Exact epidemic models on graphs using graph-automorphism driven lumping

Péter L. Simon · Michael Taylor · Istvan Z. Kiss

Abstract   The dynamics of disease transmission strongly depends on the properties of the population contact network. Pair-approximation models and individual-based network simulation have been used extensively to model contact networks with non-trivial properties. In this paper, using a continuous time Markov chain, we start from the exact formulation of a simple epidemic model on an arbitrary contact network and rigorously derive and prove some known results that were previously mainly justified based on some biological hypotheses. The main result of the paper is the illustration of the link between graph automorphisms and the process of lumping whereby the number of equations in a system of linear differential equations can be significantly reduced. The main advantage of lumping is that the simplified lumped system is not an approximation of the original system but rather an exact version of this. For a special class of graphs, we show how the lumped system can be obtained by using graph automorphisms. Finally, we discuss the advantages and possible applications of exact epidemic models and lumping.

Keywords   Network · Epidemic · Markov chain · Lumping · Graph automorphism

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

1 Introduction

The spread and persistence of infectious diseases is a result of the complex interaction between the behaviour of individual epidemiological units (e.g. individual, city,

P. L. Simon
Institute of Mathematics, Eötvös Loránd University Budapest, Budapest, Hungary

M. Taylor · I. Z. Kiss (✉)
Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9RF, UK
e-mail: i.z.kiss@sussex.ac.uk

county etc), disease characteristics and various control programmes that are aimed at halting disease transmission or bringing infection prevalence to as low a level as possible (Anderson and May 1991). The main aim of many models is to gain insight into how diseases transmit and to identify the most effective strategies for their prevention and control. The early work by Kermack and McKendrick (1927) forms the basis of differential-equation-based models which lie at the heart of modern quantitative epidemiology. Traditionally, mathematical epidemiology is based on differential equation models (Anderson and May 1991; Diekmann and Heesterbeek 2000) and these operate on the basis of some strong simplifying assumptions about the behaviour of the individuals and the biology of the disease. While these models do not explicitly model contact-network structure, they allow us to improve our understanding of the disease transmission process and to derive threshold quantities such as the basic reproduction number R_0 and critical vaccination threshold (Diekmann and Heesterbeek 2000) in terms of key biological parameters such as rate of infection and infectious period.

A key component in any disease transmission model is the population contact structure (i.e. the collection of disease causing contacts between individual epidemiological units) (Ball et al. 1997; Diekmann et al. 1998). In most cases, this is highly heterogeneous with strong correlations and non-trivial large scale structure (Hufnagel et al. 2004; Kao et al. 2006; Kiss et al. 2006a) which can make it difficult to use compartmental differential equations. In an attempt to increase model accuracy, in the context of ecological and epidemiological modelling, various pair-approximation (Keeling 1999; Keeling et al. 1997; Rand 1999; Sato et al. 1994; van Baalen 2000) and individual-based models (Keeling and Eames 2005) have been developed. These more sophisticated models can more accurately capture contact network properties but often at the expense of limited analytical tractability. Pair-approximation models focus on the interaction between pairs of individuals and assume that changes in the status of individuals depends on the status of their neighbours. A key factor in such models is the closing relation which is used to truncate and close equations at the level of pairs. This is a useful approach which works well for certain classes of networks, although its precise domain of applicability remains unknown.

Individual-based models where each individual is represented as a node and potentially infectious links between these as edges is a more flexible approach that can account for many complexities in the contact structure (Keeling and Eames 2005; Kenah and Robins 2007; Newman 2002). However, this comes at the price of little analytical tractability, unless some strong limiting assumptions (i.e. proportionate mixing), are made. Individual-based network models fall in two broad categories. First, those where network data is available and can be used to specify the contact network (Dent et al. 2008; Green et al. 2006; Kao et al. 2006; Kiss et al. 2006a), and second, theoretical network models that focus on understanding the impact of specific network properties on outbreak threshold, final epidemic size or prevalence level and the efficacy of control measures (Keeling 1999, 2005; Kiss et al. 2005, 2008; May and Lloyd 2001). The latter models are used to establish some general principles while the former are driven, at least partially, by the real-time predictive modelling of human (SARS, Hufnagel et al. 2004; Lipsitch et al. 2003; Meyers et al. 2005; and the current swine-flu outbreak, Smith et al. 2009) and animal disease (foot-and-mouth

disease, Ferguson et al. 2001; Kao et al. 2006; Avian Influenza, Dent et al. 2008)
outbreaks.

Modelling disease transmission in finite populations using stochastic simulations
alone poses many challenges. For example, to tease apart rare events from average
system behaviour many replicate simulations are needed. A more rigorous mod-
elling alternative as discussed by Keeling and Ross (2008) is the use of Markov
chains (Kemeny and Snell 1976) where the future state of the population depends
on current state only (memoryless process) and the probability of the population
being in any of the possible states at a given time is determined by a system of
linear differential equations that can be constructed based on the rates of transition
between states. In this paper, using a continuous time Markov chain (CTMC), we
start from the exact formulation of a simple epidemic model on an arbitrary con-
tact network. We rigorously derive and prove some known results that were previ-
ously mainly justified directly from biological hypotheses rather than being deduced
from an exact system of equations (Rand 1999). We revisit lumping (see Appen-
dix A in Diekmann and Heesterbeek 2000 and Filliger and Hongler 2008; Jacobi
and Görnerup 2009) and illustrate how graph automorphisms can be used to guide
the lumping process which leads to a significant reduction in the large number of
equations ($2^N - 1$ for a model with two disease states, such as the SIS model, on a
network with N nodes) of the original exact system. For a special class of graphs,
we illustrate how lumping can be directly inferred from graph automorphisms and
we combine lumping on a complete graph with simple mathematical techniques to
show that, in the limit of large networks, the solution of the exact Markov chain model
converges to that of the ODE-based mean-field model. In addition, we discuss the
advantages, disadvantages and possible applications of exact epidemic models and
lumping.

2 Model

2.1 Disease dynamics and the network of contacts

The SIS type dynamics (Anderson and May 1991) is considered on a network with
N nodes and with adjacency matrix $G = (g_{ij})_{i,j=1,2,...,N} \in \{0, 1\}^{N^2}$ where $g_{ij} = 1$ if
node i and j are connected, and $g_{ij} = 0$ otherwise. Here, we only consider networks
with bi-directional edges and without self-loops. This requirement translates to the
following two constraints upon G: $G = G^T$ and $g_{ii} = 0$. The dynamics of trans-
mission has two key stages: transmission of the disease and recovery of infectious
individuals. Infection starting from an initial index case is transmitted at rate τ across
every (S, I) edge. This is followed by the recovery of infectious individuals at rate γ.
Upon recovery, infectious individuals become susceptible again. Both infection and
recovery are modelled as independent Poisson processes. For example, in a small time
interval δt, a susceptible individual with k_I infectious neighbours becomes infected
with probability $1 - \exp(-k_I \tau \delta t)$. Similarly, $1 - \exp(-\gamma \delta t)$ represents the recovery
probability of a single infectious individual, and this is independent of the status of its
neighbours.

2.2 Formulation of the disease transmission model

At any point in time, every node can be either susceptible (S) or infected and infectious (I), and hence, the state of the system is given by a vector of length N with all entries either S or I (or zero and one). The state space of system is given by $\mathcal{S} = \{S, I\}^N$ or $\mathcal{S} = \{0, 1\}^N$ and the transmission dynamics on the network can be formulated in terms of a transition matrix between all possible states. In the case of continuous time, this matrix is also known as the infinitesimal generator matrix (Brauer et al. 2008; Daley and Gani 1999; Kemeny and Snell 1976). Based on the generator matrix, it is straightforward to write down the Kolmogorov differential equations that uniquely determine the probability of the system being in a particular state at a given time (Brauer et al. 2008; Daley and Gani 1999). However, in practice even for small network size, this approach becomes unfeasible due to the large number of equations (i.e. $2^N - 1$). As detailed in the introduction, using various techniques, it is possible to reduce the number of states (i.e. number of equations) and derive models that are either equivalent to the original system or are approximations that in the limit of large graphs (i.e. $N \to \infty$) become exact.

2.2.1 Kolmogorov equations in the case of an arbitrary graph

The 2^N elements of the state space can be conveniently divided into $N + 1$ subsets as follows: (a) $\mathcal{S}^0$ is the subset with one single element, namely the state with all nodes susceptible: $\mathcal{S}^0 = (S, S, \ldots, S)$, (b) $\mathcal{S}^k$ is the subset of all states with k infected nodes (on a graph with N nodes, k infected nodes can be arranged in $\binom{N}{k}$ distinct configurations), and (c) $\mathcal{S}^N$ is the subset with one single element, namely the state with all nodes infected: $\mathcal{S}^N = (I, I, \ldots, I)$. The elements of the subset $\mathcal{S}^k$ are denoted by $\mathcal{S}_1^k, \mathcal{S}_2^k, \ldots, \mathcal{S}_{c_k}^k$, where $c_k = \binom{N}{k}$. The status of the lth node of state $\mathcal{S}_j^k$ will be denoted by $\mathcal{S}_j^k(l)$, thus $\mathcal{S}_j^k(l) = S$ or $\mathcal{S}_j^k(l) = I$. The state of the system can change in two ways:

-   *A node becoming infected:* an S node becomes an I node, that is an $\mathcal{S}_j^k \to \mathcal{S}_i^{k+1}$ type transition, where j and i are chosen such that $\exists\, l$ for which $\mathcal{S}_j^k(l) = S$, $\mathcal{S}_i^{k+1}(l) = I$, and $\mathcal{S}_j^k(m) = \mathcal{S}_i^{k+1}(m)$ for $\forall m \neq l$. Moreover, $\exists\, r \neq l$ such that $\mathcal{S}_j^k(r) = I$ and $g_{lr} = 1$ (i.e. there is an (S, I) type edge between nodes labelled l and r).
-   *The recovery of a node:* an I node becomes an S node, that is an $\mathcal{S}_j^k \to \mathcal{S}_i^{k-1}$ type transition, where j and i are chosen such that $\exists\, l$ for which $\mathcal{S}_j^k(l) = I$, $\mathcal{S}_i^{k-1}(l) = S$, and $\mathcal{S}_j^k(m) = \mathcal{S}_i^{k-1}(m)$ for $\forall m \neq l$. This means that states $\mathcal{S}_j^k$ and $\mathcal{S}_i^{k-1}$ may differ only at one position, this is the l'th position.

