

## Exact deterministic representation of Markovian SIR epidemics on networks with and without loops

Article (Accepted Version)

Kiss, Istvan Z, Morris, Charles G, Sélley, Fanni, Simon, Péter L and Wilkinson, Robert R (2015) Exact deterministic representation of Markovian SIR epidemics on networks with and without loops. *Journal of Mathematical Biology*, 70 (3). pp. 437-464. ISSN 0303-6812

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/58841/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# Exact deterministic representation of Markovian *SIR* epidemics on networks with and without loops

Istvan Z. Kiss<sup>1,\*</sup>, Charles G. Morris<sup>1</sup>, Fanni Sélley<sup>2</sup>,  
Péter L. Simon<sup>2</sup> & Robert R. Wilkinson<sup>3</sup>

November 16, 2015

<sup>1</sup> School of Mathematical and Physical Sciences, Department of Mathematics, University of  
Sussex, Falmer, Brighton BN1 9QH, UK

<sup>2</sup> Institute of Mathematics, Eötvös Loránd University Budapest, and Numerical Analysis and  
Large Networks Research Group, Hungarian Academy of Sciences, Hungary

<sup>3</sup> Department of Mathematical Sciences, The University of Liverpool, Peach Street, Liverpool  
L69 7ZL, UK

\* corresponding author  
email: i.z.kiss@sussex.ac.uk

## Abstract

In a previous paper Sharkey et al. [14] proved the exactness of closures at the level of triples for Markovian *SIR* (susceptible-infected-removed) dynamics on tree-like networks. This resulted in a deterministic representation of the epidemic dynamics on the network that can be numerically evaluated. In this paper, we extend this modelling framework to certain classes of networks exhibiting loops. We show that closures where the loops are kept intact are exact, and lead to a simplified and numerically solvable system of ODEs (ordinary-differential-equations). The findings of the paper lead us to a generalisation of closures that are based on partitioning the network around nodes that are cut-vertices (i.e. the removal of such a node leads to the network breaking down into at least two disjointed components or subnetworks). Exploiting this structural property of the network yields some natural closures, where the evolution of a particular state can typically be exactly given in terms of the corresponding or projected states on the subnetworks and the cut-vertex. A byproduct of this analysis is an alternative probabilistic proof of the exactness of the closures for tree-like networks presented in Sharkey et al. [14]. In this paper we also elaborate on how the main result can be applied to more realistic networks, for which we write down the ODEs explicitly and compare output from these to results from simulation. Furthermore, we give a general, recipe-like method of how to apply the reduction by closures technique for arbitrary networks, and give an upper bound on the maximum number of equations needed for an exact representation.

**Keywords:** master equation, network, closure, loop, cut-vertex

# 1 Introduction

Despite tremendous progress over the past decade or so, modelling transmission processes on networks still poses many challenges. A significant number of such models are concerned with modelling epidemics on networks, in particular *SIR* dynamics which makes the treatment of some models easier due to the linear transmission process, as opposed to *SIS* dynamics where nodes can become reinfected multiple times. There is a wealth of modelling approaches to this problem [2, 6, 7, 18] that differ in the choice of variables at which models are formulated, and whether averages are taken at the population level, or a probabilistic view is kept whereby either the full state space is considered [16], or where, modelling starts at node level [12, 13, 14].

A major further challenge is posed by extending existing results for loopless networks to networks with loops or clustered networks, where clustering is defined as the propensity that any two nodes that share a common neighbour are also connected. Obviously, clustering and the presence of loops is closely related (i.e. the presence of many closed loops of size three leads to high levels of clustering). The specific issues cluster around the generation of synthetic networks with tuneable clustering [5, 10, 19], as well as the development of low-dimensional approximate or mean-field models, with the aim to match output directly from the stochastic process, namely a solvable, exact model or stochastic simulations. Some progress in both areas has been made, with final epidemic size calculations (non-time-dependent measures) giving excellent agreement for specific clustered networks [1, 4, 9, 11]. There are also examples for good time-evolution models [6, 20], but many important and difficult questions remain.

Using the approach introduced in [12, 13, 14], we present exact, deterministic representations of Markovian *SIR* epidemics on networks with and without loops and identify the link between the structural properties of the networks and the viability of closures that allow us to write down exact systems that can be numerically evaluated. Here, the equations start at the level of nodes and consider the exact probability of nodes being susceptible, infected or recovered at a given time, see [12, 13, 14]. In particular, we show the link between nodes that are cut-vertices and edges that are bridges (both defined later), and the feasibility of closures. Assuming a network with  $N$  nodes with the weighted connectivity and transmissibility rate matrix given by  $T = (T_{ij})_{i,j=1,2,\dots,N}$ , where  $T_{ii} = 0 \forall i = 1, 2, \dots, N$ , the evolution equations for singles and pairs are given by,

$$\begin{aligned}
 \langle \dot{S}_i \rangle &= - \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle, \\
 \langle \dot{I}_i \rangle &= \sum_{j=1}^N T_{ij} \langle S_i I_j \rangle - \gamma_i \langle I_i \rangle, \\
 \langle \dot{S}_i I_j \rangle &= \sum_{k=1, k \neq i}^N T_{jk} \langle S_i S_j I_k \rangle - \sum_{k=1, k \neq j}^N T_{ik} \langle I_k S_i I_j \rangle \\
 &\quad - T_{ij} \langle S_i I_j \rangle - \gamma_j \langle S_i I_j \rangle, \\
 \langle \dot{S}_i S_j \rangle &= - \sum_{k=1, k \neq j}^N T_{ik} \langle I_k S_i S_j \rangle - \sum_{k=1, k \neq i}^N T_{jk} \langle S_i S_j I_k \rangle,
 \end{aligned} \tag{1}$$

where  $\langle A_i \rangle$  denotes the time-dependent probability for individual  $i$  being in state  $A$ , and expressions of the form  $\langle A_i B_j \rangle$  denote the time-dependent probability that individuals  $i$  and  $j$  are in states  $A$  and  $B$ , respectively. The entries of the  $T$  matrix are of the form  $T_{ij} = \tau w_{ij}$ , where  $\tau$  can be thought of as a baseline transmission rate scaled by a weight  $w_{ij}$  accounting for

the strength of the link from node  $j$  to node  $i$ . It immediately follows that  $T_{ij} = 0$  implies that there is no directed link from  $j$  to  $i$ . We assume that all processes, i.e. infection and recovery, are independent Poisson processes, with per-link infection rate denoted by  $\tau$  and absorbed in  $T$  and rates of recovery  $\gamma_j$  ( $j = 1, 2, \dots, N$ ). While this is a general model formulation from the network view point, in this paper all numerical simulations are carried out using undirected and unweighted networks with the per-contact transmission rate specified explicitly, and with the same recovery rate for all nodes.

The system above is not closed as equations for the triples are needed. In Sharkey et al. [14], the authors have proved that for tree-like networks and for some special cases of non-tree-like networks the following closures hold and are exact

$$\langle S_j \rangle \langle S_i S_j I_k \rangle = \langle S_i S_j \rangle \langle S_j I_k \rangle, \quad (2)$$

for all  $i, j, k \in \{1, 2, \dots, N\}$ , and for all  $j$  with links towards  $i$  and all  $k$  with links towards  $j$  and  $i \neq k$  (i.e.  $T_{ij} \neq 0$  and  $T_{jk} \neq 0$ ), and

$$\langle S_i \rangle \langle I_k S_i I_j \rangle = \langle I_k S_i \rangle \langle S_i I_j \rangle, \quad (3)$$

for all  $i, j, k \in \{1, 2, \dots, N\}$  and for all  $k$  and  $j$  with links towards  $i$  (i.e.  $T_{ik} \neq 0$  and  $T_{ij} \neq 0$ ), and  $j \neq k$ . Closures are exact in the sense that the closed system,

$$\begin{aligned} \langle \dot{X}_i \rangle &= - \sum_{j=1}^N T_{ij} \langle X_i Y_j \rangle, \\ \langle \dot{Y}_i \rangle &= \sum_{j=1}^N T_{ij} \langle X_i Y_j \rangle - \gamma_i \langle Y_i \rangle, \\ \langle X_i \dot{Y}_j \rangle &= \sum_{k=1, k \neq i}^N T_{jk} \frac{\langle X_i X_j \rangle \langle X_j Y_k \rangle}{\langle X_j \rangle} - \sum_{k=1, k \neq j}^N T_{ik} \frac{\langle X_i Y_k \rangle \langle X_i Y_j \rangle}{\langle X_i \rangle} \\ &\quad - T_{ij} \langle X_i Y_j \rangle - \gamma_j \langle X_i Y_j \rangle, \\ \langle X_i \dot{X}_j \rangle &= - \sum_{k=1, k \neq j}^N T_{ik} \frac{\langle Y_k X_i \rangle \langle X_i X_j \rangle}{\langle X_i \rangle} - \sum_{k=1, k \neq i}^N T_{jk} \frac{\langle X_i X_j \rangle \langle X_j Y_k \rangle}{\langle X_j \rangle}, \end{aligned} \quad (4)$$

is such that when  $\langle X_i \rangle = \langle S_i \rangle$ ,  $\langle Y_i \rangle = \langle I_i \rangle$ ,  $\langle X_i Y_j \rangle = \langle S_i I_j \rangle$  and  $\langle X_i X_j \rangle = \langle S_i S_j \rangle$  hold at  $t = 0$ , then these will hold  $\forall t > 0$ . Similar equalities hold for the pairs. To emphasise that this second system is an ‘approximation’, we use  $X$  for susceptible and  $Y$  for infected. The main result of the original paper [14], on which we now build, is the proof that for a tree-like network, the closure is exact in the sense that solving the closed system we get the same values for the probabilities.

Regarding the original exact equations (Eq. (1)) we make the following remarks:

1. *The equations emerge naturally starting from nodes and building up to higher moments. The equations do not cover every possible configuration of states across connected subnetworks. For example, the exact equations do not require knowledge of pairs such as  $\langle I_i I_j \rangle$  as these are not required by the system dynamics.*
2. *All the triples that appear are such that the middle node is susceptible. In a tree-like network this effectively means that the nodes to the ‘left’ and to the ‘right’ of the middle node have not yet communicated since the only way is through the middle susceptible node, and thus, technically their states are independent, conditional on the middle node being susceptible.*

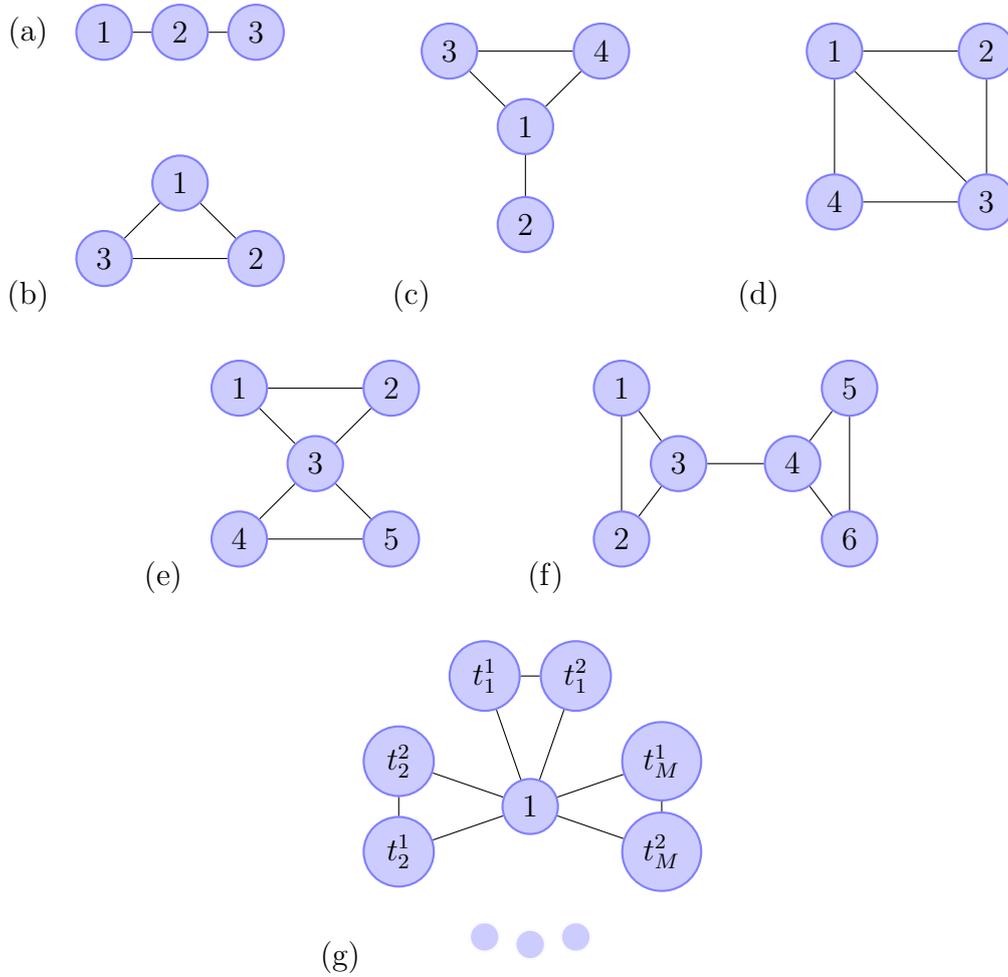


Figure 1: (a) Open triangle or line network of three nodes, (b) closed triangle, (c) lollipop, (d) toast, (e) bow tie, (f) bow tie with a bridge, and (g) star-triangle (i.e. a star of triangles) network with  $M$  triangles.

