DIFFERENTIAL EQUATION APPROXIMATIONS OF STOCHASTIC NETWORK PROCESSES: AN OPERATOR SEMIGROUP APPROACH

ANDRÁS BÁTKAI, ISTVAN Z. KISS, ESZTER SIKOLYA, AND PÉTER L. SIMON

ABSTRACT. The rigorous linking of exact stochastic models to mean-field approximations is studied. Starting from the differential equation point of view the stochastic model is identified by its Kolmogorov equations, which is a system of linear ODEs that depends on the state space size (N) and can be written as $\dot{u}_N = A_N u_N$. Our results rely on the convergence of the transition matrices A_N to an operator A. This convergence also implies that the solutions u_N converge to the solution u of $\dot{u} = Au$. The limiting ODE can be easily used to derive simpler mean-field-type models such that the moments of the stochastic process will converge uniformly to the solution of appropriately chosen mean-field equations. A bi-product of this method is the proof that the rate of convergence is $\mathcal{O}(1/N)$. In addition, it turns out that the proof holds for cases that are slightly more general than the usual density dependent one. Moreover, for Markov chains where the transition rates satisfy some sign conditions, a new approach for proving convergence to the mean-field limit is proposed. The starting point in this case is the derivation of a countable system of ordinary differential equations for all the moments. This is followed by the proof of a perturbation theorem for this infinite system, which in turn leads to an estimate for the difference between the moments and the corresponding quantities derived from the solution of the mean-field ODE.

1. INTRODUCTION

General birth-and-death models are at the basis of many real-world applications ranging from queuing theory to disease transmission models, see Grimmett and Stirzaker [9]. In particular, the analysis of such models involves the consideration of Kolmogorov equations that simply describe the evolution of the probability of a certain process being in a given state at a given time. One of the major drawbacks of this approach is the large number of equations. This is very limiting from an analysis viewpoint, and in addition, it also precludes the construction of a numerical solution of the full or complete set of equations. Using techniques such as lumping in Simon et al. [18] this can be circumvented and an equivalent exact system with a tractable number of equations can be derived. However, often this technique only works in the presence of significant system symmetries such as in the case of a simple disease transmission model on a fully connected graph where all nodes are topologically identical. This requirement rarely holds and highlights the importance of approaches that deal with the complexity of the large number of equations. Progress in this direction has been made as illustrated by the important contributions of Kurtz [8], Bobrowski [4] and Darling and Norris [6].

Here we take a dynamical system type approach, where the Kolmogorov equations are simply considered as a system of linear ODEs with a transition rate matrix having specific properties such as special tri-diagonal structure and/or well defined

functional form for the transmission rates. For example, consider a Markov chain
with finite state space $\{0, 1, \ldots, N\}$ and denote by $p_k(t)$ the probability that the
system is in state k at time t (with a given initial state that is not specified at the
moment). Assuming that starting from state k the system can move to either state
$k - 1$ or to state $k + 1$, the Kolmogorov equations of the Markov chain take the
form

(KE) $$\dot{p}_k = \beta_{k-1} p_{k-1} - \alpha_k p_k + \delta_{k+1} p_{k+1}, \quad k = 0, \ldots, N,$$

or, introducing the tri-diagonal matrix

$$A_N := \begin{pmatrix} -\alpha_0 & \delta_1 & 0 & \cdots & \cdots & 0 \\ \beta_0 & -\alpha_1 & \delta_2 & \cdots & \cdots & 0 \\ 0 & \beta_1 & -\alpha_2 & \delta_3 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_{N-2} & -\alpha_{N-1} & \delta_N \\ 0 & 0 & \cdots & 0 & \beta_{N-1} & -\alpha_N \end{pmatrix}$$

and the coloumn vector $p(t) = (p_0(t), p_1(t), \ldots, p_N(t))^T$,

$$\dot{p}(t) = A_N p(t).$$

The assumption on the tri-diagonality of the matrix can obviously be weakened,
however, most practical problems that motivate our work fall into this class. There-
fore, to be in line with applications and to make our results more transparent, the
tri-diagonal case will be considered.

Here we assume that the coefficients β_k and δ_k are *asymptotically density dependent*
in the sense that

(1.1) $$\beta_k = B_N(k), \quad \delta_k = D_N(k),$$

and the following limits exist for all $x \in [0, 1]$,

(1.2) $$\beta(x) = \lim_{N \to \infty} \frac{B_N(Nx)}{N}, \quad \delta(x) = \lim_{N \to \infty} \frac{D_N(Nx)}{N},$$

with β and δ (at least) continuous functions, and

(1.3) $$\alpha_k = \beta_k + \delta_k, \quad k = 0, \ldots, N.$$

We note that the usually used density dependence means that $\beta(x) = \frac{B_N(Nx)}{N}$ for
all N (similarly for D_N).

