IMA Journal of Applied Mathematics Advance Access published March 30, 2012

IMA Journal of Applied Mathematics (2012) 1–20
doi:10.1093/imamat/hxs001

From exact stochastic to mean-field ODE models: a new approach to prove convergence results

PÉTER L.SIMON*

Institute of Mathematics, Eötvös Loránd University, 1117 Budapest, Hungary
*Corresponding author: simonp@cs.elte.hu

AND

ISTVAN Z.KISS

Department of Mathematics, School of Mathematical and Physical Sciences, University of Sussex, Falmer, Brighton BN1 9RF, UK

[Received on 6 October 2010; revised on 2 September 2011; accepted on 12 January 2012]

In this paper, the rigorous linking of exact stochastic models to mean-field approximations is studied. Using a continuous-time Markov chain, we start from the exact formulation of a simple epidemic model on a certain class of networks, including completely connected and regular random graphs, and rigorously derive the well-known mean-field approximation that is usually justified based on biological hypotheses. We propose a unifying framework that incorporates and discusses the details of two existing proofs and we put forward a new ordinary differential equation (ODE)-based proof. The more well-known proof is based on a first-order partial differential equation approximation, while the other, more technical one, uses Martingale and Semigroup theory. We present the main steps of both proofs to investigate their applicability in different modelling contexts and to make these ideas more accessible to a broader group of applied researchers. The main result of the paper is a new ODE-based proof that may serve as a building block to prove similar convergence results for more complex networks. The new proof is based on deriving a countable system of ODEs for the moments of a distribution of interest and proving a perturbation theorem for this infinite system.

Keywords: epidemic model; network; mean-field approximation; countable system of ODEs; Markov chain

1. Introduction

Complex networks occur in a large variety of real-world systems ranging from ecology and epidemiology to neuroscience (Bansal *et al.*, 2007; Keeling & Eames, 2005; Sporns & Kötter, 2004). In most applications, networks provide the backbone on which various dynamical processes unfold. For example, infectious diseases transmit on intricate social networks, while neurons interact on non-trivial weighted and dynamical graphs. This underpinned the rapid development of research that seeks to understand how the structure/topology of the network impacts on the behaviour of different dynamics on networks (Bansal *et al.*, 2007; Keeling & Eames, 2005). The analysis of even the simplest *dynamics on networks* can be challenging mathematically, and often, results are mainly simulation based. As a result, research in this direction is fragmented into more theoretical work that explores the rigorous link between exact stochastic models and their ordinary differential equation (ODE)-based mean-field approximations (Ethier & Kurtz, 2005; Kurtz, 1970, 1971, 1980), and work that mainly relies on simulation. While simulations can be straightforward to implement, the often large number of parameters

makes the exploration of the possible behaviours difficult and generalization of simulation results is rarely possible. In an effort to increase tractability and depart from a purely simulation-based approach, various simple differential equation models have been proposed. These are all different from simple mean-filed models, which operate on the homogenous random mixing assumption, in that they capture non-trivial network features such as network heterogeneity, clustering or can accommodate dynamically evolving networks. These models range from pairwise models (Keeling, 1999; Rand, 1999; Sato *et al.*, 1994; Trapman, 2007; van Baalen, 2000) and ODE-based heterogeneous mixing models (Kiss *et al.*, 2006; Moreno *et al.*, 2002) to probability function formalism (Volz & Meyers, 2007, 2009). However, in almost all cases, the performance of these more sophisticated models is only tested by comparing ODE-based results to pure simulation. Thus, the goodness of fit is mostly performed by numerical and/or visual inspection without rigorous mathematical arguments. The major obstacle that precludes a theoretical formalism for comparison is either due to not being able to derive the Kolmogorov equations or, in the case where this is possible, these are intractable due to their sheer number.

The problem of rigourously linking exact stochastic models to mean-field approximations goes back to the early work of Kurtz (1970, 1971). Kurtz studied pure-jump density-dependent Markov processes where apart from providing a method for the derivation of the mean-field model also used solid mathematical arguments to prove the stochastic convergence of the exact to the mean-field model. His earlier results (Kurtz, 1970, 1971) relied on Trotter-type approximation theorems for operator semigroups. Later on, the results were embedded in a more general context of Martingale Theory (Ethier & Kurtz, 2005). These results have been cited and extensively used by modellers in areas such as ecology and epidemiology to justify the validity of heuristically formulated mean-field models. The existence of several approximation models, often derived based on different modelling intuitions and approaches, has recently highlighted the need to try and unify these and test their performance against the exact stochastic models (House & Keeling, 2011). Some steps in this directions have been made (Ball & Neal, 2008; Lindquist *et al.*, 2010), where authors clearly state the link between exact and mean-field models.

The present paper, in the case of a simple SIS model and a suitable class of networks, including completely connected and regular random graphs, proposes a unifying framework that incorporates and discusses the details of two existing proofs and proposes a new ODE-based proof. This complements and offers an alternative to the existing ones that are purely based on stochastic theory and partial differential equation (PDE) arguments. The paper is organized as follows. In Section 2, the model is formulated and we present the main result in general terms and discuss the three different approaches used to prove the convergence of the exact stochastic to the mean-field model. In Section 3, we give the detailed proof based on PDE arguments, while in Section 4, the proof based on stochastic theory arguments is presented. Section 5 contains the new ODE-based approach, with pluses and minuses of the three different models included in the final Section.

2. Model

Let us consider the simple SIS-type dynamics on a graph with N nodes and assume that the structure of the network allows us to determine N_{SI}, the number of SI pairs, once the number of infected nodes N_{I} is known. The simplest graph satisfying this assumption is the complete graph for which $N_{SI}(k) = k(N-k)$ if $N_{\mathrm{I}} = k$. In the case of an n-regular random graph, the widely used approximation for the number of SI pairs is $N_{SI}(k) = k(N-k)n/N$ if $N_{\mathrm{I}} = k$. Once the $N_{SI}(k)$ function is defined, then the epidemic propagation on the graph can be described by a Markov chain with state space $\{0, 1, 2, \ldots, N\}$. Denoting by $x_k(t)$ the probability of finding k infectious nodes, the Kolmogorov equation (or master equation) takes the form

$$\dot{x}_k = a_{k-1}x_{k-1} - b_k x_k + c_{k+1}x_{k+1}, \quad k = 0, 1, \ldots, N, \tag{2.1}$$

where

$$a_k = \tau N_{SI}(k), \quad c_k = \gamma k, \quad b_k = a_k + c_k, \quad a_{-1} = 0 = c_{N+1}. \tag{2.2}$$

In the case of a complete graph, it is known that the mean-filed approximation is available only if τ scales with $1/N$, hence $\tau = \beta/N$ is used, yielding $a_k = \beta k(N-k)/N$. In the case of an n-regular random graph, $a_k = n\tau k(N-k)/N$ and this can be written in the same form as for the complete graph with $\beta = n\tau$. Therefore, in the following, we assume that the master equation takes the form (2.1) and

$$a_k = \beta k(N-k)/N, \quad c_k = \gamma k, \quad b_k = a_k + c_k, \quad a_{-1} = 0 = c_{N+1}. \tag{2.3}$$