The evolution in the state space can be described by a continuous time Markov-process. Let us denote the probability of the system being in state $\mathcal{S}_j^k$ at time t by $X_j^k(t)$. Let

$$X^k(t) = (X_1^k(t), X_2^k(t), \ldots, X_{c_k}^k(t))$$

be a c_k-dimensional vector for $k = 0, 1, \ldots, N$. The above transitions determine the Kolmogorov equations (i.e. a system of linear differential equations) for the probability functions $X_j^k(t)$. In the general case of an arbitrary graph, the Kolmogorov equations can be written in terms of a matrix with the following block tridiagonal form

$$\dot{X}^k = A^k X^{k-1} + B^k X^k + C^k X^{k+1}, \qquad k = 0, 1, \ldots, N, \tag{1}$$

where A^0 and C^N are zero matrices. Thus Eq. 1 can be written in the following form

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

where

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

We note that often the transpose of P is used, and then X is a row vector, not a column. However, here we use this formulation since it is more convenient from a dynamical system point of view. The A^k matrices capture the infection while the C^k matrices describe the recovery process. These matrices depend on the structure of the network, and the transmission and recovery rates. The dimension and the entries of these matrices can be obtained using some simple bookkeeping rules.

The entry in the ith row and jth column of the matrix A^k is denoted by $A_{i,j}^k$. This element gives the rate of transition from $\mathcal{S}_j^{k-1}$ to $\mathcal{S}_i^k$. In the $\mathcal{S}^{k-1}$ class there are c_{k-1} elements, and in the $\mathcal{S}^k$ class there are c_k elements, hence matrix A^k has c_k rows and c_{k-1} columns. The entry $A_{i,j}^k$ is non-zero only in the case when the states $\mathcal{S}_j^{k-1}$ and $\mathcal{S}_i^k$ differ at one position, say at position l and $\mathcal{S}_j^{k-1}(l) = S$, $\mathcal{S}_i^k(l) = I$, and $\mathcal{S}_j^{k-1}(m) = \mathcal{S}_i^k(m)$ for $\forall m \neq l$. Moreover, we also require that there $\exists~r \neq l$ such that $\mathcal{S}_j^{k-1}(r) = I$ and $g_{lr} = 1$ (i.e. there is an (S, I) type edge between nodes labelled l and by r). In this case

$$A_{i,j}^k = \tau \cdot \#\{r \in \{1, 2, \ldots, N\} : \mathcal{S}_j^{k-1}(r) = I, g_{lr} = 1\}. \tag{2}$$

In order to better understand the role of the A^k matrix let us consider a state $\mathcal{S}_j^{k-1}$ and choose an index l such that the node at position l is S (i.e. $\mathcal{S}_j^{k-1}(l) = S$). Next, find an index $i \in \{1, 2, \ldots, c_k\}$ such that states $\mathcal{S}_j^{k-1}$ and $\mathcal{S}_i^k$ only differ at position l, that is $\mathcal{S}_i^k(l) = I$ and $\mathcal{S}_j^{k-1}(m) = \mathcal{S}_i^k(m)$ for $\forall m \neq l$. Let us denote by q the number of nodes in state $\mathcal{S}_j^{k-1}$ which are I and are connected to node l which is S. Then $A_{i,j}^k = \tau q$. Repeating this for all $l \in \{1, 2, \ldots, N\}$, such that $\mathcal{S}_j^{k-1}(l) = S$, it can be observed

that the total number of (S, I) edges in the $\mathcal{S}_j^{k-1}$ state multiplied by τ is equal to the sum of the elements in the jth column of matrix A^k, that is for $\forall j \in \{1, 2, \ldots, c_{k-1}\}$ the following equality holds

$$\sum_{i=1}^{c_k} A_{i,j}^k = \tau N_{SI}(\mathcal{S}_j^{k-1}),$$ (3)

where $N_{SI}(\mathcal{S}_j^{k-1})$ denotes the number of (S, I) edges in state $\mathcal{S}_j^{k-1}$.

The entry in the ith row and jth column of matrix C^k is denoted by $C_{i,j}^k$. This element gives the rate of transition from $\mathcal{S}_j^{k+1}$ to $\mathcal{S}_i^k$. In the $\mathcal{S}^{k+1}$ class there are c_{k+1} elements, and in the $\mathcal{S}^k$ class there are c_k elements, hence matrix C^k has c_k rows and c_{k+1} columns. The entry $C_{i,j}^k$ is non-zero only in the case when the states $\mathcal{S}_j^{k+1}$ and $\mathcal{S}_i^k$ differ at one position, say at position l such that $\mathcal{S}_j^{k+1}(l) = I, \mathcal{S}_i^k(l) = S$ and $\mathcal{S}_j^{k+1}(m) = \mathcal{S}_i^k(m)$ for $\forall m \neq l$. In this case $C_{i,j}^k = \gamma$. In state $\mathcal{S}_j^{k+1}$, $k + 1$ nodes of the graph are in state I, hence in the jth column of matrix C^k there are $k+1$ entries that are equal to γ and the remaining entries are zero. Hence, for all $j \in \{1, 2, \ldots, c_{k+1}\}$ the following equality holds

$$\sum_{i=1}^{c_k} C_{i,j}^k = \gamma(k + 1).$$ (4)

The matrix B^k is a diagonal matrix with the number of rows and columns equal to c_k. This is due to B^k only accounting for the rate of the $\mathcal{S}_i^k \to \mathcal{S}_j^k$ type transitions, where the rate of a transition from $\mathcal{S}_i^k$ to $\mathcal{S}_j^k$ is equal zero if $i \neq j$. The diagonal elements of B^k are determined such that the sum of entries in each column of P sum to zero. This gives the following expression for the entries of B,

$$B_{i,i}^k = -\sum_{j=1}^{c_{k+1}} A_{j,i}^{k+1} - \sum_{j=1}^{c_{k-1}} C_{j,i}^{k-1}.$$ (5)

In the Appendix, the rules detailed above are used to derive the Kolmogorov equations for a complete graph with $N = 3$. We note that the above rules can be conveniently programmed in a code, using *Matlab*, that will automatically generate and provide numerical solution to the full set of differential equations.

2.2.2 Approximation and simulation methods

The system given by Eq. 1 consists of 2^N linear differential equations that are impractical to solve or cannot be solved for large N. However, it is not always necessary to determine all probability functions, and in many situations, the expected values of the number, or proportion, of susceptible (S) and infectious (I) nodes or individuals is

equally valuable. These expected values at time t are denoted by $[S](t)$ and $[I](t)$ and can be expressed as follows

$$[I](t) = \sum_{k=0}^{N} k \sum_{j=1}^{c_k} X_j^k(t), \qquad [S](t) = \sum_{k=0}^{N} (N - k) \sum_{j=1}^{c_k} X_j^k(t). \qquad (6)$$

There are three different approaches to determine or approximate these exact values as given by the system in Eq. 1. The simplest way is through individual-based network simulation (Keeling and Eames 2005; Kiss et al. 2005, 2008). However, this offers little insight into the dynamics, and results obtained from simulation are difficult to generalise. An alternative is the derivation of mean-field type equations upon using pair or triple approximations (House et al. 2009; Keeling 1999; Keeling et al. 1997; Rand 1999; Sato et al. 1994; van Baalen 2000). Finally, where appropriate, by using the technique of lumping (see Appendix A in Diekmann and Heesterbeek 2000 and Filliger and Hongler 2008; Jacobi and Görnerup 2009), the 2^N equations can be reduced to a tractable number of equations that are not an approximation of the original system but an exact equivalent. The first two techniques are well known and widely used in Mathematical Epidemiology where a considerable amount of research focuses on models of disease transmission within populations with complex contact structure.

3 Mean-field and closure approximations

In this section, we report on various approximation techniques and for some mean-field type models we show how these can be derived directly from the Kolmogorov equations. We also formalise rigorously the validity of approximation models and discuss the limit in which approximations become exact.

3.1 Mean-field equations

The idea of deriving mean-field equations is based on the observation that $[S]$ and $[I]$ are linear combinations of the probability functions in Eq. 1. Hence, based on Eq. 1, it is feasible to try to derive a new system of differential equations that will uniquely define $[S]$ and $[I]$. Introducing $[SI]$ as the expected value of (S, I) pairs (directed edges), that is

$$[SI](t) = \sum_{k=0}^{N} \sum_{j=1}^{c_k} N_{SI}(\mathcal{S}_j^k) X_j^k(t), \qquad (7)$$

we obtain the following system:

Lemma 1 *The expected values $[S]$ and $[I]$ satisfy the following system*

$$[\dot{S}] = \gamma[I] - \tau[SI], \tag{8}$$
$$[\dot{I}] = \tau[SI] - \gamma[I]. \tag{9}$$

Despite this being a well-known system (Rand 1999), we provide a proof of this Lemma because of two reasons. On one hand, this system is usually not derived from the differential equations of the Markov-process, and on the other hand, the proof illustrates the usefulness of the block tridiagonal form of the transition matrix P.

Proof Let us introduce the row matrix $S_k = (1\ 1\ \ldots 1)$ with c_k columns. Then $\sum_{j=1}^{c_k} X_j^k = S_k X^k$ (here X^k is a column vector). Hence from Eq. 6 we obtain

$$[I](t) = \sum_{k=0}^{N} k S_k X^k, \qquad [S](t) = \sum_{k=0}^{N} (N-k) S_k X^k. \tag{10}$$

Using this notation, Eq. 5 takes the following form,

$$B_{i,i}^k = -(S_{k+1} A^{k+1})_i - (S_{k-1} C^{k-1})_i,$$

and using that B^k is a diagonal matrix $B_{i,i}^k = (S_k B^k)_i$ holds, and hence,

$$(S_k B^k)_i = -(S_{k+1} A^{k+1})_i - (S_{k-1} C^{k-1})_i$$

which is true for $\forall i = 1, \ldots c_k$. Thus the following equation is obtained

$$S_{k+1} A^{k+1} + S_k B^k + S_{k-1} C^{k-1} = 0, \tag{11}$$

and this holds for $\forall k = 0, 1, \ldots, N$. We note that, when indices are out of the relevant range (i.e. A^{N+1} and C^{-1}) matrices should be set to zero. Differentiating $[I](t)$ and using Eq. 1 we obtain

$$
\begin{aligned}
[\dot{I}] &= \sum_{k=0}^{N} k S_k \dot{X}^k = \sum_{k=0}^{N} k S_k (A^k X^{k-1} + B^k X^k + C^k X^{k+1}) \\
&= \sum_{k=1}^{N} k S_k A^k X^{k-1} + \sum_{k=0}^{N} k S_k B^k X^k + \sum_{k=0}^{N-1} k S_k C^k X^{k+1} \\
&= \sum_{k=0}^{N-1} (k+1) S_{k+1} A^{k+1} X^k + \sum_{k=0}^{N} k S_k B^k X^k + \sum_{k=1}^{N} (k-1) S_{k-1} C^{k-1} X^k \\
&= \sum_{k=0}^{N} \left((k+1) S_{k+1} A^{k+1} + k S_k B^k + (k-1) S_{k-1} C^{k-1} \right) X^k.
\end{aligned}
$$