In order to get a differential equation approximation of the Markov chain for $N \to \infty$, the random variables $X_N(t)$, forming a continuous time Markov process, with
values in

$$\left\{ 0, \frac{1}{N}, \frac{2}{N}, \ldots, 1 \right\}$$

are considered. Then

$$p_k(t) = P\left(X_N(t) = \frac{k}{N} \right)$$

can be expressed in terms of the transition probabilities as

$$p_k(t) = \sum_{j=0}^{N} p_{j,k}(t) p_j(0),$$

where the transition probabilities are

$$p_{j,k}(t) = P\left(X_N(t) = \frac{k}{N} \,\middle|\, X_N(0) = \frac{j}{N} \right).$$

Combining this with the Kolmogorov Equation (KE), it is straightforward to show that the transition matrix is given by

$$T_N(t) := [p_{j,k}(t)] = e^{A_N^\top t}.$$

As a simple calculation shows in Subsection 3.b, the approximating mean-field differential equation to the Markov chain can be written as

$$(1.4) \qquad \dot{x} = \beta(x) - \delta(x).$$

The problem of rigorously linking exact stochastic models to mean-field approximations goes back to the early work of Kurtz [13, 14]. Kurtz studied pure-jump density dependent Markov processes where apart from providing a method for the derivation of the mean-field model he also used solid mathematical arguments to prove the stochastic convergence of the exact to the mean-field model. His earlier results [13, 14] relied on Trotter type approximation theorems for operator semigroups. Later on, the results were embedded in the more general context of Martingale Theory [8]. A detailed survey of the subject from the probabilistic point of view is by Darling and Norris [6]. Building on this and similar work, McVinish and Pollett [16] have recently extended the differential equation approximation to the case of heterogeneous density dependent Markov chains, where the coefficients of the transition rates may vary with the nodes. The results by Kurtz and others in this area have been cited and extensively used by modelers in areas such as ecology and epidemiology to justify the validity of heuristically formulated mean-field models. The existence of several approximation models, often 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, see House and Keeling [11]. Some steps in these directions have been made by Ball and Neal, and Lindquist el al. [2, 15], where the authors clearly state the link between exact and mean-field models.

The probabilistic approach given by Ethier and Kurtz, and Darling and Norris in [8, 6] yields weak or stochastic convergence of the Markov chain to the solution of the differential equation. Here we have a more moderate goal, namely to prove that the expected value of the Markov process converges to the solution of the differential equation as $N \to \infty$, and to prove that the discrepancy between the two is of order $1/N$. The benefit of framing the question in this simpler or different way lies in a less technical and more accessible proof. This will mainly rely on well-known results from semigroup theory compared to the combination of highly specialist tools and results from probability theory. For the applied semigroup methods see also Bátkai, Csomós and Nickel [3] and references therein. This simpler approach, based on the expected value of the Markov chain, is also motivated by practical considerations, namely that usually the goodness of the approximation is tested by comparing the average of many individual simulations to the output from the simplified approximate model. This can be a satisfactory and sufficient comparison since, according to our knowledge, weak and stochastic converges is rarely tested.

The main result of the paper can be formulated in the following Theorem, where we assume uniformity for the convergence in (1.2).

Theorem 1. *Let*

$$y_1(t) = \sum_{k=0}^{N} \frac{k}{N} p_k(t)$$

be the expected value and let x be the solution of (1.4) with initial condition $x(0) = y_1(0)$. Let us assume that the limits in (1.2) are uniform in the sense that there

exists a number L, such that for all $x \in [0, 1]$

$$(1.5) \qquad |\beta(x) - \frac{B_N(Nx)}{N}| \leq \frac{L}{N}.$$

Then for any $t_0 > 0$ there exists a constant C, such that

$$|x(t) - y_1(t)| \leq \frac{C}{N}, \ t \in [0, t_0].$$

Since the essence of the Theorem is well-known we highlight the **novelties of our approach in the paper**.

- The proof is self-contained in the sense that no general, abstract, theorem or combination of theorems are used. The result is based on the simple fact that if the operators A_N converge to the operator A in a certain sense as $N \to \infty$, then the operator semigroup T_N generated by A_N converges to the semigroup T generated by A. This result is formulated in Lemma 5.
- The proof automatically implies the rate of convergence, namely it can be shown that the difference between the expected value and the solution of the mean-field ODE is of order $1/N$.
- Our tools make it possible to extend the above convergence results from density dependent Markov chains to the more general case of asymptotically density dependent Markov chains.
- For Markov chains where the transition rates satisfy some specific sign conditions, a completely new approach is presented. This is based on deriving a countable system of ordinary differential equations for the moments of a distribution of interest and proving a perturbation theorem for this infinite system.

The paper is structured as follows. In Section 2 we motivate our work via two examples and/or applications: an adaptive network model with random link activation and deletion, and a SIS type epidemic model on a static graph. The derivation of ODEs for the moments is presented in Section 3 together with the heuristic construction of the mean-filed equation for the first moment. In Section 4, we present our new approach and use it to prove Theorem 1. In Section 5, we present the derivation of an infinite system of ODEs for the moments (only in the density dependent case) and we show how this leads to a new approach that can be used to derive estimates for all moments, directly from the ODE. This is contrast with the usual approach where only the expected value of the Markov chain is estimated.

2. Motivation

In this Section we present two important examples that motivate our investigations. Recently, it has become more and more important to understand the relation between the dynamics on a network and the dynamics of the network, see Gross and Blasius [10]. In the case of epidemic propagation on networks it is straightforward to assume that the propagation of the epidemic has an effect on the structure of the network. For example, susceptible individuals try to cut their links in order to minimize their exposure to infection. This leads to a change in network structure which in turn impacts on how the epidemic spreads. The first step in modeling this phenomenon is an appropriate dynamic network model such as the recently proposed globally-constrained Random Link Activation-Deletion (RLAD) model.