Let us assume that initially the number of infected nodes is k_0. Thus the initial condition to (2.1) is

$$x_{k_0}(0) = 1, \quad x_k(0) = 0 \quad \text{for } k \neq k_0, \tag{2.4}$$

with the expected value of the number of infected nodes given by

$$[I](t) = \sum_{k=0}^{N} k x_k(t). \tag{2.5}$$

Differentiating $[I]$ with respect to time and using the Kolmogorov equations for x_k, one can derive the following differential equation for $[I]$:

$$[\dot{I}] = \frac{\beta}{N}[SI] - \gamma[I], \tag{2.6}$$

where $[SI]$ is the expected value of the number of SI-type edges. Equation (2.6) cannot be used to determine the expected value $[I]$ since $[SI]$ cannot be expressed in terms of $[I]$. However, the approximation

$$[SI] \approx N_{SI}([I])$$

yields a self-contained differential equation the solution of which approximates $[I]$. In the case of a complete graph, this approximation takes the form $[SI] \approx [S][I]$, and for a homogeneous random graph, it is $[SI] \approx [S][I]n/N$. Substituting this approximation into (2.6) and dividing by N, we introduce the variable $i(t)$ instead of $[I](t)/N$. Then for i, the following simple differential equation holds (for both cases, but with different meaning of β)

$$\dot{i} = \beta i(1-i) - \gamma i. \tag{2.7}$$

This equation is known as the mean-field approximation of the original Kolmogorov equation (2.1). It is well-known that $i(t)$ is a good approximation of $[I](t)/N$ in the following sense.

THEOREM 2.1 If $i(0) = [I](0)/N$, then for any $t \geqslant 0$, we have

$$\lim_{N\to\infty} \frac{[I](t)}{N} = i(t).$$

In fact, the statement of the theorem is not rigorous in this form since the type of the convergence is not specified and this will depend on the method of proof. There are basically two different methods of proof and these yield different types of limits. The two main approaches use (a) first-order PDE and (b) martingale and semigroup theory arguments.

The first-order PDE approach yields that $[I](t)/N$ tends to $i(t)$ for any fixed t. This is the most intuitive approach since it is based on the idea that for large N, the discrete distribution $x_k(t)$ can be approximated by a continuous density function. The exact statement that can be proved by using this method is presented in Theorem 3.1. The main steps of the proof can be found in the Appendix of Diekmann & Heesterbeek (2000), however not all details of the rigorous mathematical proof are presented there. In Simon *et al.* (2011), a rigorous proof is given and for sake of completeness, we briefly summarize this in Section 3.

The stochastic approach yields that the stochastic variable $I(t)/N$ (not the expected value) tends stochastically to $i(t)$. This implies that the expected value $[I](t)/N$ also tends to $i(t)$. The statement is formulated in exact terms in Theorem 4.1. The theorem is proved in several different ways in Ethier & Kurtz (2005), Kurtz (1970, 1971). The first proof was based on a Trotter-type approximation theorem for semigroups followed by a proof based on martingale theory. The proof in Ethier & Kurtz (2005), which is valid in a general context, reduces the problem to the study of Poisson processes by using the previously developed semigroup and martingale techniques. In Section 4, we present the main steps of the proof in Ethier & Kurtz (2005) applied to our special setting. This enables the reader to follow the main ideas of the stochastic proof without going into and understanding the technical details of the original proof in Ethier & Kurtz (2005).

The main purpose of this paper is to show a new ODE-based approach. We will call this an elementary approach since a self-contained proof of the theorem will be shown without using a combination of highly specialist mathematical tools from different areas, the availability of which is beyond the opportunities of the average scientist working in mathematical ecology, epidemiology or other applied research areas. Moreover, this elementary proof may lead to future work where proving similar results for more complex networks can be attempted. According to our knowledge, the above theorem has not been generalized to more complicated networks by using the two more sophisticated approaches.

Our elementary, ODE-based approach, presented in Section 5, yields that $[I](t)/N$ tends uniformly on bounded time intervals to $i(t)$. Moreover, we also give an upper estimate for the difference in terms of network size N, and we prove that $i(t)$ is an upper approximation of $[I](t)/N$. According to our knowledge, this has not been previously verified and it does not follow from the previous two approaches.

3. First-order PDE approach

In this section, the first proof of Theorem 2.1 is given. The main idea of the proof is based on the observation that for large N, the discrete distribution $x_k(t)$ can be approximated by a continuous density function $\rho(t, z)$. The rigorous version of Theorem 2.1 in this context reads as follows.

THEOREM 3.1 If $i(0) = [I](0)/N$, then for any $t > 0$, we have

$$\lim_{N\to\infty} \left| i(t) - \frac{[I](t)}{N} \right| = 0.$$

Let us 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, $\dot{x}_k$, $x_k(t)$, $x_{k-1}(t)$ and $x_{k+1}(t)$ in (2.1) can be formally change to $\partial_t \rho(t, z)$, $\rho(t, z)$, $\rho(t, z - 1/N)$ and $\rho(t, z + 1/N)$, respectively. This

leads to the following partial differential equation:

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

Now using the approximations

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

neglecting the $1/N$ and $1/N^2$ terms and writing ρ instead of $\rho(t,z)$, after some algebra, the following first-order PDE for ρ is obtained

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

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

$$\partial_t\rho = \partial_z(g\rho). \tag{3.1}$$

This first-order PDE needs an initial condition of the following type

$$\rho(0,z) = \rho_0(z). \tag{3.2}$$

Since the formal relation between the variables is $z = k/N$, the initial condition (2.4) yields

$$\rho_0(z) = 1 \quad \text{for } \frac{k_0}{N} < z < \frac{k_0+1}{N} \quad \text{and} \quad \rho_0(z) = 0 \quad \text{otherwise.}$$

Finally, the expected value of the infected nodes from the first-order PDE needs to be determined. Thus, we have to find the function corresponding to $[i_N](t) = [I](t)/N$ in (2.5). Using $z = k/N$ and changing the term $x_k(t)$ to $\rho(t,z)$, we note that the sums in (2.5) correspond to an integral. Namely, $[i_N](t)$ 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

$$N\int_0^1 z\rho(t,z)\mathrm{d}z.$$

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

$$i^*(t) = \frac{\int_0^1 z\rho(t,z)\mathrm{d}z}{\int_0^1\rho_0(z)\mathrm{d}z}. \tag{3.3}$$

The mean-field equation (2.7) can be solved explicitly and the solution is given by

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

where $i_0 = i(0)$ is the initial condition and

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

The first-order PDE (3.1) can also be solved explicitly, and then (3.3) yields

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

Having these explicit formulas for $i^*(t)$ and $i(t)$, it is easy to see that i^* is not a solution of the mean-field equation (2.7) but it can be proved that as $N \to \infty$ it tends to the solution of (2.7). Namely, we have the following Lemma.

LEMMA 3.1 Let ρ be the solution of the system (3.1)–(3.2). Let $i^*(t)$ be defined by (3.3). Let $i(t)$ be the solution of the scaled mean-field equation given by (2.7) with initial condition $i(0) = k_0/N$. Then for any $t \geqslant 0$, we have

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

The Lemma can be proved by using the explicit formulas for $i^*(t)$ and $i(t)$. Now the proof of Theorem 3.1 can be concluded as follows. We want to prove that the scaled expected value $[i_N](t)$ tends to the solution $i(t)$ 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 (2.1) as $N \to \infty$. Using this PDE, we defined the function $i^*(t)$ that corresponds to $[i_N](t)$. According to Lemma 3.1, $i^*(t)$ is close to $i(t)$ for large N. Hence, we only have to show finally that $[i_N](t)$ is close to $i^*(t)$. Thus the proof of Theorem 3.1 will be complete if the following Lemma is verified.