Now upon using Eq. 11, we obtain

$$[\dot{I}] = \sum_{k=0}^{N} \left(S_{k+1} A^{k+1} - S_{k-1} C^{k-1} \right) X^k.$$

The statement for $[I](t)$ follows from the Proposition below. The proof for $[S](t)$ is similar. $\qquad\square$

Proposition 1 1. $S_{k-1} C^{k-1} = \gamma k S_k$
2. $\sum_{k=0}^{N} S_{k-1} C^{k-1} X^k = \gamma [I]$
3. $\sum_{k=0}^{N} S_{k+1} A^{k+1} X^k = \tau [SI]$

Proof 1. According to Eq. 4, for all $j \in \{1, 2, \ldots, c_k\}$ the following equality holds

$$(S_{k-1} C^{k-1})_j = \sum_{i=1}^{c_{k-1}} C_{i,j}^{k-1} = \gamma k,$$

and this implies that $S_{k-1} C^{k-1} = \gamma k S_k$, since the jth coordinate of the left and right-hand side are equal.
2. The second statement is a direct consequence of the first part of the Proposition and Eq. 10.
3. According to Eq. 3, for all $j \in \{1, 2, \ldots, c_k\}$ the following equality holds

$$(S_{k+1} A^{k+1})_j = \sum_{i=1}^{c_{k+1}} A_{i,j}^{k+1} = \tau N_{SI}(\mathcal{S}_j^k),$$

with this yielding

$$\sum_{k=0}^{N} S_{k+1} A^{k+1} X^k = \sum_{k=0}^{N} \sum_{j=1}^{c_k} (S_{k+1} A^{k+1})_j X_j^k = \tau \sum_{k=0}^{N} \sum_{j=1}^{c_k} N_{SI}(\mathcal{S}_j^k) X_j^k(t) = \tau [SI].$$

$\qquad\square$

We note that the Lemma above holds for any network. However, the system given by Eqs. 8–9 is not closed since it contains the new variable $[SI]$. This problem can be solved in two ways. First, we can aim to provide an approximation for the expected number of $[SI]$ pairs in terms of $[S]$ and $[I]$ (i.e. pair approximation), and second, we can attempt to derive a differential equation for $[SI]$.

3.2 Closing at the level of pairs

The simplest closing relation is $[SI] \simeq [S][I]$, and this is based on the statistical independence in the state of individuals (i.e. the expected number of $[SI]$ pairs is equivalent to the product of the expected number of $[S]$ and $[I]$). Applying this relation in

Eqs. 8–9, the well known mean-field model is obtained

$$\dot{\tilde{S}} = \gamma \tilde{I} - \tau \tilde{S} \tilde{I}, \tag{12}$$

$$\dot{\tilde{I}} = \tau \tilde{S} \tilde{I} - \gamma \tilde{I}. \tag{13}$$

Such a closure relation leads to the problem of identifying graphs for which the approximation above holds. It is known that for a large, complete graph this is a good approximation when $\tau = \beta/N$ for some fixed value β (see Appendix A in Diekmann and Heesterbeek 2000 and Kurtz 1971). In order to formalize this statement more precisely it is useful to use the scaled variables:

$$\tilde{s} = \tilde{S}/N, \quad \tilde{i} = \tilde{I}/N, \quad [s] = [S]/N, \quad [i] = [I]/N. \tag{14}$$

In the new variables, with $\tau = \beta/N$, the system given by Eqs. 12–13 takes the following form

$$\dot{\tilde{s}} = \gamma \tilde{i} - \beta \tilde{s} \tilde{i}, \tag{15}$$

$$\dot{\tilde{i}} = \beta \tilde{s} \tilde{i} - \gamma \tilde{i}. \tag{16}$$

It is generally believed that, under appropriate conditions, $\tilde{i}$ is a good approximation of $[i]$. Before formalising this in a mathematically rigorous way we give details of a key aspect which seems to contradict our previous statement. This observation also justifies the need to formalise rigorously the meaning of $\tilde{i}$ being a good approximation of $[i]$.

It is known, and it can also be easily derived from Eqs. 15–16, that in the case of $\beta > \gamma$ the following holds

$$\lim_{t \to \infty} \tilde{i}(t) = 1 - \frac{\gamma}{\beta},$$

and in the case when $\beta \leq \gamma$, this limit is zero. However, the corresponding limit for $[i](t)$ is always zero, that is

$$\lim_{t \to \infty} [i](t) = 0,$$

as shown first in Picard (1965). Nåsell (1996) has shown that in the case of large N, the function $[i]$ has a quasi steady state that is close to the steady state of the ODE system given by Eqs. 15–16. Computing $[i](t)$ and $\tilde{i}(t)$ numerically for $N > 30$, the two nearly coincide on a quite long time interval. However, for very large t, $[i](t)$ will tend to zero. Nåsell (1996) also investigated the length of the time interval on which $\tilde{i}(t)$ is a good approximation of $[i](t)$, while Andersson and Djehich (1997) derived the asymptotic distribution for the extinction time in the limit of large population. The next Theorem specifies the way in which $\tilde{i}(t)$ can be considered a good approximation of $[i](t)$.

Theorem 1 *Let G be a complete graph, and let $\tau = \beta/N$ for some fixed value β. Let X be the solution of Eq. 1 and let $[S]$, $[I]$, $[s]$ and $[i]$ be as defined by Eqs. 6 and 14. Let $(\tilde{s}, \tilde{i})$ be the solutions of the system given by Eqs. 15–16 with the following initial conditions $\tilde{s}(0) = [s](0)$, $\tilde{i}(0) = [i](0)$. Then for any $t \geq 0$ we have*

$$\lim_{N \to \infty} |[s](t) - \tilde{s}(t)| = 0, \qquad \lim_{N \to \infty} |[i](t) - \tilde{i}(t)| = 0.$$

The proof of Theorem 1, as outlined in the Appendix, is based on the fact that in the case of a complete graph the system given by Eq. 1 takes the form given in Lemma 4 below. This significant reduction in dimensionality can be obtained intuitively, or by using the more rigorous formalism of lumping as explained later in the paper. Kurtz (1970) proved that the Markov chain given by the lumped system converges in probability to the solution of the mean-field equation. This statement is stronger than Theorem 1 since it implies that not only the expected value converges to the solution of the mean-field equation but also the distribution. However, the proof of this general statement uses abstract tools, namely an extension of Trotter's operator semigroup approximation theorem. Later Kurtz (1971) formalised the statement in a more abstract setting and proved a central limit theorem using martingale theory. Our proof of Theorem 1 is an alternative proof that does not use abstract semigroup and martingale theory and it is based on ideas outlined in the Appendix A in Diekmann and Heesterbeek (2000).

3.3 Closing at the level of triples

As illustrated in Eqs. 8–9, the dynamics of the number of Ss ($[S]$) and Is ($[I]$) or singles depends on the number of pairs (e.g. $[SI]$), and these in turn will depend on the number of triples. For example the number of $[SS]$ pairs decreases due to infection from outside the pair (i.e. $[SSI]$ triples). Using simple heuristic reasoning and by accounting for all within- and outside-pair transitions the following Lemma is obtained.

Lemma 2 *The expected values of $[S]$, $[I]$, $[SI]$, $[II]$ and $[SS]$ satisfy the following system of differential equations*

$$[\dot{S}] = \gamma[I] - \tau[SI], \tag{17}$$

$$[\dot{I}] = \tau[SI] - \gamma[I], \tag{18}$$

$$[\dot{SI}] = \gamma([II] - [SI]) + \tau([SSI] - [ISI] - [SI]), \tag{19}$$

$$[\dot{II}] = -2\gamma[II] + 2\tau([ISI] + [SI]), \tag{20}$$

$$[\dot{SS}] = 2\gamma[SI] - 2\tau[SSI]. \tag{21}$$

While we will not provide a proof for the above Lemma, we believe that a proof similar to that given for Lemma 1 is possible. Such a proof would again provide a rigorous derivation from the exact epidemic model as given by Eq. 1. Rand (1999) and others (Keeling et al. 1997; Keeling 1999; van Baalen 2000) provide the derivation of

the above equations based on biological hypotheses and without rigorous derivation from some form of master equation.

We can observe that in the equation for $[SI]$ new unknowns, the triples appear. Hence we have two possibilities again. Either derive differential equations for the triples, or approximate the triples with pairs and singles. These approximations are called closing relations. For randomly mixed and clustered homogeneous graphs, where every node has the same number of edges, the following closing relation has been proposed and has been extensively used in disease modelling

$$[ABC] \simeq \frac{n-1}{n} \frac{[AB][BC]}{[B]} \left((1-\phi) + \phi \frac{N}{n} \frac{[AC]}{[A][C]} \right).$$

Here n is the node degree and ϕ is the clustering coefficient (i.e. the ratio of triangles to triples over the whole graph (Keeling 1999)). Using this closing approximations in Eqs. 17–21, a nonlinear system of ODEs for $[S]$, $[I]$, $[SI]$, $[II]$ and $[SS]$ can be obtained.

It is known that closing relations depend on the structure of the graph and these relations have been discussed at length elsewhere (Keeling 1999; Keeling et al. 1997; Rand 1999; Sato et al. 1994; Sharkey 2008; van Baalen 2000). To evaluate the performance of such closing relations, the typical approach in the literature is the following. A closing relation yielding a closed ODE system is introduced heuristically, then its validity is justified numerically by comparing the values of $[S](t)$ and $[I](t)$ obtained numerically from Monte Carlo simulation to those obtained by numerically solving the ODE system. According to our knowledge, finding the exact class of graphs for which the above closing relations or others hold is an open problem. Our approach presented in this paper makes an attempt towards providing rigorous proof for the validity of some closing relations (i.e. see Theorem 1).

4 How to lump using graph automorphism

In this section we revisit the process of lumping linear ordinary differential equations and illustrate the link between network/graph automorphisms and lumping. For some particular graph types, and for an SIS type dynamics, we use lumping to derive exact models where the number of equations compared to the original complete system is significantly reduced. In this case, the lumped system is a reduced but exact version of the original system.

4.1 General definition of lumping

We start by defining lumping in general. Let us consider an n-dimensional linear system $\dot{X} = AX$, where A is an arbitrary $n \times n$ matrix.

Definition 1 The linear system $\dot{X} = AX$ (or equivalently the corresponding Markov-process) is called lumpable, if there is a partition $\{L_1, L_2, \ldots, L_m\}$ of the set $\{1, 2, \ldots, n\}$ (or equivalently of the state space) satisfying the following property.