This can be described in terms of Kolmogorov equations as follows,

$$\frac{p_k(t)}{dt} = \alpha(N - (k-1))\left(1 - \frac{k-1}{K_1^{max}}\right)p_{k-1}(t)$$
$$- \left[\alpha(N-k)\left(1 - \frac{k}{K_1^{max}}\right) + \omega k\right]p_k(t) + \omega(k+1)p_{k+1}(t),$$
$$k = 0, \ldots, N,$$

where $p_k(t)$ denotes the probability that at time t there are k activated links in the network, and N is the total number of potential edges. It is assumed that non-active links are activated independently at random at rate α and that existing links are broken independently at random at rate ω. Furthermore, the link creation is globally constrained by introducing a carrying capacity K_1^{max}, that is the network can only support a certain number of edges as given by K_1^{max}.

Using the above notation, here

$$\beta_k = \alpha(N-k)\left(1 - \frac{k}{K_1^{\max}}\right), \quad \delta_k = \omega k, \quad \alpha_k = \beta_k + \delta_k, \quad k = 0, \ldots, N,$$
$$\alpha_{-1} = \delta_{N+1} = 0$$

with $K_1^{\max}$ being of order N. These coefficients clearly satisfy (1.2) and (1.3).

The second motivation comes from epidemiology where a paradigm disease transmission model is the simple susceptible-infected-susceptible (SIS) model on a completely connected graph with N nodes, i.e. all individuals are connected to each other. From the disease dynamic viewpoint, each individual is either susceptible (S) or infected (I) – the susceptible ones can be infected at a certain rate (β) if linked to at least one infected individual and the infected ones can recover at a given rate (γ) and become susceptible again. It is known that in this case the 2^N-dimensional system of Kolmogorov equations can be lumped to a $N+1$-dimensional system, see Simon, Taylor and Kiss [18].

The lumped Kolmogorov equations take again the form (KE) with

$$\beta_k = \beta k(N-k)/N, \quad \delta_k = \gamma k, \quad \alpha_k = \beta_k + \delta_k, \quad k = 0, \ldots, N,$$
$$\beta_{-1} = \delta_{N+1} = 0.$$

These coefficients also satisfy (1.2) and (1.3). We note that in the case of a homogeneous random graph we get a similar system with a slightly different meaning of the coefficients.

3. Momentum approach

The basic idea of getting an approximating differential equation is to calculate the time derivative of the expected value by using the Kolmogorov equations. Since the obtained equation is not self-contained, it needs to be closed by using some closure approximation. In this Section we derive first equations for the derivatives of every moment, then we briefly show how to get the simplest mean-field approximation for the first moment. This is also discussed in the case of asymptotically density dependent Markov chains.

3.a. Differential equations for the moments. Introducing the moments

$$(3.1) \qquad\qquad y_n(t) = \sum_{k=0}^{N}\left(\frac{k}{N}\right)^n p_k(t), \quad n = 1, 2, \ldots,$$

($y_1(t)$ is the expected value we are mainly interested in) we follow Simon and Kiss [17] to derive the differential equations for $y_n(t)$ starting from the Kolmogorov equations (KE). To get the time derivative of y_n the following Lemma will be used.

Lemma 2. *Let r_k ($k = 0, 1, 2, \ldots$) be a sequence and let $r(t) = \sum_{k=0}^{N} r_k p_k(t)$, where $p_k(t)$ is given by (KE). Then*

$$\dot{r} = \sum_{k=0}^{N} (\beta_k(r_{k+1} - r_k) + \delta_k(r_{k-1} - r_k)) p_k.$$

Proof. From (KE) we obtain

$$\dot{r} = \sum_{k=0}^{N} r_k \dot{p}_k = \sum_{k=1}^{N} r_k \beta_{k-1} p_{k-1} - \sum_{k=0}^{N} r_k \alpha_k p_k + \sum_{k=0}^{N-1} r_k \delta_{k+1} p_{k+1}$$

$$= \sum_{k=0}^{N-1} r_{k+1} \beta_k p_k - \sum_{k=0}^{N} r_k \alpha_k p_k + \sum_{k=1}^{N} r_{k-1} \delta_k p_k.$$

Using that $\beta_N = 0$, $\delta_0 = 0$ and $\alpha_k = \beta_k + \delta_k$ we get

$$\dot{r} = \sum_{k=0}^{N} (r_{k+1}\beta_k - r_k(\beta_k + \delta_k) + r_{k-1}\delta_k) p_k = \sum_{k=0}^{N} (\beta_k(r_{k+1} - r_k) + \delta_k(r_{k-1} - r_k)) p_k.$$

$\square$

Before applying Lemma 2 with $r_k = (k/N)^n$, it is useful to define the following two new expressions

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

Let us introduce

$$(3.2) \qquad d_n(t) = \sum_{k=0}^{N} (\beta_k R_{k,n} + \delta_k Q_{k,n}) p_k(t).$$