3. *The closures will require extra variables that have previously not been needed, for example, evolution equations for (SS) pairs (e.g.  $\langle S_i S_j \rangle$ ) are needed. This is required by the closure of an (SSI) triple (e.g.  $\langle S_i S_j I_k \rangle$ ).*
4. *Additional closures beyond what the equations require may not hold (e.g.  $\langle S_i I_j I_k \rangle \langle I_j \rangle \neq \langle S_i I_j \rangle \langle I_j I_k \rangle$ ), but this is not relevant as these are not needed to derive a closed, exact system with fewer equations.*
5. *Given that we deal with a SIR model, each node can be either S, I or R. Knowing the probability of an arbitrary node  $i$  being susceptible ( $\langle S_i \rangle(t)$ ) or infected ( $\langle I_i \rangle(t)$ ) at time  $t$ , automatically gives the probability of this node being in state R ( $\langle R_i \rangle(t) = 1 - \langle S_i \rangle(t) - \langle I_i \rangle(t)$ ). This implies that **it** is enough to concentrate on working out  $\langle S_i \rangle(t)$  and  $\langle I_i \rangle(t)$  for all  $i = 1, 2, \dots, N$  and  $t \geq 0$ . However, working out the probabilities at node level, will only involve pairs in state (SI), and the probabilities for such pairs will only involve triples of type (SSI) and (ISI). Hence, strictly speaking in order to fully specify the probabilities at node level, we do not need to work out quantities such as  $\langle S_i R_j \rangle$ ,  $\langle I_i R_j \rangle$ ,  $\langle R_i R_j \rangle$  or  $\langle I_i I_j \rangle$ . We chose not to include these extra variables in order to reduce the number of equations as much as possible at the expense of only having information about what really matters, and what is usually important from an application viewpoint. Taking this avenue, and using the model as is, we could not specify the probability of an arbitrary link between nodes  $i$  and  $j$  being in any of the  $(S_i R_j)$ ,  $(I_i R_j)$ ,  $(R_i R_j)$  or  $(I_i I_j)$  states. However, this does not preclude computing exactly the prevalence and final epidemic size. If one wishes, the equations can be extended to compute these quantities at the expense of a more complex system with a higher number of equations.*

The paper is structured as follows. In Section 2 a number of candidate closures for different toy networks are tested by using the full/exact equations. It is shown that while some closures hold, there are cases where closures fail. It turns out that this provides some intuition towards identifying the link between network properties and the possibility of valid closures. This is consolidated via a Theorem, i.e. the main result of the paper, in Section 3, where the clear link between network properties and general closures is identified. In Section 4 applications of the main results are shown, and the applicability of closures in general is discussed. Finally, possible extensions, further challenges and research ideas are considered.

## 2 Background and examples

An *SIR* epidemic on an arbitrary directed and weighted network can be represented as a set of equations starting at the level of individuals and by building up to and accounting for the dependencies of these on pairs, and of the pairs on triples, and so on, until full system size is reached. At this point the system of equations will become well-defined and self-consistent. The aim in this type of modelling approach is to find closures, where higher order moments can be approximated or specified exactly in terms of combinations of lower order moments. It is now well-known and accepted that this is feasible for tree-like networks and for networks with loops, but starting from some very specific initial conditions or particular example networks [7, 14]. Here, we will show a more general approach to extend the ideas of closures to networks with loops/cycles and we will also consider specific fully worked-out examples, as well as how this approach could be generalised to and its feasibility for a larger class of networks.

### Testing closures: examples where closures do and do not work

Before we generalise the ideas of closures to networks with loops, let us consider the special

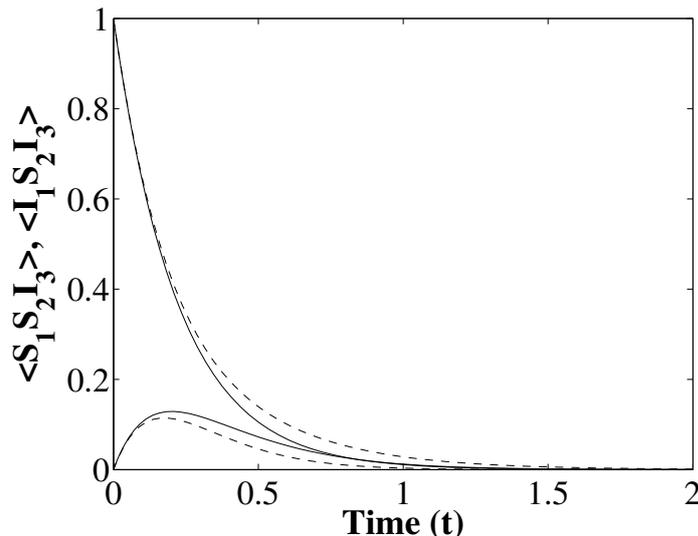


Figure 2: Testing of two closures on the closed triangle (see Fig. 1b),  $\langle S_1 S_2 I_3 \rangle = \frac{\langle S_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle}$  and  $\langle I_1 S_2 I_3 \rangle = \frac{\langle I_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle}$ , by plotting the left and right hand sides of the closures from the full system, see Appendix 6. Continuous lines represent the left hand sides, and dashed lines represent the right hand sides for the  $(SSI)$  and  $(ISI)$  triples. The difference between the continuous and dashed lines indicates that the two closures are not exact. System started such that  $\langle S_1 S_2 I_3 \rangle(0) = 1$ , and hence,  $\langle I_1 S_2 I_3 \rangle(0) = 0$ . Parameter values for all cases are  $\tau = 7/4$  and  $\gamma = 1$ .

cases of small networks with and without loops, namely the open and closed triangle, lollipop, and toast networks presented in Fig. 1. For such small networks, the full system of equations can be written down and closures can be tested. The case of an open triangle, or a line network of three nodes, is discussed explicitly in [14], where an analytical proof for the exactness of the closure is presented with calculations relying explicitly on the full-system of equations. The simplest and most obvious network with loops (or a loop) is the closed triangle, see Fig. 1b. In Appendix 6, we list the full system of differential equations, and we use this to show that triple closures, such as those presented in Eqs. (2- 3), do not hold for closed loops. In Fig. 2 we show that the  $(SSI)$ - and  $(ISI)$ -type triples cannot be closed exactly. This is based on numerically evaluating the full system, and thus, being able to compare the closure via the exact time evolution of the closures' constituent parts. This simple analysis suggests that loops cannot be closed, and it is likely that, for exact closures loops need to be kept intact.

Keeping the loops intact is a feasible approach and we illustrate this for the lollipop network (Fig. 1c). We start by generating and writing down the full set of equations for the lollipop network, see Appendix 7. These include two sets of equations. First, the naturally emerging set of equations which can be broken down into those that can, Eqs. (29-50), and cannot be closed, Eqs. (51-63). These together give rise to the natural full system (NFS). Second, the non-closable equations together with the extra variables required by the closures give rise to a reduced system (RS), see Eqs. (29-50) plus Eqs. (77-80). Figure 3 gives clear numerical evidence that closures at the full system level that keep the closed triangle complete are exact (see the left panel of Fig. 3). Moreover, in the same figure (see right panel), we compare the prevalence resulting from the NFS and the RS, and this clearly illustrates that the proposed closures are likely to be exact. By looking at the full set of equations for the lollipop network one can notice that node 1 always appears as a susceptible node whenever appearing in a triple or quadruple

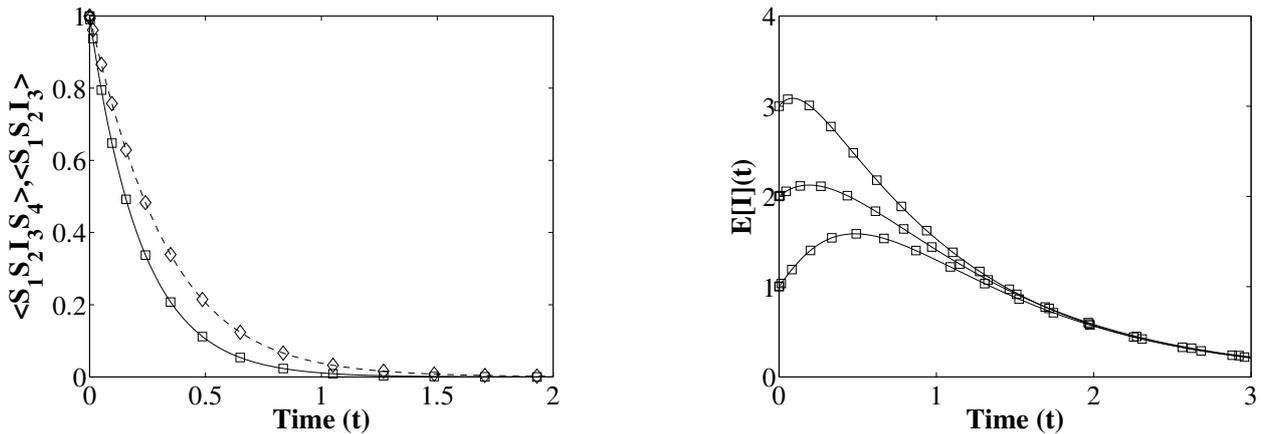


Figure 3: *Left panel:* Testing of two closures for the lollipop network (see Fig. 1c),  $\langle S_1 S_2 I_3 S_4 \rangle = \frac{\langle S_1 S_2 \rangle \langle S_1 I_3 S_4 \rangle}{\langle S_1 \rangle}$  and  $\langle S_1 S_2 I_3 \rangle = \frac{\langle S_1 S_2 \rangle \langle S_1 I_3 \rangle}{\langle S_1 \rangle}$ , by plotting the left and right hand sides of the closures from the full system, see Appendix 7. Continuous and dashed lines represent the left hand sides of the quadruple and triangle, respectively, and  $(\square)$  and  $(\diamond)$  represent the right hand sides of the corresponding closures. System started such that  $\langle S_1 S_2 I_3 S_4 \rangle(0) = 1$ . The excellent agreement shows that closures are likely to be exact. *Right panel:* The expected prevalence over time for the lollipop network from the natural full (continuous line) and reduced systems  $(\square)$ , as given in Appendix 7. Starting from three different initial conditions: (a)  $\langle S_1 I_2 I_3 I_4 \rangle(0) = 1$ , (b)  $\langle S_1 S_2 I_3 I_4 \rangle(0) = 1$  and (c)  $\langle S_1 S_2 I_3 S_4 \rangle(0) = 1$ . Parameter values for all cases are  $\tau = 7/4$  and  $\gamma = 1$ .

which is a candidate for closure. This is an important structural property of the node within the network that will be elaborated later on. We also note that two main types of closures emerge: (a) closures at the level of triples which are not part of the closed triangle and (b) closures at the level of the full system in a way in which the quad is broken down into the closed triangle and the tail-edge of the lollipop.

Moving towards networks with more or multiple loops we consider the toast network (see Fig. 1d), with full equations for this given in Appendix 8. In Fig. 4, we give examples based on the full system and simulation. The left panel of the figure clearly shows that the proposed closure does not hold. For the toast network the problem of the closure is more complex as it is possible to write down closures that hold, see the right panel of Fig. 4. However, the reduction in the number of equations due to these closures is not significant. Hence, again it is clear that the equations cannot be closed, the loops and the whole network need to be kept intact.