LEMMA 3.2 Let x_k be the solution of (2.1) satisfying the initial condition given by (2.4), and let ρ be the solution of (3.1) with initial condition given by (3.2). Let $[i_N](t) = [I](t)/N$ and let $[I](t)$ and $i^*(t)$ be defined by (2.5) and (3.3). Then for any $t \geqslant 0$, we have

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

The proof of the Lemma is based on the fact that system (2.1) can be considered as the discretization of the first-order PDE (3.1) in the variable z. It is known even for more general PDEs, see, e.g. Chapters 3 and 4 in Hundsdorfer & Verwer (2003), that the solution of the discretized system tends to that of the PDE as the step size of the discretization goes to zero, that is in our case N tends to infinity.

4. Stochastic proof of Theorem 2.1

Let us denote by $(I(t))_{t \geqslant 0}$ the stochastic process that determines the number of infected nodes at time t. In this section, we will prove that $I(t)/N$ converges stochastically to $i(t)$ as $N \to \infty$, this is formulated in the following Theorem.

THEOREM 4.1 If $i(0) = [I](0)/N$, then for any $T > 0$, there exist $K > 0$, such that for any $\delta > 0$, we have

$$P\left(\left| i(t) - \frac{I(t)}{N} \right| > \delta \right) \leqslant \frac{K}{N\delta^2}, \quad \text{for all } t \in [0, T].$$

It is important to note that this theorem is stronger than Theorem 2.1 since it implies that the expected value of $I(t)$ converges to $i(t)$.

Before going into the details of the proof, we note that this approach can be generalized to so-called density dependent Markov chains. In our case, this means that there exist two continuous functions $A, C\colon \mathbb{R} \to \mathbb{R}$, such that the transition coefficients in the Kolmogorov equation (2.1) can expressed as follows:

$$\frac{a_k}{N} = A\left(\frac{k}{N}\right), \quad \frac{c_k}{N} = C\left(\frac{k}{N}\right).$$

From (2.3), these functions are

$$A(z) = \beta z(1-z), \quad C(z) = \gamma z.$$

Following Kurtz (1970), we introduce $F(z) = A(z) - C(z)$. This is motivated by being relatively easy to derive the following equation:

$$E(I(t)) = E(I(0)) + \int_0^t E(F(I(s)))\mathrm{ds},$$

where E stands for the expected value (hence $E(I(t)) = [I](t)$). Therefore, if F and E commute (i.e. $E(F(I)) = F(E(I))$), the expected value of $[I]$ satisfies the following mean-field equation:

$$[\dot{I}] = F([I]). \tag{4.1}$$

At this stage, it is worth noting that for certain scenarios, simple arguments can be used to derive the mean-field equations without further precise mathematical arguments. Namely, when the Kolmogorov equations are numerically tractable, the precise evolution of the probability distribution over time can be computed. If this distribution proves to be unimodal and highly picked, then F and E commute at least approximately and then (4.1) follows immediately.

The main step of this approach is to prove that $I(t)$ can be expressed as follows:

$$I(t) = I(0) + Y_1\left(\int_0^t \beta I(s)\frac{S(s)}{N}\,\mathrm{ds}\right) - Y_2\left(\int_0^t \gamma I(s)\mathrm{ds}\right), \tag{4.2}$$

where Y_1 and Y_2 are standard Poisson processes (with $\lambda = 1$). The equation in this form can be found in Kurtz (1980) and in Section 2 of Chapter 11 in Ethier & Kurtz (2005). The derivation is based on Martingale and Semigroup theory and it can be found in Ethier & Kurtz (2005). The choice of this equation as a starting point is also motivated by its ease of intuitive interpretation. The Poisson process Y_1 counts the number of infections in the time interval $[0, t]$ the intensity of which can be expressed by the integral in the argument of Y_1. Similarly, the Poisson process Y_2 counts the number of recoveries in the time interval $[0, t]$, the intensity of which can be expressed by the integral in the argument of Y_2.

We note that the earlier approach of Kurtz (1970, 1971) does not use Martingale theory. In these two papers, a self-contained proof can be found and can be followed without understanding the notations and most of the preliminary work presented in Chapters 3 and 4 of the book Ethier & Kurtz (2005).

Let us introduce

$$i_N(t) = \frac{I(t)}{N} \tag{4.3}$$

and $\tilde{Y}_i(\tau) = Y_i(\tau) - \tau$, which is a Poisson process centred at its expectation, that is $E(\tilde{Y}_i(\tau)) = 0$ for all τ.

Dividing (4.2) by N, after some simple calculations, we get

$$i_N(t) = i_N(0) + \int_0^t F(i_N(s))\mathrm{d}s + \frac{1}{N}\tilde{Y}_1\left(\int_0^t \beta I(s)\frac{S(s)}{N}\,\mathrm{d}s\right) - \frac{1}{N}\tilde{Y}_2\left(\int_0^t \gamma I(s)\mathrm{d}s\right). \tag{4.4}$$

If $t \in [0, T]$, then the value of the integral in $\tilde{Y}_1$ is bounded by 0 and βNT, and the value of the integral in $\tilde{Y}_2$ is bounded by 0 and γNT. Hence, the following inequalities hold true:

$$\sup_{t\in[0,T]}\left|\tilde{Y}_1\left(\int_0^t \beta I(s)\frac{S(s)}{N}\,\mathrm{d}s\right)\right| \leqslant \tilde{Y}_1(\beta NT), \quad \sup_{t\in[0,T]}\left|\tilde{Y}_2\left(\int_0^t \gamma I(s)\mathrm{d}s\right)\right| \leqslant \tilde{Y}_2(\gamma NT). \tag{4.5}$$

The proof is now based on the following proposition, a law of large numbers type statement and can be proved by using Chebyshev's inequality like the law of large numbers.

PROPOSITION 4.1 Let $X(t)$ be a standard Poisson process (with $\lambda = 1$). Let $Y(t) = X(t) - t$ and $c > 0$ be a positive number. Then for any $\varepsilon > 0$ and for any $n \in \mathbb{N}$, the following inequality holds:

$$P\left(\frac{1}{n}|Y(cn)| > \varepsilon\right) \leqslant \frac{c}{n\varepsilon^2}.$$

Proof. It follows easily that $E(Y(t)) = 0$ and that the variance $D^2(Y(t)) = t$ for all t. Let us define $Z_n = Y(cn)/n$. Then, $E(Z_n) = 0$ and $D^2(Z_n) = c/n$ for all n. Now applying Chebyshev's inequality to Z_n, we get the desired statement. □

Using this proposition, an upper estimate for

$$y_N(t) = \frac{1}{N}\tilde{Y}_1\left(\int_0^t \beta I(s)\frac{S(s)}{N}\,\mathrm{d}s\right) - \frac{1}{N}\tilde{Y}_2\left(\int_0^t \gamma I(s)\mathrm{d}s\right) \tag{4.6}$$

can be derived as follows. From (4.5), we obtain

$$\sup_{t\in[0,T]}|y_N(t)| \leqslant \frac{1}{N}\tilde{Y}_1(\beta NT) + \frac{1}{N}\tilde{Y}_2(\gamma NT). \tag{4.7}$$