Combining these with Lemma 2 leads to

$$\dot{y}_n(t) = \sum_{k=0}^{N} \left(\frac{\beta_k}{N} \left(n\frac{k^{n-1}}{N^{n-1}} + R_{k,n} \right) + \frac{\delta_k}{N} \left(-n\frac{k^{n-1}}{N^{n-1}} + Q_{k,n} \right) \right) p_k(t)$$

hence

$$(3.3) \qquad \dot{y}_n(t) = n \cdot \sum_{k=0}^{N} \frac{\beta_k - \delta_k}{N} \cdot \left(\frac{k}{N} \right)^{n-1} \cdot p_k(t) + \frac{1}{N} d_n(t).$$

Using the binomial theorem $R_{k,n}$ and $Q_{k,n}$ can be expressed in terms of the powers of k, hence d_n can be expressed as $d_n(t) = \sum_{m=1}^{n} d_{nm} y_m(t)$ with some coefficients d_{nm}. The d_n terms contain N, hence to use the $1/N \to 0$ limit it has to be shown that d_n remains bounded as N goes to infinity. This is proved in the next lemma.

Lemma 3. *For the functions d_n the following estimates hold*

$$0 \le d_n(t) \le c \cdot \frac{n(n-1)}{2} \qquad \text{for all } t \ge 0.$$

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 $R_{k,n}$ and $Q_{k,n}$. In particular, applying the above result when $f(x) = x^n$, $x = k + 1$ and $x_0 = k$ gives

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

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

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

with some $\eta \in [k, k+1]$. Hence, $R_{k,n}$ and $Q_{k,n}$ are non-negative yielding that $d_n(t) \geq 0$. On the other hand, $\xi/N \leq 1$ and $\eta/N \leq 1$ and (1.1) lead to the inequality given below

$$\beta_k R_{k,n} + \delta_k Q_{k,n} \leq \frac{n(n-1)}{2}\left(\frac{\beta_k}{N} + \frac{\delta_k}{N}\right) \leq c \cdot \frac{n(n-1)}{2}.$$

Hence, the statement follows immediately from (3.2) and using that $\sum_{k=0}^{N} p_k(t) = 1$. □

We show two possible ways to turn (3.3) into an ODE. First, we use the approximation $E(F(X)) = F(E(X)$ (where E stands for the expected value and F is a given measurable function) to derive the mean-field approximation. Then assuming that the Markov chain is density dependent and the functions β and δ are polynomials, we derive an infinite system of ODEs for the moments.

3.b. **The mean-field approximation.** Since $d_1 = 0$, applying (3.3) for $n = 1$ we obtain

$$(3.4) \qquad \dot{y}_1(t) = \sum_{k=0}^{N}\left(\frac{B_N(k)}{N} - \frac{D_N(k)}{N}\right) \cdot p_k(t).$$

Using the asymptotic density dependence of the Markov chain (1.2), the right-hand side can be approximated by

$$\sum_{k=0}^{N}\left(\beta\left(\frac{k}{N}\right) - \delta\left(\frac{k}{N}\right)\right) \cdot p_k(t).$$

In order to make the equation "closed" the approximation

$$(3.5) \quad \sum_{k=0}^{N}\left(\beta\left(\frac{k}{N}\right) - \delta\left(\frac{k}{N}\right)\right) \cdot p_k(t) \approx \beta\left(\sum_{k=0}^{N}\frac{k}{N} \cdot p_k(t)\right) - \delta\left(\sum_{k=0}^{N}\frac{k}{N} \cdot p_k(t)\right)$$

will be used. Substituting this approximation into the equation (3.4) we obtain the following differential equation

$$(\text{ODE}) \qquad\qquad \dot{x} = \beta(x) - \delta(x).$$

This equation is known as the *mean-field approximation* of the original Kolmogorov equation (KE).

3.c. **Infinite system of ODEs in the polynomial case.** In this Subsection it is assumed that the functions in (1.1) are polynomials and the Markov chain is density dependent, that is,

$$(3.6) \qquad \frac{\beta_k}{N} = \beta\left(\frac{k}{N}\right), \quad \frac{\delta_k}{N} = \delta\left(\frac{k}{N}\right)$$

and

$$(3.7) \qquad \beta(x) = \sum_{j=0}^{l} g_j x^j, \qquad \delta(x) = \sum_{j=0}^{l} h_j x^j.$$

Using this and denoting

$$(3.8) \qquad q_j := g_j - h_j, \quad j = 0, 1, \ldots, l,$$

from (3.3) we obtain that

$$\dot{y}_n(t) = n \cdot \sum_{k=0}^{N} \sum_{j=0}^{l} q_j \left(\frac{k}{N}\right)^{n+j-1} \cdot p_k(t) + \frac{1}{N} d_n(t)$$

$$(3.9) \qquad = n \cdot \sum_{j=0}^{l} q_j y_{n+j-1}(t) + \frac{1}{N} d_n(t)$$

with

$$(3.10) \qquad 0 \le d_n(t) \le c \cdot \frac{n \cdot (n-1)}{2}.$$

Letting $N \to \infty$ on the right-hand-side, we arrive at the following system

$$\dot{f}_n(t) = n \cdot \sum_{j=0}^{l} q_j f_{n+j-1}(t)$$

$$(\text{IE}) \qquad n = 1, 2, \ldots,$$

that can be regarded as a system of "approximating" differential equations for (3.9). In Section 5 we are going to investigate how $y_1(t)$ can be approximated on finite time intervals using the solution of this infinite system.