Several tests can be performed to test the validity of closures. First, the validity of closures can be tested directly from the full system using the NFS. Second, the NFS and RS can be compared via the evolution of prevalence in time. We also note, that when the full system is available it is possible to give an analytic proof that closures hold. This involves rearranging the closure relation as a difference, for example as  $\alpha(t) = \langle \cdot \rangle \langle \cdot \rangle - \langle \cdot \rangle \langle \cdot \rangle$ . This is then followed by showing that  $\dot{\alpha}(t) = 0 \forall t \geq 0$ , which coupled with the closure holding at time  $t = 0$  gives the desired result. The calculations will involve other closure-like or  $\alpha$ -like expressions, but it will be possible to show that all such closures are such that they satisfy a  $\dot{\alpha}(t) = -C\alpha(t)$  equation, where  $C > 0$ , see Sharkey et al. [14]. For larger systems (for example the bow tie in Fig. 1e), the terms entering the closures can be evaluated from direct stochastic simulations and compared as such, see simulation examples in Figs. 5 and 6.

Obviously, the main role of closures is to reduce the number of equations and this can be successfully achieved for small or other networks with simple structure, such as the line and

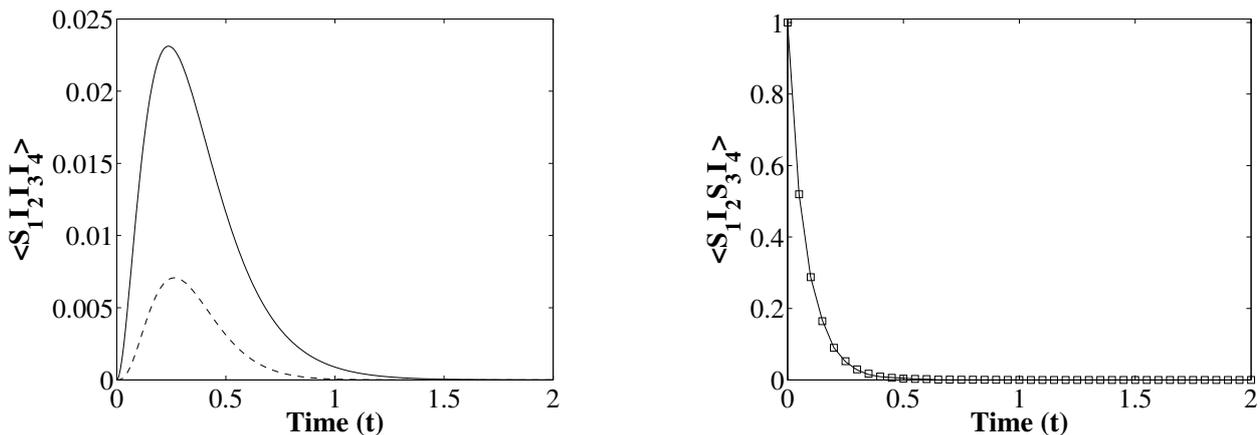


Figure 4: *Left panel:* Testing of the closure  $\langle S_1 I_2 I_3 I_4 \rangle = \frac{\langle S_1 I_2 I_3 \rangle \langle S_1 I_4 \rangle}{\langle S_1 \rangle}$  for the toast network (see Fig. 1d) by plotting  $\langle S_1 I_2 I_3 I_4 \rangle$  (continuous line) and  $\frac{\langle S_1 I_2 I_3 \rangle \langle S_1 I_4 \rangle}{\langle S_1 \rangle}$  (dashed line) from the full system given in Appendix 8. The disagreement highlights that this particular closure does not hold. *Right panel:* Testing of the closure  $\langle S_1 I_2 S_3 I_4 \rangle = \frac{\langle S_1 I_2 S_3 \rangle \langle S_1 S_3 I_4 \rangle}{\langle S_1 S_3 \rangle}$  for the same network with results based on the full system. The probability of the quadruple and the right hand side of the closure are represented by the continuous line and ( $\square$ ), respectively. This agreement shows that particular closures may still hold. System started such that  $\langle S_1 I_2 S_3 I_4 \rangle(0) = 1$  (left panel) and  $\langle S_1 I_2 S_3 I_4 \rangle(0) = 1$  (right panel). Parameter values for all cases are  $\tau = 10/4$  and  $\gamma = 1$ .

star networks. The reduction in equation numbers is illustrated in Table 1. This is by no means an exhaustive list, but simply highlights the potential benefits of good closures. While the more theoretical approach of using full system equations provides a platform to test the validity of our intuition, it is not practical for networks of realistic size. Clearly, this avenue is only useful for simple, toy examples. However, numerical simulation provides an alternative, and for the case of larger networks we can test potential closures without the need of writing down a large set of self-consistent equations. For example, based on the intuition gained so far, we test the validity of some plausible closures on the bow tie network (see Fig. 1e). Namely we consider the following closures:

$$\langle I_1 S_2 S_3 I_4 \rangle = \frac{\langle I_1 S_2 S_3 \rangle \langle S_3 I_4 \rangle}{\langle S_3 \rangle} \text{ and } \langle S_2 S_3 I_5 \rangle = \frac{\langle S_2 S_3 \rangle \langle S_3 I_5 \rangle}{\langle S_3 \rangle}. \quad (5)$$

Testing their validity is simply a matter of numerically evaluating the probability of parts of the network being in particular states at given times. This amounts to recording the presence or otherwise of given state configurations across a fixed set of nodes. Averaging over sufficient simulations, provides an excellent approximation of the desired probability of observing a given state configuration across a given part of the network. Figure 5 shows clearly that the candidate closures are likely to be exact.

The simple analysis thus far, suggests that loops cannot be closed by breaking them down to their component parts. However, the alternative, where the loops are kept closed can be considered.

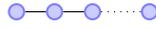
	<b>NFS</b>	<b>RS</b>
	13	10
	23	12
	$(3N^2 - N + 2)/2$	$5N - 3$
	27	17
	35	26

Table 1: Reduction in the number of equations due to closures in a number of networks.  $N$  stands for the number of nodes in a line network.

### 3 Main result for networks with loops

#### 3.1 Network structure driven closures

The analysis in this paper reveals an important relation between the structure of the network on which the epidemic is modelled and the type of closures that are feasible. Moreover, the structural properties discussed below will also serve as a good indicator of the feasibility of writing down exact equations for a given network. The two important structural properties are [3]:

**Definition 1.** Let  $G = \{V, E\}$  be a connected network. Let  $v$  be a vertex of  $G$ ,  $v \in G(V)$ . A node  $v$  is called a **cut-vertex**, iff  $G \setminus \{v\}$  is disconnected.

For our purposes we are interested in cut-vertices, i.e single nodes whose removal leads to disconnected components or subnetworks. The second edge property that is of interest and related to the nodal property is:

**Definition 2.** Let  $G = \{V, E\}$  be a connected graph. An edge  $e \in G(E)$  is called a **bridge** iff its removal increases the number of connected components. It follows that an edge is a bridge iff it is not contained in any cycle, and that the end nodes of a **bridge** are **cut-vertices**.

Examples of cut-vertices are provided in Fig. 1, namely nodes  $\{1\}$ ,  $\{3\}$ ,  $\{\{3\}, \{4\}\}$  and  $\{1\}$  are cut-vertices in the lollipop, bow tie, bow tie with a bridge and start triangle networks, respectively. Similarly, edge  $(3, 4)$  is a bridge in the bow tie with a bridge network. Further examples are provided in Fig. 8. We also note that for all our simple examples closures worked around cut-vertices, see the middle node in an open triangle and the degree 3 node in the lollipop.

#### 3.2 Main result

Based on the intuition gained from the closures on simple networks and their link to the structural properties of the network, via cut-vertices, we can state our main result that generalises the closures in Sharkey et. al. [14], and formalises the link between closures and the structural properties of the network. This leads to the following theorem:

**Theorem 1.** Let  $G = \{V, E\}$  be a network with  $N$  vertices ( $V = \{1, 2, \dots, N\}$ ) and a set of edges given by  $E$ . Consider a connected subset of vertices  $F = \{v_1, v_2, \dots, v_k\} \subseteq V$ , and assume that  $\exists v_{i^*} \in F$ , a cut-vertex in  $G$ , such that  $F \setminus \{v_{i^*}\}$  is partitioned into at least two disjointed

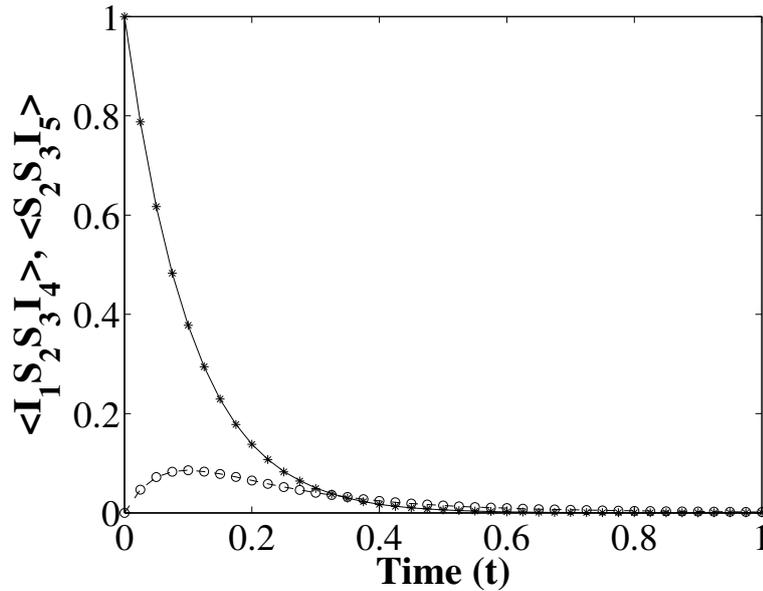


Figure 5: Testing of two closures on the bow tie network (see Fig. 1e),  $\langle I_1 S_2 S_3 I_4 \rangle = \frac{\langle I_1 S_2 S_3 \rangle \langle S_3 I_4 \rangle}{\langle S_3 \rangle}$  and  $\langle S_2 S_3 I_5 \rangle = \frac{\langle S_2 S_3 \rangle \langle S_3 I_5 \rangle}{\langle S_3 \rangle}$ , by plotting the average of  $10^5$  Gillespie-type simulations of the quadruple and the triple with continuous and dashed lines, respectively. The corresponding right hand sides of the closures are plotted with (\*) and (o), respectively. System started such that  $\langle I_1 S_2 S_3 I_4 S_5 \rangle(0) = 1$ , and hence,  $\langle S_2 S_3 I_5 \rangle(0) = 0$ . Parameter values for all cases are  $\tau = 10/4$  and  $\gamma = 1$ . The excellent agreement between the left and right hand sides of the identities shows that appropriately chosen closures are likely to hold. This plot is based exclusively on simulation results due to the complexity of writing down the full system of equations for the bow tie network.

components with vertices  $F_1 = \{v_1, v_2, \dots, v_{i-1}\}$  and  $F_2 = \{v_{i+1}, v_{i+2}, \dots, v_k\}$  belonging to any such two, distinct and disjointed components or subnetworks. Then the following equation holds:

$$\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle(t) = \frac{\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} \rangle(t) \langle S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle(t)}{\langle S_{v_{i^*}} \rangle(t)}, \quad (6)$$

where  $Z_{v_i} = S$  or  $I$  for  $\forall v_i \neq v_{i^*}$ , and  $\langle \cdot \rangle$  denotes the probability of a given subgraph being in a given state at a given time. We note that when  $\langle S_{v_{i^*}} \rangle = 0$ , the left hand side of Eq. (6) is equal to zero and the closure holds trivially.

*Proof.* We begin by noting that:

- i.  $F$  could be the entire vertex set or a strict subset of it, and
- ii. By assumption, the removal of  $v_{i^*}$  means that for  $\forall v_a \in \{F_1\}$  and  $\forall v_b \in \{F_2\}$  there are no paths in  $G$  that connect  $v_a$  and  $v_b$ .