For any j and l in $\{1, 2, \ldots, m\}$ there exist a number B_{jl} such that

$$B_{jl} = \sum_{i \in L_j} A_{ir}, \quad \text{for } \forall r \in L_l,$$

that is the sum does not depend on r whenever $r \in L_l$. The $m \times m$ matrix B is called the lumping of matrix A.

In simple terms, lumping is equivalent to defining new variables based on additive combinations of the original variables. For example, for the SIS-type dynamics on a complete graph, the new variables in the lumped system could be defined as $Y_k = X_1^k + X_2^k + \cdots + X_{c_k}^k$ for $\forall k = 0, 1, \ldots, N$. In general, the condition in the Definition above amounts to requiring that the sum of the rates of all possible transitions from any individual element of any newly defined variable to all individual elements (rates of transmission to some elements might be zero) of another specified new variable should be the same. To formalise lumping and to derive the precise grouping of the original variables, as given by the partition of the state space, the following two useful propositions are given.

Proposition 2 *If matrix B is the lumping of matrix A, then there exists an $m \times n$ matrix T, such that $TA = BT$.*

Proof Let us define matrix T by $T_{ji} = 1$ when $i \in L_j$, and let all the other entries of T be zero. Then the element in the jth row and rth column of TA is

$$(TA)_{jr} = \sum_{i=1}^{n} T_{ji} A_{ir} = \sum_{i \in L_j} A_{ir} = B_{jl},$$

where l is the index for which $r \in L_l$.

The element in the jth row and rth column of BT is

$$(BT)_{jr} = \sum_{k=1}^{m} B_{jk} T_{kr} = B_{jl},$$

where l is the index such that $r \in L_l$ given that in every column of T there is a single entry which is 1, the others are zero.   □

Proposition 3 *Let B be the lumping of matrix A and let T be the matrix for which $TA = BT$ holds. Based on this, we introduce the new, m dimensional (lumped) variable $x = TX$. This lumped variable satisfies the lumped linear ODE system $\dot{x} = Bx$.*

Proof

$$\dot{x} = T\dot{X} = TAX = BTX = Bx.$$

□

It is worth noting that the crucial ingredient for lumping is the partition of the state space. This can either be derived intuitively or attempting to search for partitions that satisfy the property stated in the Definition of lumping. Once the partition is known both B and T follow. The lumpability condition on P is a non-trivial requirement and not all systems will be lumpable.

4.2 The use of graph automorphism to lump the SIS network model

In the previous section the lumping procedure for a general linear ODE was defined and detailed. We now consider the linear system given by Eq. 1, where matrix P has a special block tridiagonal structure. In the previous general case, the state space was given as the set $\{1, 2, \ldots, n\}$. For the epidemic model, the state space has 2^N elements but instead of the set $\{1, 2, \ldots, 2^N\}$, we denote the state space by $\mathcal{S}$ as given in Sect. 2. The reason for this notation is motivated by the special structure of the state space $\mathcal{S} = \cup \mathcal{S}^k$ which is worth conserving throughout the lumping process. This is essential when using the solution of the lumped system $\dot{x} = Bx$ to determine the expected values $[S](t)$ and $[I](t)$. There are choices of lumping such that the lumped system $\dot{x} = Bx$ cannot be used to determine $[S](t)$ and $[I](t)$. For example, the trivial lumping, when T is a row matrix full of ones, is not a useful lumping. In this case, all variables are lumped into one single variable and the lumping of matrix A (i.e. B) will be a single row and single column matrix with a single zero entry. This just shows that the system is closed and that at any one point in time, the probabilities of being in any of the specified 2^N states sum to one. In order to use the lumped system to determine $[S](t)$ and $[I](t)$, the states with different number of I nodes cannot be lumped together. Hence we will use the following more restrictive definition of lumping.

Definition 2 The linear system given by Eq. 1 (or equivalently the corresponding Markov-process) is called lumpable, if there is a partition $\{L_1, L_2, \ldots, L_m\}$ of the state space $\mathcal{S}$ satisfying the following two properties.

1. For any l there exist k, such that $L_l \subseteq \mathcal{S}^k$.
2. For any j and l there exist a number Q_{jl} such that

$$Q_{jl} = \sum_{i \in L_j} P_{ir}, \quad \text{for } \forall r \in L_l, \tag{22}$$

that is the sum does not depend on r whenever $r \in L_l$.

The $m \times m$ matrix Q is called the lumping of matrix P.

We note that in Part 1 of the Definition, the state space is considered to be $\mathcal{S}$, however in Part 2, the state space is given by the set $\{1, 2, \ldots, 2^N\}$. In order to avoid this ambivalence we will introduce the following more detailed notation of the lumping partition or classes. The lumping classes that are subsets of $\mathcal{S}^k$ will be denoted by $L_1^k, L_2^k, \ldots, L_{l_k}^k$ with the condition that $\mathcal{S}^k = L_1^k \cup L_2^k \cup \cdots \cup L_{l_k}^k$, and the total number of lumping classes is $m = l_0 + l_1 + \cdots + l_N$. It is worth noting that for any

arbitrary lumping $l_0 = l_N = 1$ since the classes $\mathcal{S}^0$ and $\mathcal{S}^N$ contain a single state or element. Using this new notation for the lumping classes, the first assumption of the Definition is automatically fulfilled and the second assumption can be expressed in terms of matrices A^k and C^k as follows.

Lemma 3 *The linear system given by Eq.* 1 *is lumpable, if for* $\forall k \in \{0, 1, 2, \ldots, N\}$ *there is a partition* $L_1^k, L_2^k, \ldots, L_{l_k}^k$ *of the set* $\mathcal{S}^k$ *satisfying the following properties.*

For any $p \in \{1, 2, \ldots, l_{k-1}\}$ *and* $r \in \{1, 2, \ldots, l_k\}$ *there exist a number* $\overline{A}_{rp}^k$ *such that*

$$\overline{A}_{rp}^k = \sum_{\mathcal{S}_i^k \in L_r^k} A_{ij}^k, \quad for \; \forall \mathcal{S}_j^{k-1} \in L_p^{k-1}, \tag{23}$$

that is the sum does not depend on j *whenever* $\mathcal{S}_j^{k-1} \in L_p^{k-1}$.

For any $p \in \{1, 2, \ldots, l_{k+1}\}$ *and* $r \in \{1, 2, \ldots, l_k\}$ *there exist a number* $\overline{C}_{rp}^k$ *such that*

$$\overline{C}_{rp}^k = \sum_{\mathcal{S}_i^k \in L_r^k} C_{ij}^k, \quad for \; \forall \mathcal{S}_j^{k+1} \in L_p^{k+1}, \tag{24}$$

that is the sum does not depend on j *whenever* $\mathcal{S}_j^{k+1} \in L_p^{k+1}$.

Proof As mentioned above we only have to prove that the second condition in the definition of lumping is fulfilled. Let us take two lumping classes L_j and L_l. According to the new ordering of the lumping classes, there exists $k \in \{0, 1, 2, \ldots, N\}$ and $r \in \{1, 2, \ldots, l_k\}$ such that $L_j = L_r^k$ and there exist $h \in \{0, 1, 2, \ldots, N\}$ and $p \in \{1, 2, \ldots, l_h\}$ such that $L_l = L_p^h$. Because of the tridiagonal structure of matrix P and based on Eq. 1, there are four possible cases: $h = k - 1, h = k, h = k + 1$ or h has a different value. For all four cases, we will separately prove that Eq. 22, given in the definition of lumping, holds. In the case of $h = k - 1$, the part of matrix P corresponding to the index set in Eq. 22 is part of matrix A^k, hence Eq. 22 is equivalent to Eq. 23. In the case of $h = k + 1$ the part of matrix P corresponding to the index set in Eq. 22 is part of matrix C^k, hence Eq. 22 is equivalent to Eq. 24. In the case of $h = k$ the part of matrix P corresponding to the index set in Eq. 22 is part of matrix B^k, hence Eq. 22 is equivalent to the following. For any $p \in \{1, 2, \ldots, l_k\}$ and $r \in \{1, 2, \ldots, l_k\}$ there exist a number $\overline{B}_{rp}^k$ such that

$$\overline{B}_{rp}^k = \sum_{\mathcal{S}_i^k \in L_r^k} B_{ij}^k, \quad for \; \forall \mathcal{S}_j^k \in L_p^k. \tag{25}$$

This equation is a consequence of Eqs. 23 and 24, and taking into consideration Eq. 11 (i.e. the sum of every column of P is zero). Finally, if $h \neq k - 1, h \neq k$ and $h \neq k + 1$, then the part of matrix P corresponding to the index set in Eq. 22 has only zero entries and hence, Eq. 22 obviously holds. It is worth noting that the analysis above is based

on considering all transitions that takes the system to a state that belongs to $\mathcal{S}^k$ (i.e. through an infection from $\mathcal{S}^{k-1}$, through a recovery from $\mathcal{S}^{k+1}$, nothing happening and zero rate transitions from $\mathcal{S}^l$, when $l \in \{0, 1, 2, \ldots, N\} \backslash \{k - 1, k, k + 1\}$).   □

Before formulating one of our main results, we recall the Definition of a graph automorphism and the automorphism group of the graph.

Definition 3 Let $G = G(V, E)$ be a graph with vertices and edges given by sets $V(G)$ and $E(G)$ respectively. A bijection $\Phi : V(G) \rightarrow V(G)$ such that $(x, y) \in E(G)$ if and only if $(\Phi(x), \Phi(y)) \in E(G)$ is an automorphism of graph G. The set of all automorphisms of G, under the composition of maps, forms the automorphism group denoted by $Aut(G)$ (Yap 1986).

For graphs with labelled vertices (i.e. nodes or vertices labelled with S or I), a graph automorphism Φ needs to also conserve the labelling of the nodes in the sense that for $\forall x, y \in V(G)$ it must be that the state of nodes x and $\Phi(x)$, and y and $\Phi(y)$ are the same. We will say that the automorphism Φ takes the state $\mathcal{S}_i^k$ to the state $\mathcal{S}_j^k$ if $\mathcal{S}_i^k(l) = \mathcal{S}_j^k(\Phi(l))$. Now, we can formulate our main results that connects the automorphism group of the graph to the lumping of the Markov chain.

