that large sections of sexual networks may behave more like the low degree ($k = 2$ or $k = 3$) situations considered here where there is a strong need to accurately capture network correlations. As heterogeneity increases, pairwise models, which do not suffer such a pathological growth in dimensionality, become relatively accurate and, as was illustrated in Fig 1, even simple ODE models can be highly effective at capturing the aggregate prevalence in highly heterogeneous populations. Depending on the applied problem, we argue that a combination of the systematic approximations here can be used as a trade-off between accuracy and computational complexity.

Even if degree heterogeneity is captured by models, the network structure and infection dynamics used here are extreme simplifications of real-world behaviour: in particular, sexual

networks are dynamic (with most individuals practising serial monogamy [45]), and the natural history of infection is often far from the Markovian process with only two states (S and I) discussed here. In reality, for most STIs, there is likely to be a latent period following infection; detection, treatment and recovery will follow (non-Markovian) processes; and treatment is likely to offer some limited protection. These facets will act to prevent rapid reinfection of individuals, which, like heterogeneity, should improve the accuracy of approximation models.

Future work should clearly focus on developing more sophisticated and realistic network-based simulation models for STIs and comparing these to a range of approximation methods; however, we believe that the careful exploration of the accuracy of approximate models performed here is a key step in this process.

Supporting Information

S1 Text. Technical model definitions and methods. This Supplementary Material contains further technical model definitions and methods. Equations are numbered in continuation from the main text. Figures references and citations refer to the main text.
(PDF)

Acknowledgments

We would like to thank the WIDER group and Joss Ross for their helpful discussions and encouragement.

Author Contributions

Conceived and designed the experiments: MJK TH.

Performed the experiments: MJK TH AJC LP.

Analyzed the data: MJK TH.

Wrote the paper: MJK TH AJC LP.

References

1. Klovdahl AS. Social networks and the spread of infectious diseases: the AIDS example. Social Science & Medicine. 1985; 21(11):1203–1216. doi: 10.1016/0277-9536(85)90269-2 PMID: 3006260
2. Watts DJ, Strogatz SH. Collective dynamics of'small-world' networks. Nature. 1998; 393(6684):440–442. doi: 10.1038/30918 PMID: 9623998
3. Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y. The web of human sexual contacts. Nature. 2001; 411:907–908. doi: 10.1038/35082140 PMID: 11418846
4. Newman MEJ. Spread of epidemic disease on networks. Physical Review E. 2002; 66(1):016128. doi: 10.1103/PhysRevE.66.016128
5. Eubank S, Guclu H, Kumar V, Marathe M, Srinivasan A, Toroczkai Z, et al. Modelling disease outbreaks in realistic urban social networks. Nature. 2004; 429:180–184. doi: 10.1038/nature02541 PMID: 15141212
6. Aparicio J, Pascual M. Building epidemiological models from R0: an implicit treatment of transmission in networks. Proceedings of the Royal Society B: Biological Sciences. 2007; 274(1609):505. doi: 10.1098/rspb.2006.0057 PMID: 17476770
7. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proceedings of the National Academy of Sciences. 2009; 106(51):21484–21489. doi: 10.1073/pnas.0906910106 PMID: 20018697
8. Newman MEJ. Networks: an introduction. Oxford University Press, Inc.; 2010. doi: 10.1093/acprof:oso/9780199206650.001.0001

9. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, et al. Networks and the epidemiology of infectious disease. Interdisciplinary Perspectives on Infectious Diseases. 2011; 2011:284909. doi: 10.1155/2011/284909 PMID: 21437001

10. May RM, Lloyd AL. Infection dynamics on scale-free networks. Physical Review E. 2001; 64:art. no. 066112. doi: 10.1103/PhysRevE.64.066112

11. House T, Keeling MJ. Epidemic prediction and control in clustered populations. Journal of Theoretical Biology. 2011; 272(1):1–7. doi: 10.1016/j.jtbi.2010.12.009 PMID: 21147131

12. Ball F, Britton T, Sirl D. A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon. Journal of Mathematical Biology. 2012; 66(4–5):979–1019. doi: 10.1007/s00285-012-0609-7 PMID: 23161473

13. Ritchie M, Berthouze L, House T, Kiss IZ. Higher-order structure and epidemic dynamics in clustered networks. Journal of Theoretical Biology. 2014; 348:21–32. doi: 10.1016/j.jtbi.2014.01.025 PMID: 24486653

14. Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC. Network theory and SARS: predicting outbreak diversity. Journal of Theoretical Biology. 2005; 232(1):71–81. doi: 10.1016/j.jtbi.2004.07.026 PMID: 15498594

15. Munywoki PK, Koech DC, Agoti CN, Lewa C, Cane PA, Medley GF, et al. The Source of Respiratory Syncytial Virus Infection In Infants: A Household Cohort Study In Rural Kenya. Journal of Infectious Diseases. 2014; 209(11):1685–1692. doi: 10.1093/infdis/jit828 PMID: 24367040

16. Klovdahl AS, Potteratt JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. Social Networks and Infectious Disease—The Colorado-Springs Study. Social Science & Medicine. 1994; 38:79–88. PMID: 8146718