By definition of conditional probabilities,

$$\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle = \langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} | S_{v_{i^*}} \rangle \langle S_{v_{i^*}} \rangle, \quad (7)$$

where the conditional probability can be written as

$$\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_N} | S_{v_{i^*}} \rangle = \frac{\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} | S_{v_{i^*}} \rangle}{\langle S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_N} | S_{v_{i^*}} \rangle}. \quad (8)$$

The equality above holds due to the two subgraphs spanned by  $F_1$  and  $F_2$  being disjointed with no other links, except via  $v_{i^*}$ . Given that the cut-vertex is susceptible, it means that transmission via this route has not occurred, and thus the projection of the system state on the subgraphs spanned by  $F_1$  and  $F_2$  must be independent. Combining Eqs. (7-8) and using that

$$\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} | S_{v_{i^*}} \rangle = \langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} \rangle / \langle S_{v_{i^*}} \rangle$$

and

$$\langle S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} | S_{v_{i^*}} \rangle = \langle S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle / \langle S_{v_{i^*}} \rangle$$

gives,

$$\frac{\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle}{\langle S_{v_{i^*}} \rangle} = \frac{\langle Z_{v_1} Z_{v_2} \cdots Z_{v_{i-1}} S_{v_{i^*}} \rangle \langle S_{v_{i^*}} Z_{v_{i+1}} Z_{v_{i+2}} \cdots Z_{v_k} \rangle}{\langle S_{v_{i^*}} \rangle^2}, \quad (9)$$

which is equivalent to the general closure specified in Eq. (6).  $\square$

It is straightforward to see that all our simple intuitive closures that were exact are special cases of this main result.

### Special case for tree-like networks

In a tree-like network all nodes (except those with degree one) are cut-vertices, and noting that the equations for triples in Eqs. (1) are all such that the middle node is susceptible, it follows that

$$\langle S_i S_j I_k \rangle = \frac{\langle S_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle} \quad \text{and} \quad \langle I_i S_j I_k \rangle = \frac{\langle I_i S_j \rangle \langle S_j I_k \rangle}{\langle S_j \rangle}. \quad (10)$$

## 4 Applications of the main result

Below, we give an example for a specific network, the star-triangle network Fig. 1g, where we write down and program in the exact system and compare the results based on the ODEs to results from Gillespie-type simulation. This is complemented by giving an upper bound on the number of equations needed for an exact representation for networks with non-overlapping loops of at most size 3, which means that overlap via edges is not allowed. More importantly, we generalise the applicability of the reduction by closure technique for arbitrary networks, and also provide an upper bound on the number of equations needed for an exact system.

### 4.1 Star-triangle network

Here we give details of the usefulness of the general result on the link between cut-vertices and closures. To do this we consider the model example of a star-triangle network, as given in Fig. 1g. Suppose that the triangles are labeled 1 to  $M$ , the central node is labeled 1, and the exterior nodes (i.e. the  $i^{\text{th}}$  triangle except the central node) in the  $i^{\text{th}}$  triangle are  $t_i^1, t_i^2$ . For this setup the reduced, exact system can be written down as shown in Appendix 9. By considering the reduced system, it is straightforward to see that the number of equations depends on the number of triangles. This dependency can be quantified by a multiplicative factor which gives the number of necessary equations for individual sub-triangles. However, we note that there is a significant difference in equation numbers when considering an isolated or single subnetwork by itself or as part of a bigger network. Namely, in this case, due to node 1 being a cut-vertex, the network will break down into  $M$  disjointed triangles, upon its removal. For the purpose of generating the reduced system at the network level, these disjointed triangles should not be considered as isolated triangles, but rather as being part of the whole network. An isolated triangle needs 18 equations (i.e. 6 equations for the nodes as each node can be  $S$  or  $I$ , 6 equations for the edges as each edge can be  $SI$  or  $IS$ , and 6 triples since out of the eight possible configurations with nodes being  $S$  or  $I$ , the  $SSS$  and  $III$  triples are dynamically unimportant, see Appendix 6). However, when a triangle is considered as part of a whole network, the 18 equations need to be extended to include differential equations for  $SS$ -type edges, where at least one of the end nodes is a cut-vertex. This in fact accounts for infection from outside the triangle. Similarly, the equations for triples need to account for the  $SSS$  triple in order to capture infection coming via a cut-vertex from outside. This extension procedure, for the current setup, requires 3 extra equations, two for the edges and one for the triples. Obviously, an equation for  $\langle S_{i_1} S_{i_2} \rangle$  is not needed, as this pair cannot become infected from outside. Hence, to summarise, the reduced system seems to need  $21 \cdot M$  equations. However, the cut-vertex (i.e. node 1), has a multiplicity  $M$ , and therefore,  $\langle S_1 \rangle$  and  $\langle I_1 \rangle$  appear  $M$  times even though these are only needed once. Thus, the final number of equations in the reduced system is given by  $21 \cdot M - 2 \cdot (M - 1) = 19 \cdot M + 2$ .

Generating these  $19 \cdot M + 2$  equations in a systematic way leads to an exact representation of the  $SIR$  dynamics on the star triangle network. Figure 6 show results from comparing the numerical solutions of the resulting system of ODEs to simulation results. The plot shows excellent agreement and supports the main result, namely that closures are exact, and hence, the ODE representation is exact. The same figure shows that distributing the same amount of initially infected nodes differently leads to different dynamics, and as expected, distributing the index cases in a more random, or less regular way, leads to a larger epidemic in this very structured network.

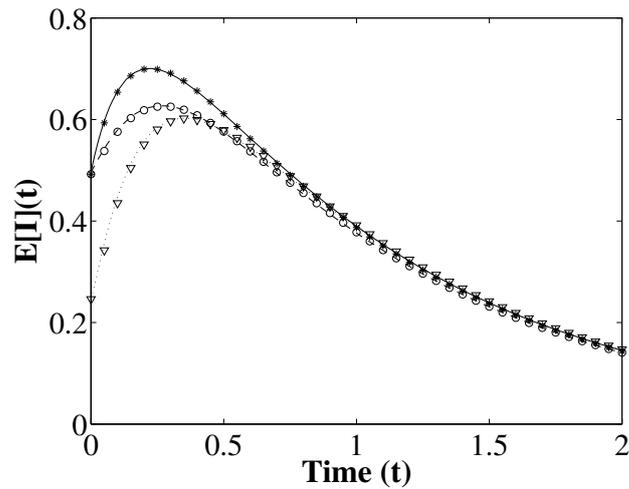


Figure 6: Testing of the exact system and closures for the star network with 64 triangles (see Fig. 1g) by plotting the prevalence of infection based on the deterministic closed system (lines), as well as the average of  $10^4$  Gillespie-type simulations (markers). System started with the central node susceptible, and with (a) one  $S$  and one  $I$  node in each triangle (continuous line and  $\star$ ), (b) two  $S$  nodes in half of the triangles and two  $I$  nodes in the rest (dashed line and  $\circ$ ), and (c) half of the triangles completely susceptible while the other half as in (a) (dotted line and  $\nabla$ ). This amounts to the same number of initially infected nodes, but distributed differently, for (a) and (b), and half as many infected nodes for (c). Parameter values for all cases are  $\tau = 3$  and  $\gamma = 1$ . The agreement between the exact model and simulations confirms the correctness of the exact model which can be difficult and error prone to derive for more complex networks.

## 4.2 General networks with loops of maximum size 3

The star-triangle discussed in the previous section is part of the family of networks with loops no larger than 3. Despite its very specific topology, the result with respect to the number of equations needed to derive an exact system can be extended to provide a much more general statement for a wider class of networks. More specifically, if a network has loops no larger than size 3 (which implies that triangles cannot have overlapping edges), see Figs. 1g and 7, we can give an upper bound on the size of the system of equations describing the system dynamics.

**Theorem 2.** *Consider a network with  $N$  nodes,  $E$  edges,  $T$  triangles and no larger loops. The number of equations needed to fully describe the system dynamics is less than  $2N + 3E + 7T \leq 10N$ .*

*Proof.* The exact reduced system can be built starting from system (1), where not all equations are listed. The equations for singles and pairs need to be augmented by equations for triples, quadruples and up to full system size. We will show that equations for most of the variables describing higher-order network structure are not needed, because these can be expressed as a nonlinear combinations of other variables describing lower-order network structure by appropriate closures. More specifically, it will be shown that differential equations are needed only for the states of nodes, edges and triangles. The number of equations required for these is determined below.

The equations for the probability of each node being either infected or susceptible are obviously needed. This results in  $2N$  equations that contain certain edge state probabilities. In system (1) there are three equations for each edge, since an edge can be in the states  $SI$ ,  $IS$  and  $SS$ . We note that the probability that an edge is in state  $II$  does not occur in system (1). This results in  $3E$  differential equations for the edges.

The differential equations for the edge-probabilities involve triples in system (1). These triples can be either open paths of length 3 or triangles. In the case of a path of length 3, the middle node is always susceptible, this can be verified as follows. If the  $(i, j)$  edge is in state  $S_i I_j$ , and node  $k$  can infect node  $i$  (provided that it is connected to it), then the equation for  $\langle S_i I_j \rangle$  will depend on  $\langle I_k S_i I_j \rangle$ . Similarly, the previously susceptible node  $j$  can get infected by a node  $l$  (provided that it is connected to it). Thus, the term  $\langle S_i S_j I_l \rangle$  also appears in the same equation. Since there are no loops longer than three, the middle node in a such a triple is a cut vertex, hence our main result guarantees that triples have exact closures,

$$\langle X_i S_j Y_k \rangle = \frac{\langle X_i S_j \rangle \langle S_j Y_k \rangle}{\langle S_j \rangle}.$$

Hence, equations for such triples are not needed as they can be expressed in terms of already existing singles and edges. Triples can also be triangles. In this case a closure is not possible, see previous sections and Appendix 6. Let us now determine the number of differential equations needed for triangles. Every triangle can be in  $2^3 = 8$  different states. However, in the differential equations of pairs not all of these states occur (see Appendix 6). Namely, take an edge of state  $S_i I_j$  in a triangle. Similarly to the case of paths of length three, the equation for  $\langle S_i I_j \rangle$  contains  $\langle I_k S_i I_j \rangle$  and  $\langle S_i S_j I_k \rangle$ , where  $k$  is the third node of the triangle. Every type of triple containing exactly one infected node or exactly one susceptible node appears this way, because nodes  $i, j$  and  $k$  are arbitrary within the triangle. Thus the 6 states require six equations (the equations for the states  $SSS$  and  $III$  are not needed). However, we will show below that an equation for the  $SSS$  state will be needed, hence for a triangle we need 7 equations. Thus up to now the number of differential equations in the reduced system is  $2N + 3E + 7T$ .

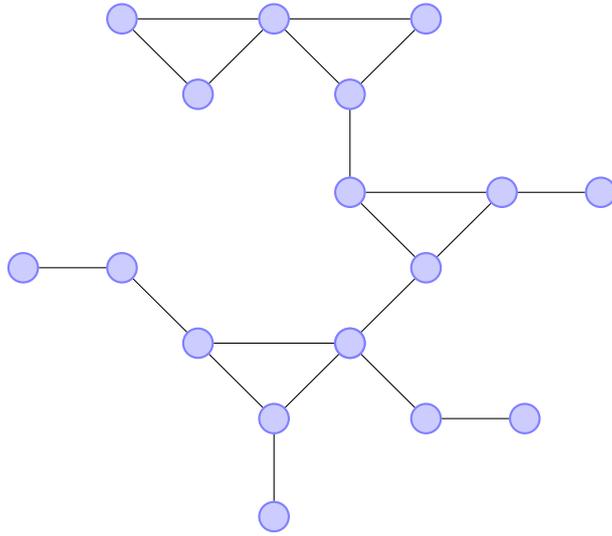


Figure 7: An example of a network with maximum loop size 3. Overlap of triangles via edges is not allowed, but triangles can overlap via single nodes.

Next we will show that no other variables, and hence no equations, are needed because there are no loops longer than three. In the equations for triples the probability of quadruples appear. Because of the structure of the graph these could be paths of length 4 or lollipops. But a path of length 4 cannot appear in the equation of a triple, since all of our triples that cannot be closed are triangles. A lollipop can appear in the equation of a path of length 3 or a triangle. In either case, the natural cut vertex of the lollipop is susceptible, because if it was susceptible in the triple it stays that way, and if it was infected, it is susceptible in the lollipop and the fourth, new node is infected. So by our main result these subgraphs have closures. This proves that no quadruples and higher order structures appear in the reduced exact system. We note that for these closures we need  $\langle S_i S_j S_k \rangle$  for all  $i \neq j \neq k$  making a triangle, since in every triple we have an  $S_i S_j I_k$  state, where  $k$  becomes the susceptible cut vertex for a lollipop (and the stem-node, or the node with three connections, is infected).