Theorem 2 *Let us introduce the following equivalence relation in the state space $\mathcal{S}$. The states $\mathcal{S}_i^k$ and $\mathcal{S}_j^k$ are equivalent if there is an automorphism of the graph that takes $\mathcal{S}_i^k$ to $\mathcal{S}_j^k$. Then the classes of this equivalence relation yield a lumping of the system defined by Eq. 1.*

Proof Let us denote the lumping classes as in Lemma 3. We will prove that Eq. 23 holds. The proof of Eq. 24 is similar, hence it is not detailed here. Let $k \in \{0, 1, 2, \ldots, N\}$, $p \in \{1, 2, \ldots, l_{k-1}\}$ and $r \in \{1, 2, \ldots, l_k\}$ be arbitrary numbers. Let $\mathcal{S}_{i_1}^k \in L_r^k$, $\mathcal{S}_{j_1}^{k-1} \in L_p^{k-1}$ and let $z = A_{i_1 j_1}^k \neq 0$. This means that the states $\mathcal{S}_{i_1}^k$ and $\mathcal{S}_{j_1}^{k-1}$ differ only at one node, say at node numbered by u, and $\mathcal{S}_{i_1}^k(u) = I$, $\mathcal{S}_{j_1}^{k-1}(u) = S$. Moreover, according to Eq. 2 the number of I nodes in the state $\mathcal{S}_{j_1}^{k-1}$ that are connected to the node numbered by u is equal to z/τ (see the definition of matrix A). Now let us consider another element in the lumping class L_p^{k-1}, that is let $\mathcal{S}_{j_2}^{k-1} \in L_p^{k-1}$. Then there is an automorphism of the graph that takes the state $\mathcal{S}_{j_1}^{k-1}$ into the state $\mathcal{S}_{j_2}^{k-1}$. Let us apply this automorphism to the state $\mathcal{S}_{i_1}^k$. Then we get the state $\mathcal{S}_{i_2}^k \in L_r^k$. This means that the relation between the states $\mathcal{S}_{i_2}^k$ and $\mathcal{S}_{j_2}^{k-1}$ is the same as the relation of the states $\mathcal{S}_{i_1}^k$ and $\mathcal{S}_{j_1}^{k-1}$. Hence we have $A_{i_2 j_2}^k = z$. Thus, if there is a non zero element z in the j_1th column of the matrix A^k in a row belonging to the lumping class L_r^k, then this element z appears also in the j_2th column of the matrix A^k in a (possibly different) row belonging to the same lumping class L_r^k, which proves Eq. 23.   □

This Theorem proves that graph automorphisms can be used to identify states that can be lumped together. This can be formalised further using group theory arguments. In the case when the automorphism group (i.e. $Aut(G)$) is known, finding the lumping classes is equivalent to working out the orbits of the elements of the state space with

respect to the group $Aut(G)$. For example, the orbit of an element $\mathcal{S}_j^k \in \mathcal{S}^k$ is given by the set $Aut(G)(\mathcal{S}_j^k) = \{\Phi(\mathcal{S}_j^k) : \Phi \in Aut(G)\} \subseteq \mathcal{S}^k$. The orbit of the element will contain all the individual states that can be lumped together and thus defines a lumping class. Finding further lumping classes can be done by working out the orbits of elements in $\mathcal{S}_k$ that are not yet part of any previously computed orbits. The number of elements in an orbit is bounded from above by the number of elements in the automorphism group (i.e. $|Aut(G)|$). Thus, the number of elements in a lumping class is also bounded from above by $|Aut(G)|$ which in turn means that the number of lumping classes is bounded from below by $\frac{2^N}{|Aut(G)|}$. It is worth noting that this is a weak lower bound and it is only meaningful when $2^N > |Aut(G)|$ (this is not the case for a completely connected graph). When the inequality holds, the lower bound can serve as an indicator of the effectiveness of lumping in reducing the dimensionality of the original exact system.

4.3 Lumping for some classes of graphs

In this section we show some applications of the above theorem and show how the lumping can be carried out when the graph has some special structure.

4.3.1 Lumping for the complete graph

First we show that for a complete graph the 2^N-dimensional system given by Eq. 1 can be lumped to an $N+1$-dimensional system. The lumped system is well-known in the literature, however, according to our knowledge, its derivation from the full 2^N-dimensional system is not available.

The automorphism group of the complete graph is the permutation group $\mathbf{S_N}$, hence there is an automorphism between any two states in $\mathcal{S}^l$. Hence, the orbit of any element from $\mathcal{S}^l$ is the same and is equal to $\mathcal{S}^l$ itself. Therefore all the states with l infected nodes can be lumped together. This means that there are $N+1$ lumping classes: $L_l = \mathcal{S}^l$ for $\forall l \in \{0, 1, \ldots, N\}$. That is we have $m = N + 1$ and the lumping matrix T takes the form

$$T = \begin{pmatrix} S_0 & 0 & 0 & 0 & 0 \\ 0 & S_1 & 0 & 0 & 0 \\ 0 & 0 & S_2 & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & S_N \end{pmatrix}$$

where $S_k = (1, 1, \ldots, 1)$ is a vector of length c_k. Thus the new (scalar valued) unknown functions are introduced as

$$x^k = \sum_{j=1}^{c_k} X_j^k = S_k X^k, \quad k = 1, \ldots, N.$$

Lemma 4 *If G is a complete graph, then the x^k functions satisfy the following differential equations.*

$$\dot{x}^0 = \gamma x^1,$$
$$\dot{x}^k = (k+1)\gamma x^{k+1} + (k-1)(N-k+1)\tau x^{k-1} - (k(N-k)\tau + k\gamma)x^k,$$
$$\dot{x}^N = (N-1)\tau x^{N-1} - N\gamma x^N,$$

for $k = 1, 2, \ldots, N - 1$.

Proof The system for X^k in Eq. 1 consists of c_k equations. Adding these equations we obtain

$$\dot{x}^k = S_k A^k X^{k-1} + S_k B^k X^k + S_k C^k X^{k+1}.$$

Now the statement follows from Proposition 1 and from the Proposition below.   □

Proposition 4 1. $S_k A^k = (k-1)(N-k+1)\tau S_{k-1}.$
2. $S_k B^k = -(k(N-k)\tau + k\gamma)S_k.$

Proof 1. In a complete graph for $\forall j \in \{1, 2, \ldots, c_k\}$ we have that $N_{SI}(\mathcal{S}_j^k) = k(N-k)$ since every node in state I is connected to every other node in state S. Hence according to Eq. 3, for $\forall j \in \{1, 2, \ldots, c_{k-1}\}$ the following relation holds

$$(S_k A^k)_j = \sum_{i=1}^{c_k} A_{i,j}^k = \tau N_{SI}(\mathcal{S}_j^{k-1}) = (k-1)(N-k+1)\tau.$$

This proves that $S_k A^k = (k-1)(N-k+1)\tau S_{k-1}$ since the jth coordinate of the left and right-hand sides are equal.
2. The second statement follows from Proposition 1 and from Eq. 11.   □

In the Appendix, the lumping process is derived and explained for a complete graph with $N = 3$ nodes.

4.3.2 Lumping for the star graph

Let us consider a star like graph with N nodes. In this graph a single central node is connected to all other nodes with no further connections. Let the Nth node be the center of the star. Thus, for example $SS \ldots SI$ denotes the state when the central node is infected and the other nodes are susceptible. We will show that in the case of a star graph, the 2^N-dimensional system as defined by Eq. 1 can be lumped to a $2N$-dimensional system.

The automorphism group of the star graph is the permutation group $\mathbf{S_{N-1}}$, since an automorphism leaves the central node unchanged and it can permutate the remaining $N - 1$ nodes in an arbitrary way. However, the labelling of the nodes have to be taken into account. Therefore two states can be taken into each other by an automorphism if and only if the center is in the same state and the number of I's among the non-central

nodes is the same. Hence, for $l = 1, 2, \ldots, N-1$ the set $\mathcal{S}^l$ of states of the graph, such that there are l nodes in state I, can be lumped into two classes: in the first class the central node is I and there are $l-1$ non-central I nodes, in the second class the central node is S and there are l non-central I nodes. In the case of $l = 0$ and $l = N$ there is obviously only one class. This means that altogether there are $2N = 2(N-1)+2$ lumping classes: $L_1 = \mathcal{S}^0$, $L_{2N} = \mathcal{S}^N$, $L_{2k} = \{\mathcal{S}_j^k : \mathcal{S}_j^k(N) = I\}$, $L_{2k+1} = \{\mathcal{S}_j^k : \mathcal{S}_j^k(N) = S\}$ for all $k \in \{1, 2, \ldots, N-1\}$.

Thus the lumped variables can be introduced as

$$x^0 = X^0, \quad x^N = X^N, \quad x_1^k = \sum_{\mathcal{S}_j^k(N)=I} X_j^k, \quad x_2^k = \sum_{\mathcal{S}_j^k(N)=S} X_j^k,$$

for $k = 1, \ldots, N-1$. Similarly to Lemma 4, we can prove the following Lemma,

Lemma 5 *If G is a star graph, then the x_i^k functions satisfy the following differential equations.*

$$\dot{x}^0 = \gamma(x_1^1 + x_2^1),$$
$$\dot{x}_1^k = (N-k+1)\tau x_1^{k-1} + (k-1)\tau x_2^{k-1} - (k\gamma + (N-k)\tau)x_1^k + k\gamma x_1^{k+1},$$
$$\dot{x}_2^k = -k(\gamma + \tau)x_2^k + \gamma(x_1^{k+1} + (k+1)x_2^{k+1}),$$
$$\dot{x}^N = \tau x_1^{N-1} + (N-1)\tau x_2^{N-1} - N\gamma x^N,$$

for $k = 1, 2, \ldots, N-1$, where the following notations are used: $x_1^0 = 0, x_2^0 = x^0, x_1^N = x^N, x_2^N = 0$.

4.3.3 Lumping for the household graph

Let us consider the simplest graph with a so-called household structure. This graph consists of two types of nodes, inner and outer nodes. Outer nodes have only within household connections while inner nodes have within household as well as connections to other households. We consider the simplest case where each household has two nodes, an inner and outer node. The inner nodes of all households form a complete graph with $N/2$ nodes (N is an even number), and every outer node is connected to an inner node. Thus the degree of all inner nodes is $\frac{N}{2}$ and the degree of all outer nodes is 1. We will show that in the case of this household-type graph the 2^N-dimensional system given by Eq. 1 can be lumped to an $\binom{N/2+3}{3}$-dimensional system.

The automorphism group of this graph is the permutation group $\mathbf{S_{N/2}}$, since an automorphism can permute the inner nodes in an arbitrary way, and once the automorphism is given on the inner nodes, its effect on the outer nodes is determined uniquely. In order to determine the lumping classes let us note first that there may be four different types of households in this graph: SI households, in which the inner node is S and the outer node is I, IS households, SS households and II households. Therefore two states of the whole graph can commute through an automorphism if and only if the number of SI, IS, SS and II type households is the same in the two

distinct states of the complete system. Hence, to obtain all different states we have to choose (with repetition) $N/2$ households out of the four different types. Thus, the number of different states (i.e. states where the number of SS, SI, IS and II-type households is different but their distribution can be arbitrary) can be obtained as the number of combinations with repetitions: $\binom{4+N/2-1}{N/2} = \binom{N/2+3}{3}$. Hence, states with the same number of SS, SI, IS and II-type households can be lumped into one newly defined lumped variable.