Thus, if

$$\sup_{t\in[0,T]}|y_N(t)| > \varepsilon,$$

then at least one of the inequalities

$$\frac{1}{N}\tilde{Y}_1(\beta NT) > \frac{\varepsilon}{2} \quad \text{or} \quad \frac{1}{N}\tilde{Y}_2(\gamma NT) > \frac{\varepsilon}{2}$$

holds. Hence, $P(\sup_{t\in[0,T]}|y_N(t)| > \varepsilon)$ can be estimated by the probability of the larger. Therefore, it can be obviously estimated by the sum of the two probabilities

$$P\left(\sup_{t\in[0,T]}|y_N(t)| > \varepsilon\right) \leqslant P\left(\frac{1}{N}\tilde{Y}_1(\beta NT) > \frac{\varepsilon}{2}\right) + P\left(\frac{1}{N}\tilde{Y}_2(\gamma NT) > \frac{\varepsilon}{2}\right). \tag{4.8}$$

Thus, using Proposition 4.1, we obtain

$$P\left(\sup_{t\in[0,T]}|y_N(t)|>\varepsilon\right)\leqslant\frac{4(\beta+\gamma)T}{N\varepsilon^2}. \tag{4.9}$$

Now the difference of $i_N(t)$ and $i(t)$ can be estimated (the latter is defined by (2.7)).

PROPOSITION 4.2 Let $i(t)$ be the solution of (2.7) and let $i_N(t)$ be given by (4.3). Let us denote by M the Lipschitz constant of F on $[0,1]$. If $i_N(0)=i(0)$, then for all $t\geqslant 0$, the following inequality holds:

$$|i_N(t)-i(t)|\leqslant|y_N(t)|\mathrm{e}^{Mt}.$$

Proof. The functions i_N and i satisfy

$$i_N(t)=i_N(0)+\int_0^t F(i_N(s))\mathrm{d}s+y_N(t)$$

and

$$i(t)=i(0)+\int_0^t F(i(s))\mathrm{d}s.$$

Subtracting the two equations, using the initial conditions and the Lipschitz constant of F, we obtain

$$|i_N(t)-i(t)|\leqslant|y_n(t)|+\int_0^t M|i_N(s)-i(s)|\mathrm{d}s.$$

Using Gronwall's lemma, the statement follows easily. □

Thus, if

$$\sup_{t\in[0,T]}|i_N(t)-i(t)|>\delta,$$

then

$$\sup_{t\in[0,T]}|y_N(t)|>\delta\mathrm{e}^{-MT}.$$

Hence,

$$P\left(\sup_{t\in[0,T]}|i_N(t)-i(t)|>\delta\right)\leqslant P\left(\sup_{t\in[0,T]}|y_N(t)|>\delta\mathrm{e}^{-MT}\right).$$

Finally, we can use the estimate in (4.9) to get

$$P\left(\sup_{t\in[0,T]}|i_N(t)-i(t)|>\delta\right)\leqslant\frac{4(\beta+\gamma)T\mathrm{e}^{2MT}}{N\delta^2},$$

and this proves Theorem 4.1.

5. ODE-based proof of Theorem 2.1

In this section, the main result of the paper is formulated and proved. This is an ODE-based proof where the evolution equations of the moments of the distribution form a countable system of ODEs. The proof only uses ODE techniques and a perturbation theorem for the infinite system is presented.

THEOREM 5.1 Let i be the solution of (2.7) with initial condition $i(0) = [I](0)/N$, and let $[I]$ be given by (2.5) through the master equation (2.1). Then for any $T > 0$ there exist $K > 0$, such that

$$\left| i(t) - \frac{[I](t)}{N} \right| \leqslant \frac{K}{N}, \qquad \text{for all } t \in [0, T].$$

In fact, we have $0 \leqslant i(t) - \frac{[I](t)}{N} \leqslant \frac{K}{N}$ for $t \in [0, T]$, that is $i(t)$ is an upper approximation of $[I](t)/N$.

The approximation (2.7) of (2.6) is based on the moment closure technique. Thus, to keep an exact system, all higher-order moments must be considered and this leads to a countable (infinite) system of ODEs.

5.1 *Moment equations and their approximations*

Let us introduce the jth moment of the probability distribution $x_k(t)$ (i.e. the probability of finding states with k infectious nodes, where $k = 0, 1, \ldots, N$)

$$y_j(t) = \sum_{k=0}^{N} \left(\frac{k}{N}\right)^j x_k(t). \tag{5.1}$$

To derive differential equations for the moments, the following Proposition is given.

PROPOSITION 5.1 Let r_k $(k = 0, 1, 2, \ldots)$ be a sequence and let $r(t) = \sum_{k=0}^{N} r_k x_k(t)$, where $x_k(t)$ is given by (2.1). Then

$$\dot{r}(t) = \sum_{k=0}^{N} (a_k(r_{k+1} - r_k) + c_k(r_{k-1} - r_k))x_k(t).$$

Proof. From (2.1), we obtain

$$\begin{aligned}\dot{r}(t) &= \sum_{k=0}^{N} r_k \dot{x}_k(t) = \sum_{k=1}^{N} r_k a_{k-1} x_{k-1}(t) - \sum_{k=0}^{N} r_k b_k x_k(t) + \sum_{k=0}^{N-1} r_k c_{k+1} x_{k+1}(t) \\ &= \sum_{k=0}^{N-1} r_{k+1} a_k x_k(t) - \sum_{k=0}^{N} r_k b_k x_k(t) + \sum_{k=1}^{N} r_{k-1} c_k x_k(t).\end{aligned}$$

Using that $a_N = 0$, $c_0 = 0$ and $b_k = a_k + c_k$, we get

$$\dot{r}(t) = \sum_{k=0}^{N} \left(r_{k+1} a_k - r_k(a_k + c_k) + r_{k-1} c_k\right) x_k(t) = \sum_{k=0}^{N} (a_k(r_{k+1} - r_k) + c_k(r_{k-1} - r_k))x_k(t).$$

□

Before applying Proposition 5.1 with $r_k = (k/N)^j$, it is useful to define the following two new expressions:

$$R_{k,j} = \frac{(k+1)^j - k^j - jk^{j-1}}{N^{j-1}}, \quad Q_{k,j} = \frac{(k-1)^j - k^j + jk^{j-1}}{N^{j-1}}.$$

Combining these with Proposition 5.1 leads to

$$\dot{y}_j(t) = \sum_{k=0}^{N} \left(\frac{a_k}{N} \left(j \frac{k^{j-1}}{N^{j-1}} + R_{k,j} \right) + \frac{c_k}{N} \left(-j \frac{k^{j-1}}{N^{j-1}} + Q_{k,j} \right) \right) x_k(t).$$

From (2.3), we get that

$$\frac{a_k}{N} - \frac{c_k}{N} = (\beta - \gamma) \frac{k}{N} - \beta \frac{k^2}{N^2},$$

and therefore,

$$\dot{y}_j(t) = \sum_{k=0}^{N} \left(j(\beta - \gamma) \frac{k^j}{N^j} - j\beta \frac{k^{j+1}}{N^{j+1}} \right) x_k(t) + \sum_{k=0}^{N} \left(\frac{a_k}{N} R_{k,j} + \frac{c_k}{N} Q_{k,j} \right) x_k(t).$$

Hence,

$$\dot{y}_j(t) = j(\beta - \gamma) y_j(t) - j\beta y_{j+1}(t) + \frac{1}{N} d_j(t), \tag{5.2}$$

where

$$d_j(t) = \sum_{k=0}^{N} (a_k R_{k,j} + c_k Q_{k,j}) x_k(t). \tag{5.3}$$

Using the binomial theorem $R_{k,j}$ and $Q_{k,j}$ can be expressed in terms of the powers of k, hence d_j can be expressed as $d_j(t) = \sum_{l=1}^{j} d_{jl} y_l(t)$ with some coefficients d_{jl}. Therefore, system (5.2) is an infinite homogeneous linear system for the moments y_j. This homogeneous linear system is not written in the usual matrix form because it is useful to separate the $\mathrm{O}\left(\frac{1}{N}\right)$ terms in order to handle the large N limit. The d_j terms contain N, hence to use the $1/N \to 0$ limit it has to be shown that d_j remains bounded as N goes to infinity. This is proved in the next Proposition.