Finally, the number of equations is estimated by the size of the graph. Much like trees, networks with loops of maximum size 3 are relatively sparse. The spanning tree of such networks has  $N - 1$  edges. There are some triangles in the network, and these triangles cannot share an edge, otherwise there would be loops of size greater than 3. So every edge is in exactly one triangle. In this case,  $E = N - 1 + T$ , since two edges of the triangle were already counted in the spanning tree.  $T \leq \frac{N}{2}$  since to construct such a network one always has to add at least two extra nodes to get a new triangle. Taking into account all of the above yields

$$2N + 3E + 7T = 2N + 3(N - 1 + T) + 7T \leq 2N + 3N + \frac{3}{2}N + \frac{7}{2}N = 10N.$$

□

Consider for example the lollipop network. In this case,  $2N + 3E + 7T = 2 \cdot 4 + 3 \cdot 4 + 7 \cdot 1 = 27$ . In Appendix 7 we have 26 equations describing the dynamics, since we do not need  $\langle S_3 S_4 \rangle$  for any of the closures. For the star-triangle network, see sections above and Appendix 9, of  $M$  triangles one needs a maximum of  $2 \cdot (1 + 2M) + 3 \cdot 3 \cdot M + 7 \cdot M = 2 + 20 \cdot M$  equations. Actually, the exact number is smaller since we do not need  $\langle S_{i_1} S_{i_2} \rangle$  in any of the triangles. So the number of equations is  $2 \cdot (1 + 2 \cdot M) + 3 \cdot 3 \cdot M + 7 \cdot M - M = 2 + 19 \cdot M$ .

### 4.3 Feasibility of the reduction by closure technique for general networks

Here, we provide a recipe-like approach to establish the feasibility of writing down an exact representation for a given network. To achieve this for a given network  $G = \{V, E\}$ , the following steps should be taken:

1. Find all cut-vertices of  $G$  by using the *depth-first search* algorithm [15], and denote these by  $C = \{v_{i_1}, v_{i_2}, \dots, v_{i_L}\} \subset V$ . This algorithm runs in polynomial time in  $(|E| + |V|)$ .
2. Splice the original network into independent subnetworks (each subnetwork is connected, but any two are disconnected) as determined by the number and properties of cut-vertices. Let us assume that this procedure leads to a family of distinct subnetworks denoted by  $G_1, G_2, \dots, G_P$ , where  $G_i = \{V_i, E_i\}$  with  $i = 1, 2, \dots, P$ , and each of these with frequency or counts given by  $f_1, f_2, \dots, f_P$ , respectively. This can be done in a way in which the cut-vertices are maintained in all subnetworks that they generate, see Fig. 8. Let us denote by  $Ind(v_{i_j})$ , where  $j = 1, 2, \dots, L$ , the number of subnetworks that cut-vertex  $v_{i_j}$  belongs to. *The subnetwork is in fact a generalisation of the loop concept in that it needs to be connected and with no further cut-vertices. As indicated by our results, closures within loops or subnetworks will not be exact.*
3. The relation between the distinct subnetworks  $P$ , their frequency, and the number of nodes in the subnetworks (e.g.  $|V_1|, |V_2|, \dots, |V_P|$ ) will determine the number of equations needed for a full, exact representation. This relation is made more precise by the corresponding multipliers  $m_1, m_2, \dots, m_P$  which simply denote the number of equations needed to describe exactly the corresponding subnetworks, e.g. an edge needs 7 equations (4 equations for the nodes and 3 equations at pair level). In a similar way a triangle needs 22 equations (6 equations for the nodes, 9 equations for the edges and 7 equations for the triangles), a cycle graph with 4 nodes needs 45 equations, and the toast network needs 57 equations. Hence an upper bound for the number of equations needed to describe the epidemic dynamics exactly is given by

$$N_{EQ}(G) = \sum_{i=1}^P m_i f_i - 2 \sum_{j=1}^L (Ind(v_{i_j}) - 1).$$

The formula simply takes a sum across the number of equations needed for all subnetworks and adjusts this to account for the unnecessary multiplications caused by cut-vertices being part of multiple subnetworks. Applying this procedure for the simplest cases of tree-like networks gives  $N_{EQ}(G) = 2 \cdot |V| + 3 \cdot |E|$ . Moreover, for tree-like networks with triangles only, the removal of all cut-vertices will leave subnetworks of two distinct types, namely the edge and triangle, yielding  $N_{EQ}(G) = 2 \cdot |V| + 3 \cdot |E| + 7 \cdot T$  equations, where as before  $T$  is the number of triangles in the network. The reason for  $N_{EQ}(G)$  being an upper bound is due to accounting for all  $SS$  pairs regardless of whether all of these are needed to link to other subnetworks. Similarly, other fully susceptible arrangements at higher level may be needed to account appropriately for outside infections (see the detailed explanation in the *Star-triangle network* section). While the formula could be further improved, the exact overestimate depends in a non-trivial way on the interaction between the structure of the network and epidemic dynamics. Our investigations show that removing the unnecessary  $SS$  variables will not considerably decrease the number of equations.

	<b>Isolated</b>	<b>Part of a network</b>
	6	7
	18	22
	36	45
	47	57
	58	69

Table 2: The number of equations for a class of subnetworks that are isolated *versus* being part of a network.

We note that for the decomposition of a network there are two extreme scenarios: (a) the network has many cut-vertices and a large number of subnetworks of a few different types and (b) the network decomposes to relatively few distinct subnetwork, but of large sizes. The tree-like networks are a good example for scenario (a), where the only subnetwork is the edge, where each edge requires only 3 equations. Thus few equations per subnetwork but many subnetworks. More structured networks will typically have distinct subnetworks of larger sizes which for an exact description will require a larger number of equations. Thus fewer subnetworks, but many more equations per subnetwork. More importantly, it is non-trivial to find a simple relation between subnetworks and the number of equations needed for an exact description, and this may require further attention and work. It is straightforward to see that the desirable scenario for an exact representation is scenario (a), and it is likely that in this case an exact description is possible. Complexity quickly increases from 22 equations needed for a closed triangle to 57 equations for a subnetwork equivalent to a toast network, see Table 2 that gives equation numbers for these and further examples of small networks. Thus both scenarios require a large number of equations. Generating and implementing the equations needed for an exact description is prone to error and we highly recommend the development of an algorithmic approach, where equations can be generated automatically rather than manually. While in the present project we adopted a manual, direct approach, future work will consider the implementation of an automated procedure for as general a situation as possible. The description above, illustrates clearly that the family of networks with many cut-vertices are more amenable to this approach, and it is likely that for networks with few cut-vertices, the task of writing down an exact system may be out of reach.

## 5 Discussion

In this paper we extended results for tree-like networks and Markovian *SIR* epidemics [14] to networks with loops, and provided an important link between the structural properties of the network and the feasibility of writing down an exact representation of the epidemic on the network. The results are built up in a methodical way starting from the simplest networks or network motifs, and to enhance clarity we give the full system of equations whenever this is possible. The proof of the main results is in fact an alternative to and more general than the proof provided in [14], and its usefulness and generality in reducing the number of equations in the exact system is illustrated by our concrete, worked out examples for tree-like networks that contain non-overlapping (via edges) loops of maximum size 3. We note that one could



consider the idea of extending the cut-vertex concept to cut-sets. For example, for the toast network, the removal of the diagonal nodes results in the network breaking into two disconnected components. However, this will unfortunately lead to a negligible reduction in the number of equations, since closures based on this cut-set will only hold for the situation when the diagonal edge is in the  $SS$  state. However, there are relatively few states for triples, quadruples and higher order structures with the diagonal being in the  $SS$  state, and the derivation of the natural system will require many more equations for states with the diagonal not being fully susceptible. Hence, closures will only exist for a limited number of states, and this frustrates the effort of obtaining an exact system with fewer equations. In future work, we will concentrate on exploring the previous idea further and on improving the upper estimate on the number of equations in a reduced, exact system, as well as making the upper estimate more explicit.

As the size of subnetworks increases so does the number of equations and at a much faster rate. However, for cycle graphs the exact system contains relatively few equations which we now briefly explain. In the case of a cycle graph with  $N$  nodes,  $2N$  equations for the nodes (each node can be  $S$  or  $I$ ) and  $2N$  equations for the edges (each edge can be in one of the two states  $SI$  or  $IS$ ) are needed. In the differential equations for the edges two types of triples occur, namely  $SSI$  (there are  $2N$  of these) and  $ISI$  triples (there are  $N$  of these), hence there are a total of  $3N$  equations for the triples. Similarly, there are  $3N$   $k$ -motifs (for  $k = 4, 5, \dots, N-1$ ), namely those of type  $S \dots SI$  and those of type  $IS \dots SI$ . Finally, there are a total of  $2N$   $N$ -motifs,  $N$  of them are of type  $S \dots SI$  and  $N$  of them are of type  $IS \dots SI$ . Therefore, the number of equations in the full system of a cycle graph with  $N$  nodes is  $2N + 2N + 3N(N-3) + 2N = 3N(N-1)$ . It is worth noting that this system can be lumped to  $2N-1$  equations by introducing a single variable for each motif type. Namely, the first lumped variable will be  $\langle S \rangle = \sum_{i=1}^N \langle S_i \rangle$  the expected number of  $\langle S \rangle$  nodes. Similarly, the second lumped variable is  $\langle I \rangle = \sum_{i=1}^N \langle I_i \rangle$  the expected number of  $I$  nodes. Then in a similar way there will be a lumped variable for  $k$ -motifs of type  $S \dots SI$  and another lumped variable for  $k$ -motifs of type  $IS \dots SI$ . In total the number of lumped variables will sum up to  $2N-1$ , meaning that the exact value of the prevalence can be given by solving a system of  $2N-1$  ODEs. Hence, it is feasible to extend our results to tree-like networks with no-overlapping loops of size greater than three.

While progress in modelling epidemic dynamics on networks with loops (which usually involves clustering) has been made [1, 4, 11, 17, 20], many challenges remain. These challenges are both around generating clustered networks, and tuning the amount of clustering and implicitly the number and type of different loops [1, 8, 10, 17, 19], and especially around providing a description of the time-evolution of the epidemic. Progress in determining the final epidemic size (time-evolution not needed) has been good but models describing the time-evolution are more challenging. In this paper, we make the first steps in providing a well-grounded and rigorous modelling alternative. Although the models presented are unlikely to pertain to analytical analysis, they could provide a valuable platform to investigate the effect of intervention or control on nodes, edges or subparts of the network. The exact system, in this case, could give precise information about the impact of isolating nodes, links or decreasing their potential of transmitting, or increasing the recovery rate of some targeted or specific nodes.

## Acknowledgements

Péter L. Simon acknowledges support from OTKA (grant no. 81403). Robert R. Wilkinson acknowledges support from EPSRC (DTA studentship).