4.3.4 Lumping for the cycle graph: non-trivial lumping

For the completely connected and star graph, lumping can also be carried out intuitively. However, lumping based on intuition alone can become unwieldy even for a relatively simple graph such as the cycle graph (C_N) where N nodes are connected in a close chain such that each node connects to the two nearest neighbours only. The automorphisms of the cycle graph can be given in terms of all possible *rotations* and *reflections* of the graph. These together, N of each, give $2N$ automorphisms, and hence $|Aut(C_N)| = 2N$. The automorphism group of the cycle graph is also known as the dihedral group D_N. Carrying out lumping based on intuition alone is prone to error and it is desirable to use the automorphism group to work out lumping classes rigorously. For small graphs, these calculations can be done on paper or for larger graphs calculations can be easily carried out using some suitable software or programming language. To illustrate the process, let us consider the case of a cycle graph with $N = 5$ nodes. Here, $|D_5| = 10$ is made up of five rotations including the identity and five reflections. The lumping classes L_i ($i = 1, 2, \ldots, m$), with m yet to be determined, have to be such that for $\forall i = 1, 2, \ldots, m, \exists k \in \{0, 1, \ldots, N\}$ such that $L_i \subseteq S^k$ (as required by Definition 2). Based on this the first lumping class is trivial $L_1 = \{(SSSSS)\}$. This is followed by considering $S^1 = \{(SSSSI), (SSSIS), (SSISS), (SISSS), (ISSSS)\}$ when it is easy to show that the orbit of the first element $(SSSSI)$ is $Aut(C_n)((SSSSI)) = \{\Phi((SSSSI)) : \Phi \in Aut(C_N)\} = S^1$. Hence $L_2 = S^1$. This is a desirable situation since all elements in S^1 can be lumped together to give a new variable. The situation changes when S^2 is considered

$$S^2 = \{(SSSII), (SSISI), (SISSI), (ISSSI), (SSIIS), (SISIS), (ISSIS),$$
$$(SIISS), (ISISS), (IISSS)\}.$$

The orbit of the first element $(SSSII) \in S^2$ is

$$Aut(C_n)((SSSII)) = \{\Phi((SSSII)) : \Phi \in Aut(C_n)\}$$
$$= \{(SSSII), (ISSSI), (SSIIS), (SIISSS), (IISSS)\} = L_3.$$

In this case the orbit of $(SSSII)$ only captures 5 out of the 10 possible states in S^2. This means that *rotations* and *reflections* map $(SSSII)$ onto identical configurations and this increases the number of lumping classes and the reduction in dimensionality of the system is not so significant. The remaining five elements form a new lumping

class $L_4 = \{(SSISI), (SISSI), (SISIS), (ISSIS), (ISISS)\}$. Based on similar argument four more lumping classes can be identified. This means that the original system with $2^5 = 32$ equations can be reduced to a system with only 8 equations. Using similar arguments, it is easy to show that for $N = 6$, $N = 7$ and $N = 8$ the exact system can be lumped from 64, 128 and 256 to 13, 18 and 30 equations, respectively. Given that for the cycle graph $2^N > |Aut(G)| = 2N$, the argument presented in Sect. 4.2 can be used to show that the number of lumping classes for the cycle graph is bounded from below by $\frac{2^N}{2N} = 2^{N-1}(1/N)$. This indicates that the number of equations in the lumped system is much larger than polynomial in N.
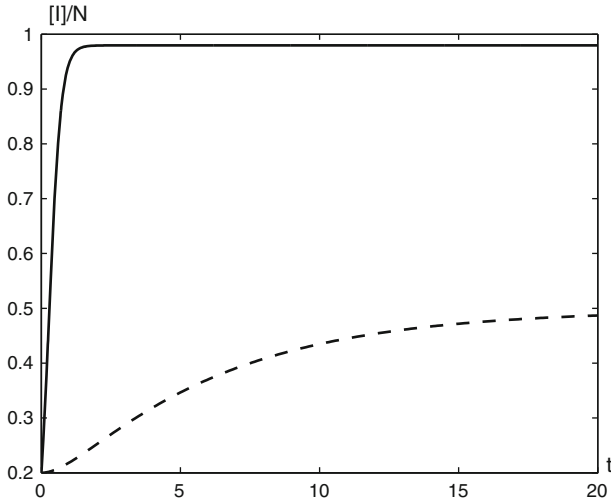
5 Discussion

In this paper, using a continuous time Markov chain, we have formulated an *SIS* type epidemic model on an arbitrary graph and discussed various approximation methods that allow us to reduce the number of equations from $2^N - 1$ to a feasible number that can either be solved analytically and/or numerically. Exploring the special block tridiagonal nature of the transition matrix P, we have rigorously derived mean-field approximations that previously have only been derived heuristically based on biological hypotheses. The key result of the paper is the precise formulation of lumping in terms of graph automorphisms and their group. Lumping is a powerful method that can be used to reduce large linear ODE models to a tractable number of equations which are still exact and can be used to evaluate the time evolution of quantities of interest as required by the specific modelling context. For example, Table 1 illustrates that for graphs with a high number of automorphisms, the exact equations given by Eq. 1 can be more effectively lumped into fewer equations that are easier to analyse.

For the complete and star graph, the time evolution of $[i] = [I]/N$ is illustrated in Fig. 1. This plot is based on the lumped system which is used to compute the expected level of prevalence over time. This is in contrast with many such results which are either based on some form of mean-field equations or individual-based network simulations. Figure 1 shows that the complete graph allows the infection to spread very quickly to a high proportion of nodes. For the star graph, the spread is much slower and only a much smaller proportion of the network will become infected. Here, the difference between prevalence levels can be mainly explained by the marked difference in average connectivity, $\frac{N+2}{4}$ for the complete graph versus $2 - 2/N$ for the star graph. The use of lumping enabled us to compare the disease spread on different graphs and to investigate the effect of the structure of the graph on the spread of the infection. Based on mean-field approximations, equations for the prevalence level can be derived. For the complete and star graph, these read as follows:

$$\lim_{t \to \infty} \tilde{i}_{comp}(t) = 1 - \frac{\gamma}{N\tau},$$

$$\lim_{t \to \infty} \tilde{i}_{star}(t) = 1 - \frac{\gamma}{\tau + \gamma} + \frac{\gamma}{(N-1)\tau + \gamma} - \frac{2\gamma}{N} \left(\frac{\gamma}{\tau(\tau + \gamma)} + \frac{1}{(N-1)\tau + \gamma} \right).$$

Fig. 1 Prevalence of infection as a function of time using the lumped equations for a complete (*solid line*) and a star (*dashed line*) graph with $N = 50$, $\tau = 0.1$ and $\gamma = 0.1$
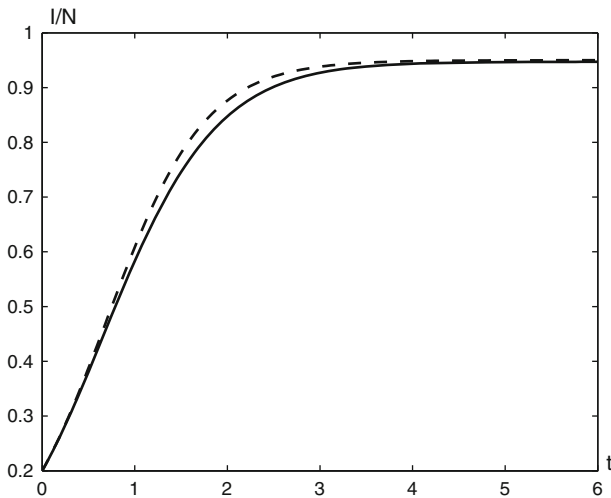
Fig. 2 Prevalence of infection as a function of time for a complete graph using the lumped equations ($[i](t)$ *solid line*) and the mean-field model ($\tilde{i}(t)$ *dashed line*). Here, $N = 20$, $\tau = 0.1$ and $\gamma = 0.1$

It is worth noting that even for a small graph with $N = 50$, the numerical values of the exact solution $[i](t)$ and that of the mean-field approximation $\tilde{i}(t)$ are close to each other. In the case of a complete graph, the quasi steady state of $[i]$ is 0.9796 and the steady state of $\tilde{i}$ is $\lim_{t \to \infty} \tilde{i}_{comp}(t) = 1 - \frac{\gamma}{N\tau} = 0.9800$, when $\gamma = 0.1$ and $\tau = 0.1$. In the case of a star graph with the same parameters, these values are 0.4967 versus 0.4992. In Fig. 2, we show that even for small networks $N = 20$, results from the mean-field and exact models agree well, with a small further increase in N making the

Table 1 Summary of lumping on special graphs

Graph type	Aut(G)	# of eq. in the lumped system
Complete	$\mathbf{S_N}$	$N+1$
Star	$\mathbf{S_{N-1}}$	$2N$
Household	$\mathbf{S_{N/2}}$	$\binom{N/2+3}{3}$

two almost indistinguishable. This is somewhat in contrast with the general belief that mean-field and pair-approximation models become exact in the limit of large graphs.

Due to their great flexibility network models are used widely in many different areas ranging from neuro-networks to evolutionary dynamics. For example, Broom and Rychtář (2008) have recently considered the fixation probability of a mutant on special classes of non-directed graphs and have shown that the total number of mutant-resident formations is proportional to the inverse of the cardinality of $Aut(G)$ (i.e. $|Aut(G)|$). In this paper, in line with results derived by Broom and Rychtář (2008), we have also illustrated how the richness of the automorphism group impacts on the reduction in dimensionality of the exact system (see Table 1). For graphs with a high number of automorphisms, such as completely connected or star graphs, the 2^N equations of the exact system can be reduced to $O(N)$. It is worth noting that for a more complex dynamics such as an SIR model, the number of equations in the reduced system on a completely connected graph will be of $O(N^2)$. While for graphs with less symmetry (i.e. smaller number of automorphisms) the reduction in dimensionality will not be as significant, the use of automorphism-driven lumping still provides a rigorous way to identify the lumping classes. Even though constructing the automorphism group of an arbitrary graph is a challenging problem and belongs to the class *NP* of computational complexity (Brandes and Erlebach 2005), there are many special classes of graphs which are both relevant from an application point of view and possess a computable or known automorphism group. More importantly exact models for small populations such as households, hospital wards, schools, workplaces and small holding farms can play an important role in accounting correctly for stochastic effects that are more important in small populations. Therefore, exploiting the structure of the graph is key when formulating models where the derivation of analytical results is needed to strengthen or confirm simulation results.