PROPOSITION 5.2 For the functions d_j, the following estimates hold.

$$0 \leqslant d_j(t) \leqslant \frac{j(j-1)}{2}(\beta + \gamma), \quad \text{for all } t \geqslant 0.$$

Proof. Taylor's theorem, with second degree remainder in Lagrange form, states that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(\xi) \frac{(x - x_0)^2}{2},$$

where ξ is between x_0 and x. This simple result can be used to find estimates for both $R_{k,j}$ and $Q_{k,j}$. In particular, applying the above result when $f(x) = x^j$, $x = k+1$ and $x_0 = k$ gives

$$R_{k,j} = \frac{j(j-1)}{2}\frac{\xi^{j-2}}{N^{j-1}}$$

with some $\xi \in [k, k+1]$. Similarly, when $x = k-1$ and $x_0 = k$, we obtain

$$Q_{k,j} = \frac{j(j-1)}{2}\frac{\eta^{j-2}}{N^{j-1}}$$

with some $\eta \in [k, k+1]$. Hence, $R_{k,j}$ and $Q_{k,j}$ are non-negative yielding that $d_j(t) \geqslant 0$. On the other hand, using (2.3) and that $\xi/N \leqslant 1$ and $\eta/N \leqslant 1$ leads to the inequality given below

$$a_k R_{k,j} + c_k Q_{k,j} \leqslant \frac{j(j-1)}{2}\left(\frac{a_k}{N} + \frac{c_k}{N}\right) \leqslant \frac{j(j-1)}{2}(\beta + \gamma).$$

Hence, the statement follows immediately from (5.3) and using that $\sum_{k=0}^{N} x_k(t) = 1$. □

The exact equations for the moments (5.2) are now setup such that the limit of $N \to \infty$ can be considered. This leads to the following system:

$$\dot{z}_j(t) = j(\beta - \gamma)z_j(t) - j\beta z_{j+1}(t), \tag{5.4}$$

with the same initial condition as for y_j, that is $z_j(0) = k_0^j/N^j$. It is worth noting that a solution of system (5.4) can be obtained in the form $z_j = z^j$. Substituting this expression for z_j in (5.4), we get the following equation for z:

$$\dot{z} = (\beta - \gamma)z - \beta z^2,$$

with initial condition $z(0) = k_0/N$. This differential equation is the same as (2.7) for i. Hence, the approximating equations for the moments (5.4) are not only more tractable but they allow to recover the mean-field equations. However, $y_1(t) = [I](t)/N$ and z are not identical. The former comes from the exact system, while z is based on the approximating equations obtained from the exact system in the limit of $N \to \infty$. Therefore, the relation between the two needs to be formally established, see Theorem 5.2. The following two statements prove that indeed, $z_1 = z = i$ is the only uniformly bounded solution of equation (5.4) and that z_1 is a good approximation to y_1 for $t \in [0, T]$ and for N large. The lemma and theorem given below play a crucial role in completing the proof of Theorem 5.1. To increase the clarity and transparency of the proof a diagram linking all propositions, lemmas and theorems is given in Fig. 1.

LEMMA 5.1 System (5.4) subject to the initial condition $z_j(0) = k_0^j/N^j$ has a unique uniformly bounded solution, where uniform boundedness means that there exists M such that $|z_j(t)| \leqslant M$ for all j. This solution can be given as $z_j = z_1^j$ and $z_1(t) = i(t)$.

THEOREM 5.2 Let us assume that the solutions of systems (5.2) and (5.4) satisfy the same initial condition $y_j(0) = k_0^j/N^j = z_j(0)$. Then for any $T > 0$, there exist $K > 0$, such that

$$0 \leqslant z_1(t) - y_1(t) \leqslant \frac{K}{N} \quad \text{for } t \in [0, T].$$

The rather technical proof of the lemma is postponed to the Appendix.

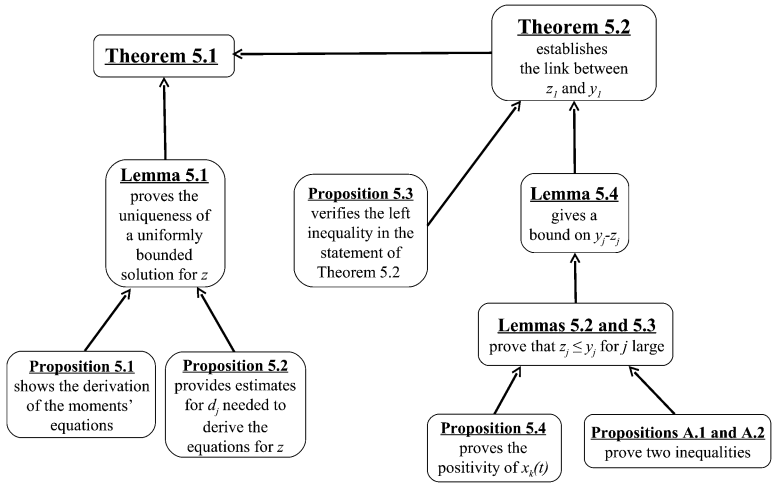

FIG. 1. The flow of the proof of Theorem 5.1

5.2 *Proof of Theorem 5.2*

In this subsection, we prove Theorem 5.2. This together with with Lemma 5.1 yields our main result formulated in Theorem 5.1. Throughout this section, let y_j be given by (5.1), and let z_j be the unique solution of (5.4) subject to the initial condition $z_j(0) = k_0^j/N^j$. Then the following proposition verifies the left inequality in the statement of Theorem 5.2.

PROPOSITION 5.3 Under the above conditions, we have that $y_1(t) \leqslant z_1(t)$ for all $t \geqslant 0$.

Proof. Since the variance $(y_2 - y_1^2)$ is non-negative, it follows that $y_1^2(t) \leqslant y_2(t)$ for all t. Since $d_1 = 0$, the first equation of system (5.2) now reads as

$$\dot{y}_1 = (\beta - \gamma)y_1 - \beta y_2.$$

Hence, $\dot{y}_1 \leqslant (\beta - \gamma)y_1 - \beta y_1^2$. If there exists $t_2 > 0$ such that $y_1(t_2) > z_1(t_2)$, then there exists $t_1 < t_2$ for which $y_1(t_1) = z_1(t_1)$ and $y_1(t) > z_1(t)$ for all $t \in (t_1, t_2]$. Let $v(t) = y_1(t) - z_1(t)$ for $t \in [t_1, t_2]$. Then using the function $F(x) = (\beta - \gamma)x - \beta x^2$ gives

$$\dot{v} = \dot{y}_1 - \dot{z}_1 \leqslant f(y_1) - f(z_1) \leqslant M(y_1 - z_1) = Mv,$$

where M is the Lipschitz constant of F on the interval $[0, 1]$. Applying Gronwall's lemma to v, we get $v(t) \leqslant 0$ for all $t \in [t_1, t_2]$, which is a contradiction. □

In the next two Lemmas, it will be proved that if j is large enough, then $z_j \leqslant y_j$. This result will be heavily used in the proof of Lemma 5.4.