17. Pellis L, House T, Keeling MJ. Exact and approximate moment closures for non-Markovian network epidemics. Journal of Theoretical Biology. 2015; 382:160–177. doi: 10.1016/j.jtbi.2015.04.039 PMID: 25975999

18. Read J, Eames KTD, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. Journal of the Royal Society Interface. 2008; 5:1001–1007. doi: 10.1098/rsif.2008.0013 PMID: 18319209

19. Ball F, Neal P. Network epidemic models with two levels of mixing. Mathematical Biosciences. 2008; 212(1):69–87. doi: 10.1016/j.mbs.2008.01.001 PMID: 18280521

20. Volz E. SIR dynamics in random networks with heterogeneous connectivity. Journal of Mathematical Biology. 2008; 56:293–310. doi: 10.1007/s00285-007-0116-4 PMID: 17668212

21. Miller JC. A note on a paper by Erik Volz: SIR dynamics in random networks. Journal of Mathematical Biology. 2011; 62(3):349–358. doi: 10.1007/s00285-010-0337-9 PMID: 20309549

22. Lindquist J, Ma J, van den Driessche P, Willeboordse FH. Effective degree network disease models. Journal of Mathematical Biology. 2011; 62(2):143–164. doi: 10.1007/s00285-010-0331-2 PMID: 20179932

23. Keeling MJ. The effects of local spatial structure on epidemiological invasions. Proceedings Biological sciences / The Royal Society. 1999; 266(1421):859–867. doi: 10.1098/rspb.1999.0716 PMID: 10343409

24. House T, Keeling MJ. Insights from unifying modern approximations to infections on networks. Journal of the Royal Society Interface. 2011; 8(54):67–73. doi: 10.1098/rsif.2010.0179 PMID: 20538755

25. Sharkey KJ, Kiss IZ, Wilkinson RR, Simon PL. Exact equations for SIR epidemics on tree graphs. Bulletin of Mathematical Biology. 2015; 77(4):614–645. doi: 10.1007/s11538-013-9923-5 PMID: 24347252

26. Sharkey KJ, Wilkinson RR. Complete hierarchies of SIR models on arbitrary networks with exact and approximate moment closure. Mathematical Biosciences. 2015; 264:74–85. doi: 10.1016/j.mbs.2015.03.008 PMID: 25829147

27. Kiss IZ, Morris CG, Sélley F, Simon PL, Wilkinson RR. Exact deterministic representation of Markovian SIR epidemics on networks with and without loops. Journal of Mathematical Biology. 2015; 70(3):437–464. doi: 10.1007/s00285-014-0772-0 PMID: 24590574

28. Pellis L, House T, Keeling MJ. Exact and approximate moment closures for non-Markovian network epidemics. Journal of Theoretical Biology. 2015; 382:160–177. doi: 10.1016/j.jtbi.2015.04.039 PMID: 25975999

29. Decreusefond L, Dhersin JS, Moyal P, Tran VC. Large graph limit for an SIR process in random network with heterogeneous connectivity. The Annals of Applied Probability. 2012; 22(2):541–575.

30. Bohman T, Picollelli M. SIR epidemics on random graphs with a fixed degree sequence. Random Structures and Algorithms. 2012; 41(2):179–214. doi: 10.1002/rsa.20401

31. Barbour A, Reinert G. Approximating the epidemic curve. Electronic Journal of Probability. 2013; 18 (54):1–30. doi: 10.1214/EJP.v18-2557

32. Janson S, Luczak M, Windridge P. Law of large numbers for the SIR epidemic on a random graph with given degrees. Random Structures and Algorithms. 2014; 45(4):726–763. doi: 10.1002/rsa.20575

33. Kamp C, Moslonka-Lefebvre M, Alizon S. Epidemic spread on weighted networks. PLoS Computational Biololgy. 2013; 9(12):e1003352. doi: 10.1371/journal.pcbi.1003352

34. Miller JC, Slim AC, Volz EM. Edge-based compartmental modelling for infectious disease spread. Journal of the Royal Society Interface. 2012; 9(70):890–906. doi: 10.1098/rsif.2011.0403 PMID: 21976638

35. Karrer B, Newman MEJ. Message passing approach for general epidemic models. Physical Review E. 2010; 82(1):016101. doi: 10.1103/PhysRevE.82.016101 PMID: 20866683

36. Karrer B, Newman MEJ, Zdeborová L. Percolation on sparse networks. Physical Review Letters. 2014; 113(20):208702. doi: 10.1103/PhysRevLett.113.208702 PMID: 25432059

37. Wilkinson RR, Sharkey KJ. Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks. Physical Review E. 2014; 89(2):022808. doi: 10.1103/PhysRevE.89.022808 PMID: 25353535

38. Kamp C. Untangling the Interplay between Epidemic Spread and Transmission Network Dynamics. PLoS Computational Biology. 2010; 6(11):1–9. doi: 10.1371/journal.pcbi.1000984

39. Leung KY, Diekmann O. Dangerous connections: on binding site models of infectious disease dynamics. Journal of Mathematical Biology. 2016;. doi: 10.1007/s00285-016-1037-x PMID: 27324477

40. Ferguson NM, P GG. More realistic models of sexually transmitted disease transmission dynamics—Sexual partnership networks, pair models, and moment closure. Sexually Transmitted Diseases. 2000; 27:600–609. PMID: 11099075

41. Berger N, Borgs C, Chayes JT, Saberi A. On the spread of viruses on the internet. In: Proceedings of the 16th Symposium on Discrete Algorithms; 2005.