It is not often that mean-field and pair-approximation models are compared directly to the solution of the exact system (Keeling and Ross 2008) and, at least for small or large graphs with special structure, graph automorphism driven lumping can be used to rigorously test the performance of various approximation models. These models can both simplify calculations and underpin the derivation of some general theoretical results. Hence, improving the testing and validation of such models is an important part of the modelling exercise in many areas such as epidemiology, ecology and other areas where network models are used. This study also links well-known and well-studied problems in graph theory (Gross and Yellen 2003; Yap 1986) (e.g. graph automorphisms and their group) to real-life applications and we foresee that other results from graph theory will become applicable in the context of network model formulation and analysis.

6 Appendix

6.1 Kolmogorov equations in the case of a complete graph with $N = 3$

The aim is to derive the full set of Kolmogorov differential equations using the special block tridiagonal matrix form introduced in Sect. 2.2.1. This small graph has a state space with 2^3 elements, and following previously introduced notation, this can divided in the following subsets, $X^0 = X_{SSS}$, $X^1 = (X_{SSI}, X_{SIS}, X_{ISS})$, $X^2 = (X_{SII}, X_{ISI}, X_{IIS})$, $X^3 = X_{III}$ with $X = (X^0, X^1, X^2, X^3)$. For $N = 3$, P is given by

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

Taking into account the structure of the network (i.e. each node connected to every other node) and using the bookkeeping rules presented in Sect. 2.2.1, the entries of matrix P are given by

$$B^0 = (0), \quad C^0 = (\gamma, \gamma, \gamma),$$

$$A^1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad B^1 = \begin{pmatrix} -2\tau - \gamma & 0 & 0 \\ 0 & -2\tau - \gamma & 0 \\ 0 & 0 & -2\tau - \gamma \end{pmatrix}, \quad C^1 = \begin{pmatrix} \gamma & \gamma & 0 \\ \gamma & 0 & \gamma \\ 0 & \gamma & \gamma \end{pmatrix},$$

$$A^2 = \begin{pmatrix} \tau & \tau & 0 \\ \tau & 0 & \tau \\ 0 & \tau & \tau \end{pmatrix}, \quad B^2 = \begin{pmatrix} -2\tau - 2\gamma & 0 & 0 \\ 0 & -2\tau - 2\gamma & 0 \\ 0 & 0 & -2\tau - 2\gamma \end{pmatrix}, \quad C^2 = \begin{pmatrix} \gamma \\ \gamma \\ \gamma \end{pmatrix},$$

$$A^3 = (2\tau, 2\tau, 2\tau), \quad B^3 = (-3\gamma).$$

For example, here the entries in the first column of matrix A^2 (τ, τ, 0) correspond to the rates of the following transitions, $SSI \rightarrow SII$, $SSI \rightarrow ISI$ and $SSI \rightarrow IIS$. Given that our graph is a completely connected triangle with starting state SSI, the transition through an infection always happens at rate τ since there is only one source of infection. Based on the definition of the sub-matrices above, $\dot{X} = PX$ is equivalent to the following system

$$\dot{X}^0 = B^0 X^0 + C^0 X^1,$$
$$\dot{X}^1 = A^1 X^0 + B^1 X^1 + C^1 X^2,$$
$$\dot{X}^2 = A^2 X^1 + B^2 X^2 + C^2 X^3,$$
$$\dot{X}^3 = A^3 X^2 + B^3 X^3.$$

In terms of the most intuitive notation, this system is equivalent to

$$\dot{X}_{SSS} = \gamma(X_{SSI} + X_{SIS} + X_{ISS}),$$
$$\dot{X}_{SSI} = \gamma(X_{SII} + X_{ISI}) - (2\tau + \gamma)X_{SSI},$$
$$\dot{X}_{SIS} = \gamma(X_{SII} + X_{IIS}) - (2\tau + \gamma)X_{SIS},$$
$$\dot{X}_{ISS} = \gamma(X_{ISI} + X_{IIS}) - (2\tau + \gamma)X_{ISS},$$
$$\dot{X}_{SII} = \gamma X_{III} + \tau(X_{SSI} + X_{SIS}) - 2(\tau + \gamma)X_{SII},$$
$$\dot{X}_{ISI} = \gamma X_{III} + \tau(X_{SSI} + X_{ISS}) - 2(\tau + \gamma)X_{ISI},$$
$$\dot{X}_{IIS} = \gamma X_{III} + \tau(X_{SIS} + X_{ISS}) - 2(\tau + \gamma)X_{IIS},$$
$$\dot{X}_{III} = -3\gamma X_{III} + 2\tau(X_{SII} + X_{ISI} + X_{IIS}),$$

and from either formulation, the transition matrix is

$$P = \begin{pmatrix} 0 & \gamma & \gamma & \gamma & 0 & 0 & 0 & 0 \\ 0 & -2\tau - \gamma & 0 & 0 & \gamma & \gamma & 0 & 0 \\ 0 & 0 & -2\tau - \gamma & 0 & \gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & -2\tau - \gamma & 0 & \gamma & \gamma & 0 \\ 0 & \tau & \tau & 0 & -2\tau - 2\gamma & 0 & 0 & \gamma \\ 0 & \tau & 0 & \tau & 0 & -2\tau - 2\gamma & 0 & \gamma \\ 0 & 0 & \tau & \tau & 0 & 0 & -2\tau - 2\gamma & \gamma \\ 0 & 0 & 0 & 0 & 2\tau & 2\tau & 2\tau & -3\gamma \end{pmatrix}.$$

6.2 Lumping in the case of a complete graph with $N = 3$

Here, we show how lumping is carried out in the case of a complete graph with three nodes ($N = 3$). In this case, the differential equations for X^0, X^1, X^2, X^3 are given above. Let us introduce the new (scalar valued) unknown functions

$$x^0 = X^0, \quad x^1 = X_1^1 + X_2^1 + X_3^1, \quad x^2 = X_1^2 + X_2^2 + X_3^2, \quad x^3 = X^3.$$

By adding the differential equations corresponding to X_1^1, X_2^1, X_3^1, and those corresponding to X_1^2, X_2^2, X_3^2, we obtain the following four dimensional system for x^0, x^1, x^2, x^3:

$$\dot{x}^0 = \gamma x^1,$$
$$\dot{x}^1 = 2\gamma x^2 - (2\tau + \gamma)x^1,$$
$$\dot{x}^2 = 3\gamma x^3 + 2\tau x^1 - 2(\tau + \gamma)x^2,$$
$$\dot{x}^3 = 2\tau x^2 - 3\gamma x^3.$$

It is easy to see that matrix Q, as defined in Eq. 22, is given by

$$Q = \begin{pmatrix} 0 & \gamma & 0 & 0 \\ 0 & -2\tau - \gamma & 2\gamma & 0 \\ 0 & 2\tau & -2(\tau + \gamma) & 3\gamma \\ 0 & 0 & 2\tau & -3\gamma \end{pmatrix},$$

with entries based on the sum of columns in matrices A^k, B^k and C^k as they appear in the original matrix P. For example, if only the $x^1 \rightarrow x^2$ transition is considered, the original differential equations reduce to

$$\dot{X}^2 = \begin{pmatrix} \dot{X}_1^2 \\ \dot{X}_2^2 \\ \dot{X}_3^2 \end{pmatrix} = \begin{pmatrix} \tau & \tau & 0 \\ \tau & 0 & \tau \\ 0 & \tau & \tau \end{pmatrix} \begin{pmatrix} X_1^1 \\ X_2^1 \\ X_3^1 \end{pmatrix} = A^2 X^1.$$

If all the equations above are added, we obtain

$$\dot{x}^2 = (\dot{X}_1^2 + \dot{X}_2^2 + \dot{X}_3^2) = 2\tau(X_1^1 + X_2^1 + X_3^1) = 2\tau x^1,$$

where the 2τ represents the sum of each individual column in the A^2 matrix. The lumping condition requires that each column in these matrices sum to the same value and this allows us to lump groups of the original variables into newly defined variables and create a new system with fewer equations. Upon lumping, the original 8-dimensional system can be reduced to a four-dimensional one, and using this, the desired expected values $[S](t)$ and $[I](t)$ can still be determined as

$$[I](t) = x^1(t) + 2x^2(t) + 3x^3(t), \qquad [S](t) = 3x^0(t) + 2x^1(t) + x^2(t).$$

6.3 Proof of Theorem 1

We have seen that for a complete graph we can introduce $x^k(t)$ as the total probability of states with k infected nodes at time t, and these functions satisfy the following system

$$\dot{x}^k = (k+1)\gamma x^{k+1} + (k-1)(N-k+1)\tau x^{k-1} - (k(N-k)\tau + k\gamma)x^k, \quad (26)$$

for $k = 0, \ldots, N$, with $x^{-1} = 0$ and $x^{N+1} = 0$. The scaled expected value of susceptible and infected nodes are

$$[s](t) = \frac{1}{N} \sum_{k=0}^{N} (N-k)x^k(t), \qquad [i](t) = \frac{1}{N} \sum_{k=0}^{N} kx^k(t). \qquad (27)$$

Our aim here is to investigate the limit of large N and compare these to the solution of the scaled mean-field equations as given by Eqs. 15–16.