LEMMA 5.2 There exist $j_0 \in \mathbb{N}$ and $\delta > 0$, such that

$$z_j(t) \leqslant y_j(t), \qquad \text{for all } j \geqslant j_0, \quad t \in [0, \delta].$$

Proof. In order to derive an upper estimate for z_j, we exploit the fact that using $z_2 = z_1^2$ the function z_1 can be explicitly determined from the first equation of system (5.4), $\dot{z}_1 = (\beta - \gamma)z_1 - \beta z_1^2$. Introducing $q = k_0/N = z_1(0)$ and $\alpha = \beta - \gamma$ gives

$$z_1(t) = \frac{\alpha q}{(\alpha - \beta q)\exp(-\alpha t) + \beta q}.$$

To estimate this expression, two different cases need to be considered.

Case 1. If $\alpha - \beta q < 0$, then z_1 is decreasing.

In this case, let us choose a number $t' > 0$ and $\alpha' < \alpha$ such that

$$\exp(-\alpha t) \leqslant 1 - \alpha' t, \quad \text{for all } t \in [0, t'].$$

Then for all $t \in [0, t']$

$$z_1(t) \leqslant \frac{\alpha q}{(\alpha - \beta q)(1 - \alpha' t) + \beta q} = \frac{q}{1 + ct},$$

where $c = -\alpha'(\alpha - \beta q)/\alpha > 0$. Hence,

$$z_j(t) \leqslant \frac{q^j}{(1 + ct)^j}, \quad \text{for all } t \in [0, t']. \tag{5.5}$$

A trivial lower estimate for y_j is $y_j(t) \geqslant (k_0/N)^j x_{k_0}(t)$. In order to get a lower estimate for $x_{k_0}(t)$, let us multiply (2.1) by $\mathrm{e}^{b_k t}$ and integrate from 0 to t. This gives

$$x_k(t)\mathrm{e}^{b_k t} = x_k(0) + a_{k-1}\int_0^t x_{k-1}(s)\mathrm{e}^{b_k s}\mathrm{d}s + c_{k+1}\int_0^t x_{k+1}(s)\mathrm{e}^{b_k s}\mathrm{d}s. \tag{5.6}$$

In the case when $k = k_0$ and upon using the initial condition ($x_{k_0}(0) = 1$), it follows that $x_{k_0}(t) \geqslant \mathrm{e}^{-b_{k_0} t}$ for all $t > 0$. From $\mathrm{e}^{-b_{k_0} t} \geqslant 1 - b_{k_0} t$, it follows that

$$y_j(t) \geqslant q^j(1 - b_{k_0} t) \quad \text{for all } t \geqslant 0. \tag{5.7}$$

Proposition A.1, stated and proved in the Appendix, can now be applied when $d = b_{k_0}$. For an arbitrary $t_0 < 1/b_{k_0}$, the index j_0 is chosen according to the proposition. Let $\delta = \min\{t', t_0\}$. Then for all $j \geqslant j_0$ and $t \in [0, \delta]$, from (5.5) and (5.7), it follows that

$$z_j(t) \leqslant \frac{q^j}{(1 + ct)^j} \leqslant q^j(1 - b_{k_0} t) \leqslant y_j(t).$$

Case 2. If $\alpha - \beta q \geqslant 0$, then z_1 is non-decreasing.

The proof is similar hence it is presented only briefly.

The upper estimate for $z_1(t)$ in the interval $[0, 1/\alpha]$ is

$$z_j(t) \leqslant \frac{q^j}{(1 - ct)^j}, \tag{5.8}$$

where $c = \alpha - \beta q$.

The lower estimate for y_j is based on the observation that $y_j(t) \geqslant (k_0/N)^j x_{k_0}(t) + ((k_0+1)/N)^j x_{k_0+1}(t)$. Deriving lower estimates for $x_{k_0}(t)$ and for $x_{k_0+1}(t)$, it follows that there exists $j_1 \in \mathbb{N}$ and $t' > 0$, such that

$$y_j(t) \geqslant q^j(1 + d^j t), \quad \text{for all } t \in [0, t'], \quad j \geqslant j_1, \tag{5.9}$$

where $d \in (1, 1 + 1/Nq)$.

Then applying Proposition A.2, which is stated and proved in the Appendix, we get the desired statement. □

The next proposition is needed in the proof of Lemma 5.3.

PROPOSITION 5.4 For all $k \in \{0, 1, \ldots, N\}$ and for all $t > 0$, we have $x_k(t) > 0$.

Proof. In the case $k = k_0$ and upon using the initial condition ($x_{k_0}(0) = 1$), from (5.6), it follows that $x_{k_0}(t) \geqslant \mathrm{e}^{-b_{k_0} t} > 0$ for all $t > 0$. The statement for $k > k_0$ can be proved by induction. Assuming that $x_{k-1}(t) > 0$, (5.6) gives

$$x_k(t) \geqslant a_{k-1}\, \mathrm{e}^{-b_k t} \int_0^t x_{k-1}(s) \mathrm{e}^{b_k s} \mathrm{d}s > 0.$$

Using a similar argument, the statement for $k < k_0$ follows easily. □

LEMMA 5.3 For any $T > 0$, there exists $j_1 \in \mathbb{N}$, such that

$$z_j(t) \leqslant y_j(t), \quad \text{for all } j \geqslant j_1, \quad t \in [0, T].$$

Proof. Using that $z_j = i^j$, an upper bound for z_j can be derived. It follows easily from (2.7) that if $i(0) > 1 - \gamma/\beta$, then i is a decreasing function. If the opposite inequality holds, then i is an increasing function. Hence, $q = \max\{k_0/N, 1 - \gamma/\beta\}$ is an upper bound for i, that is $i(t) \leqslant q$ for all $t \geqslant 0$. Therefore,

$$z_j(t) \leqslant q^j, \quad \text{for all } t \geqslant 0. \tag{5.10}$$

A lower bound on y_j can now be derived. Let us start by choosing $k \in \{0, 1, \ldots, N\}$ such that $k/N > q$ holds and introduce j_0 and δ according to Lemma 5.2. Let r be given by

$$r = \min\{x_k(t) : t \in [\delta, T]\} > 0.$$

The positivity of r is guaranteed by Proposition 5.4. Finally, let us choose $j_1 \geqslant j_0$ in such a way that $r(k/N)^j > q^j$ for all $j \geqslant j_1$. Then for all $t \in [\delta, T]$ the following inequality holds:

$$y_j(t) \geqslant \left(\frac{k}{N}\right)^j x_k(t) \geqslant \left(\frac{k}{N}\right)^j r > q^j \geqslant z_j(t).$$

On the other hand, from Lemma 5.2, it follows that $z_j(t) \leqslant y_j(t)$ for $t \in [0, \delta]$ since $j \geqslant j_1 \geqslant j_0$. □

To formulate our final Lemma, a new variable is introduced together with its corresponding evolution equation. For all $j \in \mathbb{N}$ and $j \geqslant 1$, u_j is defined by

$$u_j = y_j - z_j.$$

Subtracting (5.2) and (5.4) gives

$$\dot{u}_j(t) = j(\beta - \gamma)u_j(t) - j\beta u_{j+1}(t) + \frac{1}{N}d_j(t), \tag{5.11}$$

where the initial condition is $u_j(0) = 0$.