42. Durrett R. Random Graph Dynamics. Cambridge University Press; 2007. doi: 10.1017/CBO9780511546594

43. Chatterjee S, Durrett R. Contact processes on random graphs with power law degree distributions have critical value 0. The Annals of Probability. 2009; 37(6):2332–2356. doi: 10.1214/09-AOP471

44. Durrett R. Some features of the spread of epidemics and information on a random graph. Proceedings of the National Academy of Sciences. 2010; 107(10):4491–4498. doi: 10.1073/pnas.0914402107 PMID: 20167800

45. Eames KTD, Keeling MJ. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. Proceedings of the National Academy of Sciences. 2002; 99(20):13330–13335. doi: 10.1073/pnas.202244299 PMID: 12271127

46. Edwards R, Kim S, van den Driessche P. A multigroup model for a heterosexually transmitted disease. Mathematical Biosciences. 2010; 224(2):87–94. doi: 10.1016/j.mbs.2009.12.008 PMID: 20043928

47. Molloy M, Reed B. A Critical-Point for Random Graphs with a Given Degree Sequence. In: Random Structures & Algorithms. vol. 6. Wiley Online Library; 1995. p. 161–179. doi: 10.1002/rsa.3240060204

48. Bauch CT. A versatile ODE approximation to a network model for the spread of sexually transmitted diseases. Journal of Mathematical Biology. 2002; 45:375–395. doi: 10.1007/s002850200153 PMID: 12424529

49. Taylor M, Simon PL, Green DM, House T, Kiss IZ. From Markovian to pairwise epidemic models and the performance of moment closure approximations. Journal of Mathematical Biology. 2012; 64:1021–1042. doi: 10.1007/s00285-011-0443-3 PMID: 21671029

50. Taylor TJ, Kiss IZ. Interdependency and hierarchy of exact and approximate epidemic models on networks. Journal of Mathematical Biology. 2014; 69:183–211. doi: 10.1007/s00285-013-0699-x PMID: 23739839

51. Simon PL, Taylor M, Kiss IZ. Exact epidemic models on graphs using graph-automorphism driven lumping. Journal of Mathematical Biology. 2011; 62(4):479–508. doi: 10.1007/s00285-010-0344-x PMID: 20425114

52. Kirkwood JG. Statistical Mechanics of Fluid Mixtures. Journal Chemical Physics. 1935; 3:300. doi: 10.1063/1.1749657

53. Miller JC, Kiss IZ. Epidemic Spread in Networks: Existing Methods and Current Challenges. Mathematical Modelling of Natural Phenomena. 2014; 9(2):4–42. doi: 10.1051/mmnp/20149202 PMID: 25580063

54. House T, Davies G, Danon L, Keeling MJ. A motif-based approach to network epidemics. Bulletin of Mathematical Biology. 2009; 71(7):1693–1706. doi: 10.1007/s11538-009-9420-z PMID: 19396497

55. Durrett R. On the Growth of One Dimensional Contact Processes. The Annales of Probability. 1980; 8:890–907. doi: 10.1214/aop/1176994619

56. Griffeath D. The basic contact processes. Stochastic Processes & their Applications. 1981; 11:151–185. doi: 10.1016/0304-4149(81)90002-8

57. de Mendonca J. Precise critical exponents for the basic contact process. J Phys A: Math Gen. 1999; 32: L467–L473.

58. Floyd W, Kay L, Shapiro M. A Covering-Graph Approach to Epidemics on SIS and SIS-Like Networks. Bulletin of Mathematical Biology. 2012; 74(1):175–189. doi: 10.1007/s11538-011-9684-y PMID: 21989564

59. Lee HK, Shim PS, Noh JD. Epidemic threshold of the susceptible-infected-susceptible model on complex networks. Physical Review E. 2013; 87:062812. doi: 10.1103/PhysRevE.87.062812 PMID: 23848734

60. Wilkinson RR, Sharkey KJ. An Exact Relationship Between Invasion Probability and Endemic Prevalence for Markovian SIS Dynamics on Networks. PLoS ONE. 2013; 8(7):1–8. doi: 10.1371/journal.pone.0069028 PMID: 23935916

61. Simon PL, Kiss IZ. Super compact pairwise model for SIS epidemic on heterogeneous networks. Journal of Complex Networks. 2016; 4:187–200. doi: 10.1093/comnet/cnv018

62. Gupta S, Anderson RM, May RM, et al. Networks of sexual contacts: implications for the pattern of spread of HIV. AIDS. 1989; 3(12):807. PMID: 2517202

63. Anderson RM, May RM. Infectious diseases of humans. Oxford University Press; 1992.

64. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton University Press; 2008.