Remark 4. All the results for the infinite system obtained from here on remain valid in the asymptotically density dependent case when

$$\frac{\beta_k}{N} = \sum_{j=0}^{l} \tilde{g}_j(N) k^j, \qquad \frac{\delta_k}{N} = \sum_{j=0}^{l} \tilde{h}_j(N) k^j$$

and

$$\tilde{g}_j(N) = \frac{g_j}{N^j} + \mathcal{O}\left(\frac{1}{N^{j+1}}\right), \quad \tilde{h}_j(N) = \frac{h_j}{N^j} + \mathcal{O}\left(\frac{1}{N^{j+1}}\right), \quad j = 0, \ldots, l.$$

The only difference is that in (3.9) we obtain

$$\dot{y}(t) = n \cdot \sum_{j=0}^{l} q_j y_{n+j-1}(t) + \frac{1}{N} d_n(t) + \mathcal{O}\left(\frac{1}{N}\right).$$

4. Proof of Theorem 1

In this Section we prove that the solution of (ODE) is an $\mathcal{O}(1/N)$ approximation of the expected value of the Markov chain, that is we prove Theorem 1.

Let us introduce the matrix A_N as in Section 1

$$
A_N := \begin{pmatrix}
-\alpha_0 & \delta_1 & 0 & \cdots & \cdots & 0 \\
\beta_0 & -\alpha_1 & \delta_2 & \cdots & \cdots & 0 \\
0 & \beta_1 & -\alpha_2 & \delta_3 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \beta_{N-2} & -\alpha_{N-1} & \delta_N \\
0 & 0 & \cdots & 0 & \beta_{N-1} & -\alpha_N
\end{pmatrix}.
$$

Then the operator families $(T_N(t))_{t \geq 0}$ defined as

$$
T_N(t) := [p_{j,k}(t)] = \mathrm{e}^{A_N^\top t}
$$

form uniformly continuous semigroups on $\mathbb{C}^{N+1}$ for each $N \in \mathbb{N}$. This yields

$$
(4.1) \qquad (T_N(t)f)\left(\frac{j}{N}\right) = \sum_{k=0}^{N} f\left(\frac{k}{N}\right) \cdot p_{j,k}(t)
$$

for $f = (f_0, \ldots, f_N) \in \mathbb{C}^{N+1}$ where we make the identification

$$
\mathbb{C}^{N+1} \equiv \left\{ f : f \text{ maps } \{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1\} \text{ to } \mathbb{R} \right\}.
$$

Assume for the sake of simplicity that on the right-hand-side of (ODE) the functions $\beta, \delta \in C^2[0,1]$. If we denote the solution of (ODE) with initial condition x_0 by $\varphi(t, x_0)$, then the operator family defined as

$$
(4.2) \qquad (T(t)f)(x_0) := f(\varphi(t, x_0)), \quad f \in C([0,1]), \ t \geq 0
$$

defines a strongly continuous operator semigroup on $C([0,1])$ (see Engel, Nagel [7, Section 3.28]) with generator $(A, D(A))$. We also know that for $f \in C^1([0,1])$

$$
(Af)(x_0) = (\beta(x_0) - \delta(x_0)) \cdot f'(x_0).
$$

The main idea is to approximate the semigroup $(T_N(t))_{t \geq 0}$ (that is, the solution of the transposed (KE)) using the semigroup $(T(t))_{t \geq 0}$ (that is, the solution of the mean-field equation (ODE)). Observe that the semigroups act on different spaces: the first one acts on $X_N := \mathbb{C}^{N+1}$, the second one on $X := C^1([0,1])$. In order to prove an approximation Theorem in a fixed space, we assume that there are linear operators

$$
(4.3) \qquad J_N : X_N \to X, \quad J_N(f) := g,
$$

$$
(4.4) \qquad P_N : X \to X_N, \quad P_N(g) := f \text{ with } f\left(\frac{k}{N}\right) = g\left(\frac{k}{N}\right), \ k = 0, \ldots, N
$$

such that $\|J_N\| \leq 1$, $\|P_N\| \leq 1$, $N \in \mathbb{N}$, and

$$
P_N J_N = I_{X_N}, \quad N \in \mathbb{N};
$$

$$
J_N P_N f \to f, \quad N \to \infty \ \forall f \in X
$$

are satisfied (see Bátkai et al. [3, Definition 3.5]). The next Lemma formulates the main approximation result in rigorous terms.

Lemma 5. *Assume that the conditions of Theorem 1 are satisfied. For $(T_N(t))_{t\geq 0}$ and $(T(t))_{t\geq 0}$ the following holds: for all $f \in C^2([0,1])$ and $t_0 > 0$ there exists $C = C(f, t_0) > 0$ such that for all $t \in [0, t_0]$*

$$(4.5) \qquad \|(P_N T(t) - T_N(t) P_N)f\| \leq \frac{C}{N},$$

where P_N denotes the projection defined in (4.4).