Our next and final Lemma gives bounds on $u_m(t)$ and yields the basis of the proof of Theorem 5.2.

LEMMA 5.4 For any $T > 0$, there exist $m \in \mathbb{N}$ and $K_m > 0$, such that

$$|u_m(t)| \leqslant \frac{K_m}{N}, \quad \text{for all } t \in [0, T].$$

Proof. According to Lemma 5.3, we can find $m \in \mathbb{N}$, such that $u_m(t) \geqslant 0$ and $u_{m+1}(t) \geqslant 0$ for all $t \in [0, T]$. Now let us consider (5.11) with $j = m$. Multiplying this equation by $\exp(-m(\beta - \gamma)t)$ and integrating from 0 to t gives

$$u_m(t)\mathrm{e}^{-m(\beta-\gamma)t} = -\beta m \int_0^t u_{m+1}(s)\mathrm{e}^{-m(\beta-\gamma)s}\mathrm{d}s + \frac{1}{N}\int_0^t d_m(s)\mathrm{e}^{-m(\beta-\gamma)s}\mathrm{d}s.$$

Combining that $u_{m+1}(t) \geqslant 0$ with the upper bound for d_m given in Proposition 5.2 results in the following inequality,

$$0 \leqslant u_m(t) \leqslant \frac{1}{N}\frac{(m-1)(\beta+\gamma)}{2(\beta-\gamma)}\mathrm{d}^{m(\beta-\gamma)t}.$$

Thus, the statement holds with $K_m = \frac{(m-1)(\beta+\gamma)}{2(\beta-\gamma)}\mathrm{e}^{m(\beta-\gamma)T}$. □

Now we are in the position to complete the proof of Theorem 5.2.

Proof of Theorem 5.2. Let us choose m and K_m according to Lemma 5.4. We prove by induction that for any $j = 1, 2, \ldots, m-1$, there exists K_j for which

$$|u_j(t)| \leqslant \frac{K_j}{N}, \quad \text{for all } t \in [0, T].$$

For $j = 1$, this together with Proposition 5.3 is exactly the statement of Theorem 5.2.

Let us assume that the statement is true for u_{j+1} and prove it for u_j. Multiplying (5.11) by $\exp(-j(\beta - \gamma)t)$ and integrating from 0 to t gives

$$u_j(t)\mathrm{e}^{-j(\beta-\gamma)t} = -\beta j \int_0^t u_{j+1}(s)\mathrm{e}^{-j(\beta-\gamma)s}\,\mathrm{d}s + \frac{1}{N}\int_0^t d_j(s)\mathrm{e}^{-j(\beta-\gamma)s}\,\mathrm{d}s.$$

Combining that $|u_{j+1}(t)| \leqslant K_{j+1}/N$ with the upper bound for d_j given in Proposition 5.2 results in the following inequality:

$$|u_j(t)| \leqslant K_j/N \quad \text{with } K_j = \frac{2\beta K_{j+1} + (j-1)(\beta+\gamma)}{2(\beta-\gamma)}\mathrm{e}^{j(\beta-\gamma)T}.$$

□

6. Discussion

Understanding the link between exact stochastic and mean-field approximation models is a challenging problem that arise often in applied research and when formulated rigorously can lead to difficult theoretical questions. Identifying the theoretical link between different modelling paradigms, such as stochastic versus ODE- or PDE-based models, requires the concurrent use of a number of different mathematical techniques. For example, Theorem 3.1 combines PDE elements with the discretization theorem for PDEs which is mainly used in Numerical Analysis. At the same time, Theorem 4.1 uses Martingale (see (4.2)) and/or semigroup theory. The concurrent use of different mathematical tools may limit the applicability of these results or can make it non-trivial to check if the assumptions of the theoretical results hold.

This paper makes two main contributions. First, it provides a unifying framework for the existing proofs and discusses the exact way in which convergence of the exact to the mean-field model holds. On the the other hand, we propose a novel proof that only relies on ODE-techniques and thus increase the transparency of our results and makes it more accessible to applied researchers. The main idea of our proof is the use of all moments of the distribution. This enabled us to keep the system exact and formulate convergence results to an approximation model based on the simplest form of moment closures. Our results rely on perturbation methods for infinite ODE systems and allowed us to theoretically identify the link between the exact model and moment closure models often derived based on heuristic arguments.

It is worth noting that the simplest method, the PDE-based approach, leads to the point-wise convergence of the expected value, while the stochastic method yields the stronger convergence in the sense that convergence also holds for the distribution. Our main result proves the uniform convergence of the expected value that in some sense lies between the two existing approaches. The technique presented in this paper could lead to further developments on several different fronts. For example, the most natural extension could be to generalize the link between exact stochastic and approximation models for networks other than fully connected or to check the validity of existing moment closure techniques that so far have only been tested via numerical simulations. At the same time, the results presented in the paper could also be extended for general dynamics and in the context of applied areas other than ecology and epidemiology.

Funding

P.L.S. acknowledges support from OTKA (81403). Funding from the European Union and the European Social Fund is also acknowledged (financial support to the project under TÁMOP-4.2.1/B-09/1/KMR). I.Z.K. acknowledges support from EPSRC (EP/H001085/1).

References

BALL, F. & NEAL, P. (2008) Network epidemic models with two levels of mixing. *Math. Biosci.*, **212**, 69–87.

BANSAL, S., GRENFELL, B. T. & MEYERS, L. A. (2007) When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface*, **4**, 879–891.

DIEKMANN, O. & HEESTERBEEK, J. A. P. (2000) *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Chichester, UK: John Wiley.

ETHIER, S. N. & KURTZ, T. G. (2005) *Markov Processes: Characterization and Convergence*. John Wiley.

HOUSE, T. & KEELING, M. J. (2011) Insights from unifying modern approximations to infections on networks. *J. R. Soc. Interface*, **8**, 67–73.

HUNDSDORFER, W. & VERWER, J. G. (2003) *Numerical Solution of Time-Dependent Advection-Diffusion-Reactions Equations*. Berlin: Springer.

KEELING, M. J. (1999) The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. Lond. B*, **266**, 859–867.

KEELING, M. J. & EAMES, K. T. D. (2005) Networks and epidemic models. *J. R. Soc. Interface*, **2**, 295–307.

KISS, I. Z., GREEN, D. M. & KAO, R. R. (2006) The effect of network heterogeneity and multiple routes of transmission on final epidemic size. *Math. Biosci.*, **203**, 124–136.

KURTZ, T. G. (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probabl.*, **7**, 49–58.

KURTZ, T. G. (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Probabl.*, **8**, 344–356.

KURTZ, T. G. (1980) *Biological Growth and Spread*. Lecture Notes on Biomathematics, vol. 38. Berlin: Springer, pp. 449–467.