The idea of comparison for large N is to introduce a continuous, time dependent density function $\rho(t, z)$ instead of the discrete distribution $x^k(t)$, with the following formal relation, $z = k/N$. Following this, in Eq. 26 we can formally change $\dot{x}^k$ to $\partial_t \rho(t, z)$, $x^k(t)$ to $\rho(t, z)$, $x^{k-1}(t)$ to $\rho(t, z - 1/N)$ and $x^{k+1}(t)$ to $\rho(t, z + 1/N)$. This leads to the following partial differential equation

$$\partial_t \rho(t, z) = (Nz + 1)\gamma \rho(t, z + 1/N) + (Nz - 1)(N - Nz + 1)\tau \rho(t, z - 1/N)$$
$$-(Nz(N - Nz)\tau + Nz\gamma)\rho(t, z).$$

Now using the approximations

$$\rho(t, z + 1/N) = \rho(t, z) + \partial_z \rho(t, z)/N, \qquad \rho(t, z - 1/N) = \rho(t, z) - \partial_z \rho(t, z)/N,$$

and after some algebra we obtain

$$\partial_t \rho(t, z) = (Nz + 1)\gamma \partial_z \rho(t, z)/N + (2Nz - N - 1)\tau \rho(t, z)$$
$$-(Nz - 1)(N - Nz + 1)\tau \partial_z \rho(t, z)/N + \gamma \rho(t, z).$$

Substituting $\tau = \beta/N$, neglecting the $1/N$ and $1/N^2$ terms and writing ρ instead of $\rho(t, z)$, we obtain the following first order partial differential equation for ρ

$$\partial_t \rho = z\gamma \partial_z \rho + (2z - 1)\beta \rho - z(1 - z)\beta \partial_z \rho + \gamma \rho.$$

Introducing the function $g(z) = \gamma z - \beta z(1 - z)$, the equation for ρ becomes

$$\partial_t \rho = \partial_z(g\rho). \tag{28}$$

This first order partial differential equation needs an initial condition of the following type

$$\rho(0, z) = \rho_0(z). \tag{29}$$

The desired initial condition can be obtained from the initial condition of Eq. 26. This latter initial condition can be written as

$$x^m(0) = 1, \quad \text{for some } m, \quad x^l(0) = 0, \quad \text{for } l \neq m, \tag{30}$$

that is at the initial instant there are m infected nodes. Since the formal relation between the variables is $z = k/N$, the above initial condition yields

$$\rho_0(z) = 1 \quad \text{for } \frac{m}{N} < z < \frac{m+1}{N} \quad \text{and} \quad \rho_0(z) = 0 \quad \text{otherwise.}$$

Finally, we want to determine the expected value of the infected and susceptible nodes from the first order PDE. Thus we have to find the functions corresponding to $[s](t)$

and $[i](t)$ in Eq. 27. Using $z = k/N$ and changing the term $x^k(t)$ to $\rho(t, z)$, we note that the sums in Eq. 27 correspond to integrals. Namely, $[i](t)$ corresponds to

$$N \sum_{k=0}^{N} \frac{k}{N} \rho\left(t, \frac{k}{N}\right) \frac{1}{N},$$

and this sum is an approximation of the integral

$$N \int_{0}^{1} z\rho(t, z)\mathrm{d}z.$$

Noticing that $\int_0^1 \rho_0(z)\mathrm{d}z = 1/N$, we can introduce $i^*(t)$ as a function corresponding to $[i](t)$ as follows

$$i^*(t) = \frac{\int_0^1 z\rho(t, z)\mathrm{d}z}{\int_0^1 \rho_0(z)\mathrm{d}z}. \tag{31}$$

The mean-field equation (Eq. 16) can be solved explicitly and the solution is given by

$$\tilde{i}(t) = \frac{B(t)i_0}{\beta - \gamma - A(t)i_0},$$

where $i_0 = \tilde{i}(0)$ is the initial condition and

$$A(t) = \beta - \beta exp((\beta - \gamma)t), \qquad B(t) = (\beta - \gamma)exp((\beta - \gamma)t).$$

The first order PDE (Eq. 28) can also be solved explicitly, and then Eq. 31 yields

$$i^*(t) = \frac{B(t)}{A(t)} \left[1 + \frac{N(\beta - \gamma)}{A(t)} \log\left(-1 + \frac{2A(t)}{2N(\beta - \gamma - A(t)i_0) - A(t)} \right) \right].$$

Having these explicit formulas for $i^*(t)$ and $\tilde{i}(t)$, it is easy to see that i^* is not a solution of the mean-field equation (Eq. 16) but it can be proved that as $N \to \infty$ it tends to the solution of Eq. 16. Namely, we have the following Lemma.

Lemma 6 *Let ρ be the solution of the system given by Eq. 28 with initial condition given by Eq. 29. Let $i^*(t)$ be defined by Eq. 31. Let $\tilde{i}(t)$ be the solution of the scaled mean-field equation given by Eq. 16 with initial condition $\tilde{i}(0) = m/N$. Then for any $t \geq 0$ we have*

$$\lim_{N \to \infty} |\tilde{i}(t) - i^*(t)| = 0.$$

The Lemma can be proved by using the explicit formulas for $i^*(t)$ and $\tilde{i}(t)$.

Now the proof of the Theorem can be concluded as follows. We want to prove that the scaled expected value $[i](t)$ tends to the solution $\tilde{i}(t)$ of the scaled mean-field equation as $N \to \infty$. In order to prove this, we introduced a first order PDE that can be considered the limit of Eq. 26 as $N \to \infty$. Using this PDE, we defined the function $i^*(t)$ that corresponds to $[i](t)$. According to Lemma 6, $i^*(t)$ is close to $\tilde{i}(t)$ for large N. Hence, we only have to show finally that $[i](t)$ is close to $i^*(t)$. Thus the proof of Theorem 1 will be complete if the following Lemma is verified.

Lemma 7 *Let x^k be the solution of Eq.* 26 *satisfying the initial condition given by Eq.* 30*, and let ρ be the solution of Eq.* 28 *with initial condition given by Eq.* 29*. Let $[i](t)$ and $i^*(t)$ be defined by Eq.* 27 *and by* Eq. 31*. Then for any $t \geq 0$ we have*

$$\lim_{N \to \infty} |[i](t) - i^*(t)| = 0.$$

The proof of the Lemma is based on the fact that the lumped system seen in Eq. 26 can be considered as the discretisation of the first order PDE seen in Eq. 28 in the variable z. It is known even for more general PDEs, see Chaps. 3 and 4 in Hundsdorfer and Verwer (2003), that the solution of the discretised system tends to that of the PDE as the step size of the discretisation goes to zero, that is in our case N tends to infinity.

References

Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press

Andersson H, Djehich B (1997) A threshold limit theorem for the stochastic logistic epidemic. J Appl Prob 35:662–670

Ball F, Mollison D, Salia-Tomba G (1997) Epidemics with two levels of mixing. Ann Appl Prob 7:46–89

Brandes U, Erlebach T (2005) Network analysis: methodological foundations. Springer-Verlag, Berlin

Brauer F, van den Driessche P, Wu J (2008) Mathematical epidemiology. In: Lecture notes in mathematics. Springer-Verlag, Berlin

Broom M, Rychtář J (2008) An analysis of the fixation probability of a mutant on special classes of non-directed graphs. Proc R Soc A 464:2609–2627

Daley DJ, Gani J (1999) Epidemic modelling: an intoduction. Cambridge University Press

Dent JE, Kao RR, Kiss IZ, Hyder K, Arnold M (2008) Contact structures in the poultry industry in Great Britain: exploring transmission routes for a potential avian influenza virus epidemic. BMC Vet Res 4:27

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Wiley, Chichester

Diekmann O, De Jong MCM, Metz JAJ (1998) A deterministic epidemic model taking account of repeated contacts between the same individuals. J Appl Prob 35:448–462

Ferguson NM, Donnelly CA, Anderson RM (2001) The foot-and mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science 292:1155–1160

Filliger L, Hongler M-O (2008) Lumping complex networks. Madeira Math Encounters XXXV. http://ccm. uma.pt/mme35/files/LumpingNetworks-Filliger.pdf

Green DM, Kiss IZ, Kao RR (2006) Modelling the initial spread of foot-and-mouth disease through animal movements. Proc R Soc B 273:2729–2735

Gross JL, Yellen J (eds) (2003) Handbook of graph theory. CRC Press, Boca Raton

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

Hufnagel L, Brockmann D, Geisel T (2004) Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci USA 101:15124–15129

Hundsdorfer W, Verwer JG (2003) Numerical solution of time-dependent advection-diffusion-reactions equations. Springer, New York

Jacobi MN, Görnerup O (2009) A spectral method for aggregating variables in linear dynamical systems with application to cellular automata renormalization. Adv Complex Syst 12:1–25

Kao RR, Danon L, Green DM, Kiss IZ (2006) Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. Proc R Soc B 273:1999–2007

Keeling MJ (1999) The effects of local spatial structure on epidemiological invasions. Proc R Soc Lond B 266:859–867

Keeling MJ (2005) The implications of network structure for epidemic dynamics. Theor Popul Biol 67:1–8

Keeling MJ, Eames KTD (2005) Networks and epidemic models. J R Soc Interface 2:295–307

Keeling MJ, Ross JV (2008) On methods for studying stochastic disease dynamics. J R Soc Interface 5:171–181

Keeling MJ, Rand DA, Morris AJ (1997) Correlation models for childhood epidemics. Proc R Soc B 264:1149–1156

Kemeny JG, Snell JL (1976) Finite Markov chains, 2nd edn. Springer, New York

Kenah E, Robins JM (2007) Network-based analysis of stochastic SIR epidemic models with random and proportionate mixing. J Theor Biol 249:706–722

Kermack WO, McKendrick AG (1927) A contribution to the mathematical study of epidemics. Proc R Soc Lond Ser A 115:700–721

Kiss IZ, Green DM, Kao RR (2005) Disease contact tracing in random and clustered networks. Proc R Soc B 272:1407–1414

Kiss IZ, Green DM, Kao RR (2006a) The network of sheep movements within Great Britain: network properties and their implications for infectious disease spread. J R Soc Interface 3:669–677

Kiss IZ, Green DM, Kao RR (2006b) The effect of network heterogeneity and multiple routes of transmission on final epidemic size. Math Biosci 203:124–136

Kiss IZ, Green DM, Kao RR (2008) The effect of network mixing patterns on epidemic dynamics and the efficacy of disease contact tracing. J R Soc Interface 5:791–799

Kurtz TG (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. J Appl Prob 7:49–58

Kurtz TG (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. J Appl Prob 8:344–356

Lipsitch M et al (2003) Transmission dynamics and control of severe acute respiratory syndrome. Science 300:1966–1970

May RM, Lloyd AL (2001) Infection dynamics on scale-free networks. Phys Rev E 64:066112

Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC (2005) Network theory and SARS: predicting outbreak diversity. J Theor Biol 232:71–81

Nåsell I (1996) The quasi-stationary distribution of the closed endemic SIS model. Adv Appl Probab 28:895–932

Newman MEJ (2002) The spread of epidemic disease on networks. Phys Rev E 66:016128

Picard P (1965) Sur les modèles stochastique logistiques en démographie. Ann Inst Henri Poincaré B II:151–172

Rand DA (1999) Correlation equations for spatial ecologies. In: McGlade J (ed) Advanced ecological theory. Blackwell, Oxford, pp 100–142

Sato K, Matsuda H, Sasaki A (1994) Pathogen invasion and host extinction in lattice structured populations. J Math Biol 32:251–268

Sharkey KJ (2008) Deterministic epidemiological models at the individual level. J Math Biol 57:311–331

Smith GJ et al (2009) Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459:1122–1125

van Baalen M (2000) Pair approximations for different spatial geometries. In: Dieckmann U, Law R, Metz JAJ (eds) The geometry of ecological interactions: simplifying complexity. Cambridge University Press, pp 359–387

Yap HP (1986) Some topics in graph theory. In: London Mathematical Society, Lecture notes series 108. Cambridge University Press, Cambridge