Proof. For the generators A_N, A, and for $f \in C^2([0,1])$ the following identities hold:

$$(P_N A f)\left(\frac{k}{N}\right) = \left(\beta\left(\frac{k}{N}\right) - \delta\left(\frac{k}{N}\right)\right) \cdot f'\left(\frac{k}{N}\right);$$

$$(A_N P_N f)\left(\frac{k}{N}\right) = \frac{\beta_k}{N} \cdot \frac{f(\frac{k+1}{N}) - f(\frac{k}{N})}{\frac{1}{N}} - \frac{\delta_k}{N} \cdot \frac{f(\frac{k}{N}) - f(\frac{k-1}{N})}{\frac{1}{N}},$$

$k = 0, \ldots, N$. The idea of proving the estimate for the semigroups is to estimate the difference of the generators. Their difference can be divided into two parts as follows.

$$(P_N A f)\left(\frac{k}{N}\right) - (A_N P_N f)\left(\frac{k}{N}\right) = \left(\beta\left(\frac{k}{N}\right) - \frac{\beta_k}{N} - \delta\left(\frac{k}{N}\right) + \frac{\delta_k}{N}\right) f'\left(\frac{k}{N}\right)$$

$$+ \frac{\beta_k}{N}\left(f'(\frac{k}{N}) - \frac{f(\frac{k+1}{N}) - f(\frac{k}{N})}{\frac{1}{N}}\right) + \frac{\delta_k}{N}\left(\frac{f(\frac{k}{N}) - f(\frac{k-1}{N})}{\frac{1}{N}} - f'(\frac{k}{N})\right).$$

The first part can be estimated by using the asymptotic density dependence (1.5) as

$$(4.6) \qquad \left|\beta\left(\frac{k}{N}\right) - \frac{\beta_k}{N}\right| = \left|\beta\left(\frac{k}{N}\right) - \frac{B_N(N\frac{k}{N})}{N}\right| \leq \frac{L}{N}$$

and

$$(4.7) \qquad \left|\delta\left(\frac{k}{N}\right) - \frac{\delta_k}{N}\right| = \left|\delta\left(\frac{k}{N}\right) - \frac{D_N(N\frac{k}{N})}{N}\right| \leq \frac{L}{N}$$

and using that f' is bounded.

The second part can be estimated by using Taylor's formula for $f \in C^2([0,1])$. We obtain that for each $k = 0, \ldots, N$ there exists $\xi_k \in (\frac{k}{N}, \frac{k+1}{N})$ such that

$$(4.8) \qquad \left|\frac{f(\frac{k+1}{N}) - f(\frac{k}{N})}{\frac{1}{N}} - f'(\frac{k}{N})\right| = \left|N \cdot \frac{f''(\xi_k)}{2} \cdot \frac{1}{N^2}\right| \leq \frac{1}{2N}\|f''\|.$$

Since β and δ are bounded on $[0,1]$ (they are in $C^2[0,1]$), we obtain from (4.6) and (4.7) that $\frac{\beta_k}{N}$, $\frac{\delta_k}{N}$ are uniformly bounded for all $k = 0, \ldots, N$ and $N \in \mathbb{N}$. Hence, for all $f \in C^2([0,1])$ there exists $K > 0$ such that

$$(4.9) \qquad \|(P_N A - A_N P_N)f\| \leq \frac{K}{N}\|f''\| + \frac{L}{N}\|f'\|.$$

It is easy to see that for all N, the operators $\widetilde{T}_N(t) := J_N T_N(t) P_N$, $t \geq 0$ form a strongly continuous semigroup on X – where J_N is defined in (4.3) – with generator $\widetilde{A}_N := J_N A_N P_N$.

Using the variation of parameters formula (see e.g. Engel, Nagel [7, Corollary III.1.7]), for the difference of the (projected) semigroups, and for $f \in C^2([0,1])$

we obtain that

$$(P_N T(t) - T_N(t) P_N) f = P_N \left(T(t)f - \widetilde{T}_N(t)f \right)$$

$$= \int_0^t P_N \widetilde{T}_N(t-s) \left(A - \widetilde{A}_N \right) T(s)f \, ds$$

$$= \int_0^t T_N(t-s) \left(P_N A - A_N P_N \right) T(s)f \, ds.$$

By the estimate in (4.9) and using that $T(t)$ maps the subspace $C^2([0,1])$ of X into itself (see Chicone [5, Theorem 1.3]), we obtain that for $f \in C^2([0,1])$ there exist $K^* > 0$ and $L^* > 0$, such that

$$\|(P_N T(t) - T_N(t) P_N) f\| \leq \int_0^t \frac{K^*}{N} \cdot \|(T(s)f)''\| + \frac{L^*}{N} \cdot \|(T(s)f)'\| \, ds.$$

Since $\beta, \delta \in C^2([0,1])$, the solution function φ of (ODE) is also a C^2-function of the initial data. Hence, for $s \in [0, t]$ there exists $K_t > 0$ such that

$$\|(T(s)f)''\| = \|(f \circ \varphi(s, \cdot))''\| = \|(f'' \circ \varphi) \cdot (\varphi')^2 + (f' \circ \varphi) \cdot \varphi''\| \leq K_t \cdot \|f\|_{C^2}$$

and similarly $\|(T(s)f)'\| \leq K_t^* \cdot \|f\|_{C^2}$. This yields that for $f \in C^2([0,1])$ and $t_0 > 0$ there exists $C = C(f, t_0) > 0$ such that for all $t \in [0, t_0]$

$$\|(P_N T(t) - T_N(t) P_N) f\| \leq \frac{C}{N}.$$

$\square$

Using this semigroup approximation Lemma we can now prove Theorem 1.