## 6 Appendix: The closed triangle network

The triangle network is a loop of three nodes numbered 1, 2 and 3, see Fig. 1b. The system dynamics are given by the following set of equations:

$$\langle \dot{I}_1 \rangle = \tau \langle S_1 I_2 \rangle + \tau \langle S_1 I_3 \rangle - \gamma \langle I_1 \rangle, \quad (11)$$

$$\langle \dot{S}_1 \rangle = -\tau \langle S_1 I_2 \rangle - \tau \langle S_1 I_3 \rangle, \quad (12)$$

$$\langle \dot{I}_2 \rangle = \tau \langle I_1 S_2 \rangle + \tau \langle S_2 I_3 \rangle - \gamma \langle I_2 \rangle, \quad (13)$$

$$\langle \dot{S}_2 \rangle = -\tau \langle I_1 S_2 \rangle - \tau \langle S_2 I_3 \rangle, \quad (14)$$

$$\langle \dot{I}_3 \rangle = \tau \langle I_2 S_3 \rangle + \tau \langle I_1 S_3 \rangle - \gamma \langle I_3 \rangle, \quad (15)$$

$$\langle \dot{S}_3 \rangle = -\tau \langle I_2 S_3 \rangle - \tau \langle I_1 S_3 \rangle, \quad (16)$$

$$\langle I_1 \dot{S}_2 \rangle = -(\tau + \gamma) \langle I_1 S_2 \rangle - \tau \langle I_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 I_3 \rangle, \quad (17)$$

$$\langle \dot{S}_1 I_2 \rangle = -(\tau + \gamma) \langle S_1 I_2 \rangle + \tau \langle S_1 S_2 I_3 \rangle - \tau \langle S_1 I_2 I_3 \rangle, \quad (18)$$

$$\langle I_2 \dot{S}_3 \rangle = -(\tau + \gamma) \langle I_2 S_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle - \tau \langle I_1 I_2 S_3 \rangle, \quad (19)$$

$$\langle \dot{S}_2 I_3 \rangle = -(\tau + \gamma) \langle S_2 I_3 \rangle - \tau \langle I_1 S_2 I_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle, \quad (20)$$

$$\langle I_1 \dot{S}_3 \rangle = -(\tau + \gamma) \langle I_1 S_3 \rangle - \tau \langle I_1 I_2 S_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle, \quad (21)$$

$$\langle \dot{S}_1 I_3 \rangle = -(\tau + \gamma) \langle S_1 I_3 \rangle - \tau \langle S_1 I_2 I_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle, \quad (22)$$

$$\langle I_1 \dot{S}_2 I_3 \rangle = -2(\tau + \gamma) \langle I_1 S_2 I_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle + \tau \langle S_1 S_2 I_3 \rangle, \quad (23)$$

$$\langle \dot{S}_1 I_2 I_3 \rangle = -2(\tau + \gamma) \langle S_1 I_2 I_3 \rangle + \tau \langle S_1 S_2 I_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle, \quad (24)$$

$$\langle I_1 \dot{I}_2 S_3 \rangle = -2(\tau + \gamma) \langle I_1 I_2 S_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle, \quad (25)$$

$$\langle \dot{S}_1 I_2 S_3 \rangle = -(2\tau + \gamma) \langle S_1 I_2 S_3 \rangle, \quad (26)$$

$$\langle \dot{S}_1 \dot{S}_2 I_3 \rangle = -(2\tau + \gamma) \langle S_1 S_2 I_3 \rangle, \quad (27)$$

$$\langle I_1 \dot{S}_2 S_3 \rangle = -(2\tau + \gamma) \langle I_1 S_2 S_3 \rangle. \quad (28)$$

Since this network has no cut-vertices, closures are not possible for any of the subsystems above. For instance,

$$\langle S_1 S_2 I_3 \rangle = \frac{\langle I_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle}, \quad \langle I_1 S_2 I_3 \rangle = \frac{\langle I_1 S_2 \rangle \langle S_2 I_3 \rangle}{\langle S_2 \rangle},$$

are closures that do not hold, see Fig. 2. We note that the evaluation of one of the closures above (the first) requires an extra equation for  $\langle S_1 S_2 \rangle$ . This is given by,

$$\langle \dot{S}_1 \dot{S}_2 \rangle = -2\tau \langle S_1 S_2 I_3 \rangle.$$

Depending on the closures that we wish to test, additional equations may be needed.

## 7 Appendix: Equations for the lollipop network

The lollipop network we consider has nodes numbered as shown in Fig. 1c. The equations describing the *SIR* model can be formulated as follows:

$$\langle \dot{I}_1 \rangle = \tau \langle S_1 I_2 \rangle + \tau \langle S_1 I_3 \rangle + \tau \langle S_1 I_4 \rangle - \gamma \langle I_1 \rangle, \quad (29)$$

$$\langle \dot{S}_1 \rangle = -\tau \langle S_1 I_2 \rangle - \tau \langle S_1 I_3 \rangle - \tau \langle S_1 I_4 \rangle, \quad (30)$$

$$\langle \dot{I}_2 \rangle = \tau \langle I_1 S_2 \rangle - \gamma \langle I_2 \rangle, \quad (31)$$

$$\langle \dot{S}_2 \rangle = -\tau \langle I_1 S_2 \rangle, \quad (32)$$

$$\langle \dot{I}_3 \rangle = \tau \langle I_1 S_3 \rangle + \tau \langle S_3 I_4 \rangle - \gamma \langle I_3 \rangle, \quad (33)$$

$$\langle \dot{S}_3 \rangle = -\tau \langle I_1 S_3 \rangle - \tau \langle S_3 I_4 \rangle, \quad (34)$$

$$\langle \dot{I}_4 \rangle = \tau \langle I_1 S_4 \rangle + \tau \langle I_3 S_4 \rangle - \gamma \langle I_4 \rangle, \quad (35)$$

$$\langle \dot{S}_4 \rangle = -\tau \langle I_1 S_4 \rangle - \tau \langle I_3 S_4 \rangle, \quad (36)$$

$$\langle I_1 \dot{S}_2 \rangle = -(\tau + \gamma) \langle I_1 S_2 \rangle + \tau \langle S_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 I_4 \rangle, \quad (37)$$

$$\langle S_1 \dot{I}_2 \rangle = -(\tau + \gamma) \langle S_1 I_2 \rangle - \tau \langle S_1 I_2 I_3 \rangle - \tau \langle S_1 I_2 I_4 \rangle, \quad (38)$$

$$\langle I_1 \dot{S}_3 \rangle = -(\tau + \gamma) \langle I_1 S_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle + \tau \langle S_1 S_3 I_4 \rangle - \tau \langle I_1 S_3 I_4 \rangle, \quad (39)$$

$$\langle S_1 \dot{I}_3 \rangle = -(\tau + \gamma) \langle S_1 I_3 \rangle - \tau \langle S_1 I_2 I_3 \rangle - \tau \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle, \quad (40)$$

$$\langle I_1 \dot{S}_4 \rangle = -(\tau + \gamma) \langle I_1 S_4 \rangle + \tau \langle S_1 I_2 S_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle - \tau \langle I_1 I_3 S_4 \rangle, \quad (41)$$

$$\langle S_1 \dot{I}_4 \rangle = -(\tau + \gamma) \langle S_1 I_4 \rangle - \tau \langle S_1 I_2 I_4 \rangle - \tau \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle, \quad (42)$$

$$\langle S_3 \dot{I}_4 \rangle = -(\tau + \gamma) \langle S_3 I_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle - \tau \langle I_1 S_3 I_4 \rangle, \quad (43)$$

$$\langle I_3 \dot{S}_4 \rangle = -(\tau + \gamma) \langle I_3 S_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle - \tau \langle I_1 I_3 S_4 \rangle, \quad (44)$$

$$\langle S_1 \dot{I}_3 I_4 \rangle = -2(\tau + \gamma) \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle, \quad (45)$$

$$\langle S_1 \dot{S}_3 I_4 \rangle = -(2\tau + \gamma) \langle S_1 S_3 I_4 \rangle - \tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (46)$$

$$\langle S_1 \dot{I}_3 S_4 \rangle = -(2\tau + \gamma) \langle S_1 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 S_4 \rangle, \quad (47)$$

$$\langle I_1 \dot{S}_3 I_4 \rangle = -2(\tau + \gamma) \langle I_1 S_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (48)$$

$$\langle I_1 \dot{I}_3 S_4 \rangle = -2(\tau + \gamma) \langle I_1 I_3 S_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle + \tau \langle S_1 I_2 I_3 S_4 \rangle, \quad (49)$$

$$\langle I_1 \dot{S}_3 S_4 \rangle = -(2\tau + \gamma) \langle I_1 S_3 S_4 \rangle + \tau \langle S_1 I_2 S_3 S_4 \rangle. \quad (50)$$

This first group of equations consist of variables (e.g. configurations of states and subgraphs) which cannot be closed or further reduced. Naturally, this first set requires equations at the levels of triples and quadruples or full system size. Note that triples which are part of the triangle cannot be closed. However, the second group of equations, i.e.

$$\langle S_1 \dot{I}_2 I_4 \rangle = -2(\tau + \gamma) \langle S_1 I_2 I_4 \rangle + \tau \langle S_1 I_2 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle, \quad (51)$$

$$\langle S_1 \dot{I}_2 I_3 \rangle = -2(\tau + \gamma) \langle S_1 I_2 I_3 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle, \quad (52)$$

$$\langle S_1 \dot{S}_2 I_3 \rangle = -(\tau + \gamma) \langle S_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle - \tau \langle S_1 S_2 I_3 I_4 \rangle, \quad (53)$$

$$\langle S_1 \dot{S}_2 I_4 \rangle = -(\tau + \gamma) \langle S_1 S_2 I_4 \rangle + \tau \langle S_1 S_2 I_3 S_4 \rangle - \tau \langle S_1 S_2 I_3 I_4 \rangle, \quad (54)$$

$$\langle S_1 \dot{I}_2 S_3 \rangle = -(\tau + \gamma) \langle S_1 I_2 S_3 \rangle - 2\tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (55)$$

$$\langle S_1 \dot{I}_2 S_4 \rangle = -(\tau + \gamma) \langle S_1 I_2 S_4 \rangle - 2\tau \langle S_1 I_2 I_3 S_4 \rangle, \quad (56)$$

$$\langle S_1 I_2 \dot{I}_3 I_4 \rangle = -3(\tau + \gamma) \langle S_1 I_2 I_3 I_4 \rangle + \tau (\langle S_1 I_2 I_3 S_4 \rangle + \langle S_1 I_2 S_3 I_4 \rangle), \quad (57)$$

$$\langle S_1 I_2 \dot{I}_3 S_4 \rangle = -(3\tau + 2\gamma) \langle S_1 I_2 I_3 S_4 \rangle, \quad (58)$$

$$\langle S_1 I_2 \dot{S}_3 I_4 \rangle = -(3\tau + 2\gamma) \langle S_1 I_2 S_3 I_4 \rangle, \quad (59)$$

$$\langle S_1 S_2 \dot{I}_3 I_4 \rangle = -2(\tau + \gamma) \langle S_1 S_2 I_3 I_4 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle + \tau \langle S_1 S_2 I_3 S_4 \rangle, \quad (60)$$

$$\langle S_1 S_2 \dot{I}_3 S_4 \rangle = -(2\tau + \gamma) \langle S_1 S_2 I_3 S_4 \rangle, \quad (61)$$

$$\langle S_1 \dot{S}_2 S_3 I_4 \rangle = -(2\tau + \gamma) \langle S_1 S_2 S_3 I_4 \rangle, \quad (62)$$

$$\langle S_1 I_2 \dot{S}_3 S_4 \rangle = -(\tau + \gamma) \langle S_1 I_2 S_3 S_4 \rangle, \quad (63)$$

can be closed by using the following identities:

$$\langle S_1 I_2 I_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 I_4 \rangle, \quad (64)$$

$$\langle S_1 I_2 I_3 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 I_3 \rangle, \quad (65)$$

$$\langle S_1 S_2 I_3 \rangle \langle S_1 \rangle = \langle S_1 S_2 \rangle \langle S_1 I_3 \rangle, \quad (66)$$

$$\langle S_1 S_2 I_4 \rangle \langle S_1 \rangle = \langle S_1 S_2 \rangle \langle S_1 I_4 \rangle, \quad (67)$$

$$\langle S_1 I_2 S_3 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 S_3 \rangle, \quad (68)$$

$$\langle S_1 I_2 S_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 S_4 \rangle, \quad (69)$$

$$\langle S_1 I_2 I_3 I_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 I_3 I_4 \rangle, \quad (70)$$

$$\langle S_1 I_2 I_3 S_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 I_3 S_4 \rangle, \quad (71)$$

$$\langle S_1 I_2 S_3 I_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 S_3 I_4 \rangle, \quad (72)$$

$$\langle S_1 S_2 I_3 I_4 \rangle \langle S_1 \rangle = \langle S_1 S_2 \rangle \langle S_1 I_3 I_4 \rangle, \quad (73)$$

$$\langle S_1 S_2 I_3 S_4 \rangle \langle S_1 \rangle = \langle S_1 S_2 \rangle \langle S_1 I_3 S_4 \rangle, \quad (74)$$

$$\langle S_1 S_2 S_3 I_4 \rangle \langle S_1 \rangle = \langle S_1 S_2 \rangle \langle S_1 S_3 I_4 \rangle, \quad (75)$$

$$\langle S_1 I_2 S_3 S_4 \rangle \langle S_1 \rangle = \langle S_1 I_2 \rangle \langle S_1 S_3 S_4 \rangle. \quad (76)$$

We note that there are two distinct types of closures. Namely, closures at the level of triples that are not part of the triangle and closures at the full system size. To complete the closed system we need the following extra equations for variables that are required by the closures. These new variables together with their equations are:

$$\langle S_1 \dot{S}_2 \rangle = -\tau \langle S_1 S_2 I_3 \rangle - \tau \langle S_1 S_2 I_4 \rangle, \quad (77)$$

$$\langle S_1 \dot{S}_3 \rangle = -\tau \langle S_1 I_2 S_3 \rangle - 2\tau \langle S_1 S_3 I_4 \rangle, \quad (78)$$

$$\langle S_1 \dot{S}_4 \rangle = -\tau \langle S_1 I_2 S_4 \rangle - 2\tau \langle S_1 I_3 S_4 \rangle, \quad (79)$$

$$\langle S_1 \dot{S}_3 S_4 \rangle = -\tau \langle S_1 I_2 S_3 S_4 \rangle. \quad (80)$$

Substituting the closures given in Eqs. (64-76) into Eqs. (51-63) together with the set of equations that cannot be closed, Eqs. (29-50), and the extra variables induced by the closures, Eqs. (77- 80), will result in a system of 26 differential equations describing the system dynamics completely. Without closures, the system is fully specified by 35 equations. Strictly speaking, we can drop the equations for  $\langle S_i \rangle$  if we are only interested in prevalence and then the equations in the full and reduced system drop to 31 and 23. Note that in the full system all  $\langle S_i \rangle$ s can be dropped, while in the reduced system we cannot drop  $\langle S_1 \rangle$  as the closures rely on it.