LINDQUIST, J., MA, J., VAN DEN DRIESSCHE, P. & WILLEBOORDSE, F. H. (2010) Effective degree network disease models. *J. Math. Biol.*, **62**, 143–164.

MORENO, Y., PASTOR-SATORRAS, R. & VESPIGNANI, A. (2002) Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B*, **26**, 521–529.

RAND, D. A. (1999) Correlation equations for spatial ecologies. *Advanced Ecological Theory* (J. McGlade ed.). Oxford: Blackwell, 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.

SIMON, P. L., TAYLOR, M. & KISS, I. Z. (2011) Exact epidemic models on graphs using graph automorphism driven lumping. *J. Math. Biol.*, **62**, 479–508.

SPORNS, O. & KÖTTER, R. (2004) Motifs in brain networks. *Plos Biol.*, **2**, e369, 1910–1918.

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

VAN BAALEN, M. (2000) Pair approximations for different spatial geometries. *The Geometry of Ecological Interactions: Simplifying Complexity* (U. Dieckmann, R. Law, J. A. J. Metz eds). Cambridge, UK: Cambridge University Press, pp. 359–387.

VOLZ, E. & MEYERS, L. A. (2007) Susceptible-infected-recovered epidemics in dynamic contact networks. *Proc. R. Soc. B*, **274**, 2925–2933.

VOLZ, E. & MEYERS, L. A. (2009) Epidemic threshold in dynamic contact networks. *J. Roy. Soc. Interface*, **6**, 233–241.

Appendix A

First, we prove Lemma 5.1. This together with Theorem 5.2 yields the proof of Theorem 5.1.

Proof of Lemma 5.1. Since system (5.4) is linear and homogeneous, it is enough to prove that the only solution with zero initial condition is the constant zero function.

The system is autonomous hence it is enough to prove that the statement is true on a time interval of length T, that is $z_j(t_0) = 0$ for all j implies z_j is constant zero on $[t_0, t_0 + T]$. This result can then be extended using induction to show that z_j is constant zero on the intervals $[kT, (k+1)T]$ for all $k \in \mathbb{N}$. Thus, it is sufficient to prove that there exists $T > 0$, such that $z_j(0) = 0$ for all j implies that z_j is constant zero on $[0, T]$.

Multiplying (5.4) by $\exp(-j(\beta - \gamma)t)$, introducing $v_j(t) = z_j(t)\exp(-j(\beta - \gamma)t)$ and denoting $\beta - \gamma$ by α leads to the following differential equation for v_j:

$$\dot{v}_j(t) = -j\beta v_{j+1}(t)\mathrm{e}^{\alpha t}. \tag{A.1}$$

It is useful to show now that conditions $v_j(0) = 0$ and $v_j(t) \leqslant M$ for all j imply that there exists $T > 0$, such that $v_j(t) = 0$ on the time interval $[0, T]$.

Integrating (A.1) and using the initial condition $v_j(0) = 0$ gives

$$v_j(t) = -j\beta \int_0^t v_{j+1}(s)\mathrm{e}^{\alpha s}\,\mathrm{d}s. \tag{A.2}$$

This equation can be used iteratively and v_1 can be expressed in terms of v_{j+1} as

$$v_1(t) = (-1)^j j \frac{\beta^j}{\alpha^{j-1}} \int_0^t v_{j+1}(s)\mathrm{e}^{\alpha s}(\mathrm{e}^{\alpha t} - \mathrm{e}^{\alpha s})^{j-1}\,\mathrm{d}s. \tag{A.3}$$

This statement can be proved by induction using (A.2). Let us now choose a number T such that $\beta(\exp(\alpha T) - 1)/\alpha < 1$. Then, for all $t \leqslant T$ and for all $s \in [0, t]$, the following inequality holds, $\beta(\exp(\alpha t) - \exp(\alpha s))/\alpha < 1$. Therefore, the right-hand side of (A.3) can be estimated by $jq^j \cdot$ constant, where $q = \beta(\exp(\alpha T) - 1)/\alpha < 1$ since v_j is uniformly bounded. Thus, the right-hand side tends to zero when taking the limit $j \to \infty$. Hence, $v_1(t) = 0$ for all $t \in [0, T]$.

Using (A.1) with $j = 1$ gives that $v_2(t) = 0$ also holds for all $t \in [0, T]$. Similarly, by induction, it follows that $v_j(t) = 0$ for all $t \in [0, T]$ and for all $j \in \mathbb{N}$. This completes the proof. □

Now we prove two Propositions that were used in the proof of Lemma 5.2.

PROPOSITION A.1 For any positive numbers c and d and for all $t_0 \in (0, 1/d)$, there exists $j_0 \in \mathbb{N}$ such that for all $j \geqslant j_0$ and $t \in [0, t_0]$, the inequality $(1 + ct)^{-j} \leqslant 1 - dt$ holds.

Proof. Let $f(t) = 1/(1 - dt)$ and $g(t) = (1 + ct)^j$. We will prove that there exists $j_0 \in \mathbb{N}$ such that for all $j \geqslant j_0$ and $t \in [0, t_0]$, the inequality $f(t) \leqslant g(t)$ holds. Since $f(0) = 1 = g(0)$, it is enough to prove that $f'(t) \leqslant g'(t)$ for all $t \in [0, t_0]$. We have that

$$f'(t) = \frac{d}{(1 - dt)^2} \leqslant \frac{d}{(1 - dt_0)^2}$$

and

$$g'(t) = jc(1 + ct)^{j-1} \geqslant jc.$$

Hence, choosing a number j_0 to satisfy

$$\frac{d}{(1 - dt_0)^2} \leqslant j_0 c,$$

it follows that for all $j \geqslant j_0$ and $t \in [0, t_0]$

$$f'(t) \leqslant \frac{d}{(1 - dt_0)^2} \leqslant j_0 c \leqslant jc \leqslant g'(t).$$

□

PROPOSITION A.2 Let $c > 0$ and $d > 1$. Then for all $t_0 \in (0, (d - 1)/dc)$, there exists $j_0 \in \mathbb{N}$ such that for all $j \geqslant j_0$ and $t \in [0, t_0]$, the inequality $(1 - ct)^{-j} \leqslant 1 + d^j t$ holds.

Proof. Let $f(t) = (1 - ct)^{-j}$ and $g(t) = 1 + d^j t$. We will prove that there exists $j_0 \in \mathbb{N}$ such that for all $j \geqslant j_0$ and $t \in [0, t_0]$, the inequality $f(t) \leqslant g(t)$ holds. Since $f(0) = 1 = g(0)$, it is enough to prove that $f'(t) \leqslant g'(t)$ for all $t \in [0, t_0]$. We have that

$$f'(t) = jc(1 - ct)^{-j-1} \leqslant jc(1 - ct_0)^{-j-1}$$

and

$$g'(t) = d^j.$$

The assumption $t_0 \in (0, (d - 1)/dc)$ implies $d(1 - ct_0) > 1$, hence there exists a number j_0 for which

$$j\frac{c}{1 - ct_0} \leqslant d^j(1 - ct_0)^j, \qquad \text{for all } j \geqslant j_0.$$

Thus, we get that for all $j \geqslant j_0$ and $t \in [0, t_0]$

$$f'(t) \leqslant \frac{jc}{(1 - ct_0)^{j+1}} \leqslant d^j = g'(t)$$

holds. □