Proof of Theorem 1. We apply Lemma 5 for $f = \mathrm{id}_{[0,1]}$. We obtain that for any $t_0 > 0$ there exists $C = C(\mathrm{id}, t_0)$ such that for all $t \in [0, t_0]$

(4.10) $$\|(P_N T(t) - T_N(t) P_N) \, \mathrm{id}\| \leq \frac{C}{N}.$$

Observe that by (3.1) and (4.1), the following relation holds

$$y_1(t) = \sum_{k=0}^N \frac{k}{N} \cdot p_k(t) = \sum_{k=0}^N \frac{k}{N} \cdot \sum_{j=0}^N p_{j,k}(t) \cdot p_j(0)$$

$$= \sum_{j=0}^N p_j(0) \sum_{k=0}^N \mathrm{id}(\frac{k}{N}) \cdot p_{j,k}(t) = \sum_{j=0}^N p_j(0) \cdot (T_N(t) P_N \mathrm{id}) \, (\frac{j}{N})$$

$$= \langle p(0), T_N(t) P_N \mathrm{id} \rangle.$$

Furthermore, using (4.2) it is easy to show that

$$(P_N T(t) \mathrm{id}) \, (\frac{k}{N}) = \varphi(t, \frac{k}{N}).$$

It can be seen that it is enough to prove the statement when the initial condition is $p_m(0) = 1$, $p_j(0) = 0$, $j \neq m$. Then $y_1(0) = \frac{m}{N}$ yielding $x(0) = \frac{m}{N}$ and therefore $x(t) = \varphi \left(t, \frac{m}{N} \right)$. Hence, combining the above facts we obtain

$$|y_1(t) - x(t)| = \left| \langle p(0), T_N(t) P_N \mathrm{id} \rangle - \varphi(t, \frac{m}{N}) \right|$$

$$= \left| (T_N(t) P_N \mathrm{id}) \, (\frac{m}{N}) - (P_N T(t) \mathrm{id}) \, (\frac{m}{N}) \right| \leq \frac{C}{N}$$

where we used (4.10). $\square$

5. Infinite system of ODEs for the moments

In this Section we consider the infinite system of ODEs (3.9) for the moments and its formal limit as $N \to \infty$ (IE). Our aim here is to prove that the solutions of the first system converge to those of the second as $N \to \infty$ and t is in a bounded time interval. That is, that the moments of the Markov chain converge to the solutions of the infinite system (IE). As a bi-product of this result we get a completely new proof of Theorem 1 under extra sign conditions on the transition rates.

In order to prove this general convergence result were are going to use the approach of Kato [12] and Banasiak et al. [1].

Lemma 6. *Denote the formal operator*

$$(\mathcal{L}f)_n := n \cdot \sum_{j=0}^{l} q_j f_{n+j-1}, \quad n = 1, 2, \ldots.$$

Let

$$\mathcal{D}(L_{\max}) := \left\{ f \in \ell^1 : \mathcal{L}f \in \ell^1 \right\}, \quad L_{\max} := \mathcal{L}|\mathcal{D}(L_{\max}).$$

Assume that $q_0, q_2, \ldots, q_l \geq 0$, $q_1 \leq 0$ *and*

(5.1) $$q_0 + q_1 + \cdots + q_l \leq 0;$$

(5.2) $$q_0 - 2q_3 - 3q_4 - 4q_5 - \cdots - (l-1)q_l \leq 0.$$

Then there exists an operator (L, D) *such that* $L \subset L_{\max}$ *and* (L, D) *is the generator of a positive strongly continuous semigroup* $(T(t))_{t \geq 0}$ *of contractions on* ℓ^1.

Proof. Building the infinite coefficient matrix for $\mathcal{L}$, the assumptions imply that all the column sums are less or equal than 0:

(5.3) $(n+1)q_0 + nq_1 + (n-1)q_2 + (n-2)q_3 + \cdots + (n-l+1)q_l \leq 0, \quad n = 1, 2, \ldots.$

Hence, using the assumptions on the signs of the q_j's, we can apply Theorem 1 of Kato in [12]. Though the assumptions of Kato's theorem are referred to the equality instead of the inequality in (5.3), the proof works without any changes in our case. Thus, we do not repeat the proof here. $\qquad \square$

Let us turn to our original question on the expected value $y_1(t)$ – that is, the first coordinate of the solution of (3.9).

Theorem 7. *Assume that the conditions of Lemma 6 are satisfied. Then there exists an appropriate Banach space of sequences* ℓ_w^1 *with norm* $\| \cdot \|_w$ *such that for any initial condition* $y_0 \in \ell_w^1$ *the solution* $y(t) := (y_n(t))_{n \in \mathbb{N}}$ *of (3.9) and the solution* $f(t) := (f_n(t))_{n \in \mathbb{N}}$ *of (IE) satisfy*

(5.4) $$\|y(t) - f(t)\|_w \leq \frac{K}{N} \quad \text{for } t \in [0, T],$$

that is the solution of (3.9) tends to the solution of (IE) in a finite time interval as $N \to \infty$.