## 8 Appendix: Equations for toast network

The evolution equations on the toast network labeled as in Fig. 1d are given by

$$\langle \dot{I}_1 \rangle = \tau \langle S_1 I_2 \rangle + \tau \langle S_1 I_3 \rangle + \tau \langle S_1 I_4 \rangle - \gamma \langle I_1 \rangle, \quad (81)$$

$$\langle \dot{S}_1 \rangle = -\tau \langle S_1 I_2 \rangle - \tau \langle S_1 I_3 \rangle - \tau \langle S_1 I_4 \rangle, \quad (82)$$

$$\langle \dot{I}_2 \rangle = \tau \langle I_1 S_2 \rangle + \tau \langle S_2 I_3 \rangle - \gamma \langle I_2 \rangle, \quad (83)$$

$$\langle \dot{S}_2 \rangle = -\tau \langle I_1 S_2 \rangle - \tau \langle S_2 I_3 \rangle, \quad (84)$$

$$\langle \dot{I}_3 \rangle = \tau \langle I_1 S_3 \rangle + \tau \langle S_3 I_4 \rangle + \tau \langle I_2 S_3 \rangle - \gamma \langle I_3 \rangle, \quad (85)$$

$$\langle \dot{S}_3 \rangle = -\tau \langle I_1 S_3 \rangle - \tau \langle S_3 I_4 \rangle - \tau \langle I_2 S_3 \rangle, \quad (86)$$

$$\langle \dot{I}_4 \rangle = \tau \langle I_1 S_4 \rangle + \tau \langle I_3 S_4 \rangle - \gamma \langle I_4 \rangle, \quad (87)$$

$$\langle \dot{S}_4 \rangle = -\tau \langle I_1 S_4 \rangle - \tau \langle I_3 S_4 \rangle, \quad (88)$$

$$\langle I_1 \dot{S}_2 \rangle = -(\tau + \gamma) \langle I_1 S_2 \rangle + \tau \langle S_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 I_4 \rangle - \tau \langle I_1 S_2 I_3 \rangle, \quad (89)$$

$$\langle S_1 \dot{I}_2 \rangle = -(\tau + \gamma) \langle S_1 I_2 \rangle + \tau \langle S_1 S_2 I_3 \rangle - \tau \langle S_1 I_2 I_3 \rangle - \tau \langle S_1 I_2 I_4 \rangle, \quad (90)$$

$$\langle I_1 \dot{S}_3 \rangle = -(\tau + \gamma) \langle I_1 S_3 \rangle + \tau \langle S_1 I_2 S_3 \rangle + \tau \langle S_1 S_3 I_4 \rangle - \tau \langle I_1 S_3 I_4 \rangle - \tau \langle I_1 I_2 S_3 \rangle, \quad (91)$$

$$\langle S_1 \dot{I}_3 \rangle = -(\tau + \gamma) \langle S_1 I_3 \rangle - \tau \langle S_1 I_2 I_3 \rangle - \tau \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle + \tau \langle S_1 I_2 S_3 \rangle, \quad (92)$$

$$\langle I_1 \dot{S}_4 \rangle = -(\tau + \gamma) \langle I_1 S_4 \rangle + \tau \langle S_1 I_2 S_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle - \tau \langle I_1 I_3 S_4 \rangle, \quad (93)$$

$$\langle S_1 \dot{I}_4 \rangle = -(\tau + \gamma) \langle S_1 I_4 \rangle - \tau \langle S_1 I_2 I_4 \rangle - \tau \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle, \quad (94)$$

$$\langle S_3 \dot{I}_4 \rangle = -(\tau + \gamma) \langle S_3 I_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle - \tau \langle I_1 S_3 I_4 \rangle - \tau \langle I_2 S_3 I_4 \rangle, \quad (95)$$

$$\langle I_3 \dot{S}_4 \rangle = -(\tau + \gamma) \langle I_3 S_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle - \tau \langle I_1 I_3 S_4 \rangle + \tau \langle I_2 S_3 S_4 \rangle, \quad (96)$$

$$\langle S_2 \dot{I}_3 \rangle = -(\tau + \gamma) \langle S_2 I_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle - \tau \langle I_1 S_2 I_3 \rangle + \tau \langle S_2 S_3 I_4 \rangle, \quad (97)$$

$$\langle I_2 \dot{S}_3 \rangle = -(\tau + \gamma) \langle I_2 S_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle - \tau \langle I_1 I_2 S_3 \rangle - \tau \langle I_2 S_3 I_4 \rangle, \quad (98)$$

$$\langle S_1 \dot{I}_2 I_4 \rangle = -2(\tau + \gamma) \langle S_1 I_2 I_4 \rangle + \tau \langle S_1 I_2 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle + \tau \langle S_1 S_2 I_3 I_4 \rangle, \quad (99)$$

$$\langle S_1 \dot{I}_2 I_3 \rangle = -2(\tau + \gamma) \langle S_1 I_2 I_3 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle + \tau \langle S_1 I_2 S_3 \rangle + \tau \langle S_1 S_2 I_3 \rangle, \quad (100)$$

$$\langle S_1 \dot{I}_3 I_4 \rangle = -2(\tau + \gamma) \langle S_1 I_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 I_4 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (101)$$

$$\langle S_1 \dot{S}_2 I_3 \rangle = -(2\tau + \gamma) \langle S_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle - \tau \langle S_1 S_2 I_3 I_4 \rangle, \quad (102)$$

$$\langle S_1 \dot{S}_2 I_4 \rangle = -(\tau + \gamma) \langle S_1 S_2 I_4 \rangle + \tau \langle S_1 S_2 I_3 S_4 \rangle - 2\tau \langle S_1 S_2 I_3 I_4 \rangle, \quad (103)$$

$$\langle S_1 \dot{I}_2 S_3 \rangle = -(2\tau + \gamma) \langle S_1 I_2 S_3 \rangle - 2\tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (104)$$

$$\langle S_1 \dot{S}_3 I_4 \rangle = -(2\tau + \gamma) \langle S_1 S_3 I_4 \rangle - 2\tau \langle S_1 I_2 S_3 I_4 \rangle, \quad (105)$$

$$\langle I_1 \dot{S}_3 I_4 \rangle = -2(\tau + \gamma) \langle I_1 S_3 I_4 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle + \tau \langle S_1 S_3 I_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle - \tau \langle I_1 I_2 S_3 I_4 \rangle, \quad (106)$$

$$\langle S_1 \dot{I}_2 S_4 \rangle = -(\tau + \gamma) \langle S_1 I_2 S_4 \rangle - 2\tau \langle S_1 I_2 I_3 S_4 \rangle + \tau \langle S_1 S_2 I_3 S_4 \rangle, \quad (107)$$

$$\langle S_1 \dot{I}_3 S_4 \rangle = -(2\tau + \gamma) \langle S_1 I_3 S_4 \rangle - \tau \langle S_1 I_2 I_3 S_4 \rangle + \tau \langle S_1 I_2 S_3 S_4 \rangle, \quad (108)$$

$$\langle I_1 \dot{I}_3 S_4 \rangle = -2(\tau + \gamma) \langle I_1 I_3 S_4 \rangle + \tau \langle S_1 I_3 S_4 \rangle + \tau \langle I_1 S_3 S_4 \rangle + \tau \langle S_1 I_2 I_3 S_4 \rangle + \tau \langle I_1 I_2 S_3 S_4 \rangle, \quad (109)$$

$$\langle I_1 \dot{S}_3 S_4 \rangle = -(2\tau + \gamma) \langle I_1 S_3 S_4 \rangle + \tau \langle S_1 I_2 S_3 S_4 \rangle - \tau \langle I_1 I_2 S_3 S_4 \rangle, \quad (110)$$

$$\langle I_1 \dot{S}_2 I_3 \rangle = -2(\tau + \gamma) \langle I_1 S_2 I_3 \rangle + \tau \langle S_1 S_2 I_3 I_4 \rangle + \tau \langle I_1 S_2 S_3 I_4 \rangle + \tau \langle S_1 S_2 I_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle, \quad (111)$$

$$\langle I_1 \dot{I}_2 S_3 \rangle = -2(\tau + \gamma) \langle I_1 I_2 S_3 \rangle + \tau \langle S_1 I_2 S_3 I_4 \rangle + \tau \langle I_1 I_2 S_3 I_4 \rangle + \tau \langle S_1 I_2 S_3 \rangle + \tau \langle I_1 S_2 S_3 \rangle, \quad (112)$$

$$\langle I_2 \dot{S}_3 I_4 \rangle = -2(\tau + \gamma) \langle I_2 S_3 I_4 \rangle + \tau \langle I_1 S_2 S_3 I_4 \rangle + \tau \langle I_1 I_2 S_3 S_4 \rangle - \tau \langle I_1 I_2 S_3 I_4 \rangle, \quad (113)$$

$$\langle I_2 \dot{S}_3 S_4 \rangle = -(\tau + \gamma) \langle I_2 S_3 S_4 \rangle + \tau \langle I_1 S_2 S_3 S_4 \rangle - 2\tau \langle I_1 I_2 S_3 S_4 \rangle, \quad (114)$$

$$\langle S_2 \dot{S}_3 I_4 \rangle = -(\tau + \gamma) \langle S_2 S_3 I_4 \rangle + \tau \langle I_1 S_2 S_3 S_4 \rangle - 2\tau \langle I_1 S_2 S_3 I_4 \rangle, \quad (115)$$

$$\langle I_1 \dot{S}_2 S_3 \rangle = -(2\tau + \gamma) \langle I_1 S_2 S_3 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle - \tau \langle I_1 S_2 S_3 I_4 \rangle, \quad (116)$$

$$\langle S_1 \dot{I}_2 I_3 I_4 \rangle = -3(\tau + \gamma) \langle S_1 I_2 I_3 I_4 \rangle + \tau (\langle S_1 I_2 I_3 S_4 \rangle + 2\langle S_1 I_2 S_3 I_4 \rangle + \langle S_1 S_2 I_3 I_4 \rangle), \quad (117)$$

$$\langle \dot{S}_1 I_2 I_3 S_4 \rangle = -(3\tau + 2\gamma) \langle S_1 I_2 I_3 S_4 \rangle + \tau (\langle S_1 S_2 I_3 S_4 \rangle + \langle S_1 I_2 S_3 S_4 \rangle), \quad (118)$$

$$\langle \dot{S}_1 I_2 S_3 I_4 \rangle = -(4\tau + 2\gamma) \langle S_1 I_2 S_3 I_4 \rangle, \quad (119)$$

$$\langle \dot{S}_1 S_2 I_3 I_4 \rangle = -(3\tau + 2\gamma) \langle S_1 S_2 I_3 I_4 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle + \tau \langle S_1 S_2 I_3 S_4 \rangle, \quad (120)$$

$$\langle \dot{S}_1 S_2 I_3 S_4 \rangle = -(3\tau + \gamma) \langle S_1 S_2 I_3 S_4 \rangle, \quad (121)$$

$$\langle \dot{S}_1 S_2 S_3 I_4 \rangle = -(2\tau + \gamma) \langle S_1 S_2 S_3 I_4 \rangle, \quad (122)$$

$$\langle \dot{S}_1 I_2 S_3 S_4 \rangle = -(2\tau + \gamma) \langle S_1 I_2 S_3 S_4 \rangle, \quad (123)$$

$$\langle \dot{I}_1 S_2 S_3 S_4 \rangle = -(3\tau + \gamma) \langle I_1 S_2 S_3 S_4 \rangle, \quad (124)$$

$$\langle \dot{I}_1 S_2 S_3 I_4 \rangle = -(3\tau + 2\gamma) \langle I_1 S_2 S_3 I_4 \rangle + \tau \langle I_1 S_2 S_3 S_4 \rangle + \tau \langle S_1 S_2 S_3 I_4 \rangle, \quad (125)$$

$$\langle \dot{I}_1 I_2 S_3 S_4 \rangle = -(3\tau + 2\gamma) \langle I_1 I_2 S_3 S_4 \rangle + \tau \langle I_1 S_2 S_3 S_4 \rangle + \tau \langle S_1 I_2 S_3 S_4 \rangle, \quad (126)$$

$$\langle \dot{I}_1 I_2 S_3 I_4 \rangle = -3(\tau + \gamma) \langle I_1 I_2 S_3 I_4 \rangle + \tau \langle I_1 S_2 S_3 I_4 \rangle + \tau \langle I_1 I_2 S_3 S_4 \rangle + 2\tau \langle S_1 I_2 S_3 I_4 \rangle. \quad (127)$$