Proof. Using the formal operator $\mathcal{L}$ as in Lemma 6, the system (3.9) can be written as

$$\dot{y}(t) = \mathcal{L}y(t) + \frac{1}{N}d(t),$$

(5.5) $$y(0) = y_0,$$

where $d(t) := (d_n(t))_{n \in \mathbb{N}}$. Denote by ℓ_w^1 a weighted ℓ^1 space of sequences such that $d(t) \in \ell_w^1$, $t \in \mathbb{R}_+$, e.g.,

$$\|x\|_w := \sum_{n=1}^{\infty} |x_n| r^n \text{ for } x = (x_n)_{n \in \mathbb{N}} \text{ with } |r| < 1.$$

In this case, by (3.10),

$$\|d(t)\|_w \le c \cdot \sum_{n=1}^{\infty} n(n-1)r^n =: K < \infty, \quad t \in \mathbb{R}_+.$$

Applying Lemma 6 on the space ℓ_w^1, we obtain an operator (L_w, D_w) – acting formally as $\mathcal{L}$ – which is the generator of a contraction semigroup $(T_w(t))_{t \ge 0}$ on ℓ_w^1. By the formula of variation of parameters for (5.5) (see, e.g., Engel and Nagel [7, Corollary III.1.7]), we have

$$(5.6) \qquad y(t) = T_w(t)y_0 + \frac{1}{N} \cdot \int_0^t T_w(t-s)d(s)\,ds$$

$$(5.7) \qquad = f(t) + \frac{1}{N} \cdot \int_0^t T_w(t-s)d(s)\,ds$$

for any $y_0 \in \ell_w^1$. Thus

$$(5.8) \qquad \|y(t) - f(t)\|_w \le \frac{1}{N} \cdot \int_0^t \|T_w(t-s)\|_w \cdot \|d(t)\|_w \, ds \le \frac{1}{N} \cdot t \cdot K,$$

and this implies (5.4). $\qquad\square$

Remark 8. We note that the sign conditions of Lemma 6 hold in the case of globally constrained random link activation-deletion process, but they do not hold in the case of an SIS epidemic on a complete graph.

Acknowledgments

References

[1] J. Banasiak, N. Lachowicz and N. Moszyński, *Semigroups for generalized brith-and-death equations in ℓ^p spaces.* Semigroup Forum **73** (2006), 175–193.

[2] F. Ball and P. Neal, *Network epidemic models with two levels of mixing.* Math. Biosci. **212** (2008), 69–87.

[3] A. Bátkai, P. Csomós and G. Nickel, *Operator splittings and spatial approximations for evolution equations.* J. Evol. Equ. **9** (2009), 613-636.

[4] A. Bobrowski, *Functional Analysis for Probability and Stochastic Processes. An Introduction.* Cambridge, 2005.

[5] C. Chicone, *Ordinary Differential Equations with Applications.* Texts in Applied Mathematics, 34. Springer, New York, 2006.

[6] R. W. R. Darling and J. R. Norris, *Differential equation approximations for Markov chains.* Probab. Surv. **5** (2008), 37-79.

[7] K.-J. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations.* Graduate Texts in Math., vol. 194, Springer-Verlag, 2000.

[8] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence.* John Wiley & Sons Ltd, USA, 2005.

[9] G. Grimmett and D. Stirzaker, *Probability and Random Processes.* Oxford University Press, Oxford, 2001.

[10] T. Gross and B. Blasius, *Adaptive coevolutionary networks: a review.* J. Roy. Soc. Interface **5** (2008), 259–271.

[11] T. House and M. J. Keeling, *Insights from unifying modern approximations to infections on networks.* J. R. Soc. Interface **8** (2011), 67–73.

[12] T. Kato, *On the semi-groups generated by Kolmogoroff's differential equations.* J. Math. Soc. Japan **6** (1954), 1–15.

[13] T. G. Kurtz, *Extensions of Trotter's operator semigroup approximation theorems.* J. Functional Analysis **3** (1969), 354–375.

[14] T. G. Kurtz, *Solutions of ordinary differential equations as limits of pure jump Markov processes.* J. Appl. Prob. **7** (1970), 49–58.

[15] J. Lindquist, J. Ma, P. van den Driessche and F. H. Willeboordse, *Effective degree network disease models.* J. Math. Biol. **62** (2011), 143–164.

[16] R. McVinish and P. K. Pollett, *The deterministic limit of heterogeneous density dependent Markov chains.* Ann. Appl., Prob., submitted.

[17] P. L. Simon and I. Z. Kiss, *From exact stochastic to mean-field ODE models: a case study of three different approaches to prove convergence results.* to appear.

[18] P. L. Simon, N. Taylor and I. Z. Kiss, *Exact epidemic models on graphs using graph automorphism driven lumping.* J. Math. Biol. **62** (2011), 479–508.

Eötvös Loránd University Budapest, Hungary

School of Mathematical and Physical Sciences, Department of Mathematics, University of Sussex, Falmer, Brighton BN1 9RF, UK
E-mail address: batka@cs.elte.hu, seszter@cs.elte.hu, simonp@cs.elte.hu
E-mail address: i.z.kiss@sussex.ac.uk