## 9 Appendix: Equations for the star-triangle network

In this section we write down the system of differential equations that are an exact representation of the *SIR* epidemic on the star triangle-network, see Fig. 1g. To simplify the notation, nodes within triangle  $i$ , i.e.  $t_i^1$  and  $t_i^2$ , are now denoted by  $i_1$  and  $i_2$ . The relevant equations are:

$$\langle \dot{S}_1 \rangle = -\tau \sum_{j=1}^M \sum_{k=1}^2 \langle S_1 I_{jk} \rangle, \quad (128)$$

$$\langle \dot{I}_1 \rangle = +\tau \sum_{j=1}^M \sum_{k=1}^2 \langle S_1 I_{jk} \rangle - \gamma \langle I_1 \rangle, \quad (129)$$

$$\langle \dot{S}_{i_1} \rangle = -\tau \langle I_1 S_{i_1} \rangle - \tau \langle S_{i_1} I_{i_2} \rangle, \quad (130)$$

$$\langle \dot{I}_{i_1} \rangle = \tau \langle I_1 S_{i_1} \rangle + \tau \langle S_{i_1} I_{i_2} \rangle - \gamma \langle I_{i_1} \rangle, \quad (131)$$

$$\langle \dot{S}_{i_2} \rangle = -\tau \langle I_1 S_{i_2} \rangle - \tau \langle I_{i_1} S_{i_2} \rangle, \quad (132)$$

$$\langle \dot{I}_{i_2} \rangle = \tau \langle I_1 S_{i_2} \rangle + \tau \langle I_{i_1} S_{i_2} \rangle - \gamma \langle I_{i_2} \rangle, \quad (133)$$

$$\langle \dot{S}_{i_1} I_{i_2} \rangle = -(\tau + \gamma) \langle S_{i_1} I_{i_2} \rangle + \tau \langle I_1 S_{i_1} S_{i_2} \rangle - \tau \langle I_1 S_{i_1} I_{i_2} \rangle, \quad (134)$$

$$\langle \dot{I}_{i_1} S_{i_2} \rangle = -(\tau + \gamma) \langle I_{i_1} S_{i_2} \rangle + \tau \langle I_1 S_{i_1} S_{i_2} \rangle - \tau \langle I_1 I_{i_1} S_{i_2} \rangle, \quad (135)$$

$$\langle \dot{I}_1 S_{i_1} \rangle = -(\tau + \gamma) \langle I_1 S_{i_1} \rangle + \tau \langle S_1 S_{i_1} I_{i_2} \rangle - \tau \langle I_1 S_{i_1} I_{i_2} \rangle + \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (136)$$

$$\langle \dot{I}_1 S_{i_2} \rangle = -(\tau + \gamma) \langle I_1 S_{i_2} \rangle + \tau \langle S_1 I_{i_1} S_{i_2} \rangle - \tau \langle I_1 I_{i_1} S_{i_2} \rangle + \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (137)$$

$$\langle \dot{S}_1 I_{i_1} \rangle = -(\tau + \gamma) \langle S_1 I_{i_1} \rangle - \tau \langle S_1 I_{i_1} I_{i_2} \rangle + \tau \langle S_1 S_{i_1} I_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 I_{i_1} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (138)$$

$$\langle \dot{S}_1 I_{i_2} \rangle = -(\tau + \gamma) \langle S_1 I_{i_2} \rangle - \tau \langle S_1 I_{i_1} I_{i_2} \rangle + \tau \langle S_1 I_{i_1} S_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 I_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (139)$$

$$\langle \dot{S}_1 S_{i_1} \rangle = -\tau \langle S_1 S_{i_1} I_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (140)$$

$$\langle S_1 \dot{S}_{i_2} \rangle = -\tau \langle S_1 I_{i_1} S_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (141)$$

$$\langle I_1 \dot{S}_{i_1} S_{i_2} \rangle = -(2\tau + \gamma) \langle I_1 S_{i_1} S_{i_2} \rangle + \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (142)$$

$$\begin{aligned} \langle I_1 \dot{S}_{i_1} I_{i_2} \rangle &= -2(\tau + \gamma) \langle I_1 S_{i_1} I_{i_2} \rangle + \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} I_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle} \\ &\quad + \tau \langle I_1 S_{i_1} S_{i_2} \rangle + \tau \langle S_1 S_{i_1} I_{i_2} \rangle, \end{aligned} \quad (143)$$

$$\begin{aligned} \langle I_1 \dot{I}_{i_1} S_{i_2} \rangle &= -2(\tau + \gamma) \langle I_1 I_{i_1} S_{i_2} \rangle + \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 I_{i_1} S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle} \\ &\quad + \tau \langle I_1 S_{i_1} S_{i_2} \rangle + \tau \langle S_1 I_{i_1} S_{i_2} \rangle, \end{aligned} \quad (144)$$

$$\langle S_1 \dot{S}_{i_1} I_{i_2} \rangle = -(2\tau + \gamma) \langle S_1 S_{i_1} I_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} I_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (145)$$

$$\langle S_1 \dot{I}_{i_1} S_{i_2} \rangle = -(2\tau + \gamma) \langle S_1 I_{i_1} S_{i_2} \rangle - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 I_{i_1} S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (146)$$

$$\begin{aligned} \langle S_1 \dot{I}_{i_1} I_{i_2} \rangle &= -2(\tau + \gamma) \langle S_1 I_{i_1} I_{i_2} \rangle + \tau \langle S_1 S_{i_1} I_{i_2} \rangle + \tau \langle S_1 I_{i_1} S_{i_2} \rangle \\ &\quad - \tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 I_{i_1} I_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \end{aligned} \quad (147)$$

$$\langle S_1 \dot{S}_{i_1} S_{i_2} \rangle = -\tau \sum_{j=1, j \neq i}^M \sum_{k=1}^2 \frac{\langle S_1 S_{i_1} S_{i_2} \rangle \langle S_1 I_{jk} \rangle}{\langle S_1 \rangle}, \quad (148)$$

where, in Eqs. (136 – 148) we have used the following closures:

$$\begin{aligned} \langle S_1 S_{i_1} S_{i_2} I_{k_j} \rangle \langle S_1 \rangle &= \langle S_1 S_{i_1} S_{i_2} \rangle \langle S_1 I_{k_j} \rangle, \\ \langle S_1 S_{i_1} I_{k_j} \rangle \langle S_1 \rangle &= \langle S_1 S_{i_1} \rangle \langle S_1 I_{k_j} \rangle, \\ \langle S_1 S_{i_2} I_{k_j} \rangle \langle S_1 \rangle &= \langle S_1 S_{i_2} \rangle \langle S_1 I_{k_j} \rangle, \\ \langle S_1 S_{i_1} I_{i_2} I_{k_j} \rangle \langle S_1 \rangle &= \langle S_1 S_{i_1} I_{i_2} \rangle \langle S_1 I_{k_j} \rangle, \end{aligned}$$

and

$$\langle S_1 I_{i_1} S_{i_2} I_{k_j} \rangle \langle S_1 \rangle = \langle S_1 I_{i_1} S_{i_2} \rangle \langle S_1 I_{k_j} \rangle,$$

for  $i, k = 1, 2, \dots, M$ ,  $i \neq k$ , and  $j = 1, 2$ .

These closures are of two main type, namely:

1. closure of a triple which is not a triangle ( $i \neq j$ ):

$$\langle X_{i_l} S_1 Y_{j_k} \rangle \langle S_1 \rangle = \langle X_{i_l} S_1 \rangle \langle S_1 Y_{j_k} \rangle \quad (149)$$

where  $l, k = 1, 2$ ,  $i, j = 1, \dots, M$  and  $X, Y$  are either  $S$  or  $I$  in some particular combination.

2. closure of a quadruple containing a triangle:

$$\langle X_{i_1} Y_{i_2} S_1 Z_{j_k} \rangle \langle S_1 \rangle = \langle X_{i_1} Y_{i_2} S_1 \rangle \langle S_1 Z_{j_k} \rangle \quad (150)$$

where  $k = 1, 2$ ,  $i, j = 1, \dots, M$  and  $X, Y, Z$  are either  $S$  or  $I$  in some particular combination.

## References

- [1] F. Ball, D. Sirl & P. Trapman (2010) Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.* **224** 5373.
- [2] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross & M. C. Vernon (2012) “Networks and the Epidemiology of Infectious Disease”. *Interdisciplinary Perspectives on Infectious Diseases* **284909** special issue “Network Perspectives on Infectious Disease Dynamics”.
- [3] R. Diestel. Graph theory. Springer-Verlag, Heidelberg, New York, 2005.
- [4] J.P. Gleeson (2009) Bond percolation on a class of clustered random networks. *Phys. Rev. E.* **80** 036107.
- [5] D.M. Green & I.Z. Kiss (2010) Large-scale properties of clustered networks: Implications for disease dynamics. *Journal of Biological Dynamics* **4** 431-445.
- [6] T. House & M.J. Keeling (2011) Insights from unifying modern approximations to infections on networks. *J. Roy. Soc. Interface* **8** 67-73.
- [7] B. Karrer & M.E.J. Newman (2010) Message passing approach for general epidemic models. *Phys. Rev. E.* **82** 016101.
- [8] I.Z. Kiss & D.M. Green (2008) Comment on “Properties of highly clustered networks”. *Phys. Rev. E* **78** 048101.
- [9] J.C. Miller (2009) Percolation and epidemics in random clustered networks. *Phys. Rev. E* **80** 020901.
- [10] M.E.J. Newman (2003) Properties of highly clustered networks. *Phys. Rev. E* **68** 026121.
- [11] M.E.J. Newman (2009) Random graphs with clustering. *Phys. Rev. Lett.* **103** 058701.
- [12] K.J. Sharkey (2008) Deterministic epidemiological models at the individual level. *J. Math. Biol.* **57** 311-331.
- [13] K.J. Sharkey (2011) Deterministic epidemic models on contact networks: Correlations and unbiological terms. *Theor. Popul. Biol.* **79** 115-129.
- [14] K.J. Sharkey, I.Z. Kiss, R.R. Wilkinson & P.L. Simon (2012) Exact equations for SIR epidemics on tree graphs. *Bull. Math. Biol.* DOI [10.1007/s11538-013-9923-5](https://doi.org/10.1007/s11538-013-9923-5) .
- [15] R. Sedgewick. Algorithms in C: graph algorithms. Addison-Wesley, 2002.
- [16] P.L. Simon, M. Taylor & I.Z. Kiss (2011) Exact epidemic models on graphs using graph-automorphism driven lumping. *J. Math. Biol.* **62** 479-508.
- [17] P. Trapman (2007) On analytical approaches to epidemics on networks. *Theor. Popul. Biol.* **71** 160-173.
- [18] T.J. Taylor & I.Z. Kiss (2013) Interdependency and hierarchy of exact and approximate epidemic models on networks. Published online in *J. Math. Biol.* DOI [10.1007/s00285-013-0699-x](https://doi.org/10.1007/s00285-013-0699-x).

- [19] E.M. Volz (2004) Random networks with tunable degree distribution and clustering *Phys. Rev. E* **70** 056115.
- [20] E.M. Volz, J.C. Miller, A. Galvani & L.A. Meyers (2011) Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput. Biol.* **7** e1002042.