State space collapse for critical multistage epidemics

Florian Simatos

STATE SPACE COLLAPSE FOR CRITICAL MULTISTAGE EPIDEMICS

FLORIAN SIMATOS

ABSTRACT. We study a multistage epidemic model which generalizes the SIR model and where infected individuals go through $K \geq 1$ stages of the epidemic before being removed. An infected individual in stage $k \in \{1,\dots,K\}$ may infect a susceptible individual, who directly goes to stage k of the epidemic; or it may go to the next stage $k+1$ of the epidemic. For this model, we identify the critical regime in which we establish diffusion approximations. Surprisingly, the limiting diffusion exhibits an unusual form of state space collapse which we analyze in detail.

CONTENTS

1. INTRODUCTION

A multistage model. In this paper we study an epidemic model proposed by Antal and Krapivsky [2] which generalizes the classical SIR model. Similarly as for the SIR model, there is a closed population where each individual is either *susceptible*, *infected* or *removed*. In addition, we assume that the disease which is spread progresses through $K \geq 1$ stages. For short, we will say that an (infected) individual is in stage $k \in \{1,\dots,K\}$ if it possesses the disease in stage k. It will also be convenient to say that an individual is in stage 0 if it is susceptible and in stage $K+1$ if it is removed.

Two types of transitions may occur within our model: either (1) an infected individual in stage $k \in \{1,\dots,K\}$ tries to infect a susceptible individual; or (2) an infected individual in stage $k \in \{1,\dots,K\}$ progresses to the next stage $k+1$. We consider the mean-field regime where in case (1) above, the infected individual samples an individual uniformly at random from the whole population: if the sampled individual is susceptible, then it becomes infected and starts the infection at the same stage than the individual who infected him or her.

In view of its epidemiological interpretation and because it generalizes the SIR model (which corresponds to the single-stage case $K=1$), it is natural to interpret this model as

DEPARTMENT OF MATHEMATICS & COMPUTER SCIENCE, EINDHOVEN UNIVERSITY OF TECHNOLOGY, P.O. BOX 513, 5600 MB EINDHOVEN, THE NETHERLANDS

E-mail address: f.simatos@tue.nl.

Date: July 10, 2014.

an epidemic model. From that viewpoint, our model differs from multistage epidemic models previously proposed. For HIV/AIDS for instance, Hyman et al. [10] proposed a multistage (or staged-progression) model where a newly infected individual starts the epidemic in stage one. We believe that the techniques used in the present paper can be adapted to study this case as well.

Besides its natural epidemiological interpretation, this model may also bear insight into other areas of applied probability. We may for instance think of cell population dynamic, e.g., when a cancer tumor progresses, cells accumulate mutations that are passed on to newly infected cells. In this case the stage of an individual corresponds to the number of mutations. Another potential application of this model is in communication networks: in this case we may think of users accumulating information and, upon meeting with a user with no information, sharing all the information the first user has. Such kind of epidemic spreading of information is actually at the heart of modern peer-to-peer systems.

Scaling limits. Epidemic models can broadly be categorized into discrete and continuous ones, see for instance Kendall [13]. Discrete models are individual-based models that focus on the inherent stochasticity of the dynamic, and are often described by finite state space Markov processes. Continuous models correspond to large population approximations: they may be either deterministic (corresponding to some notion of averaging in the large population regime) or stochastic. In the first case, the continuous model is typically described by ordinary differential equations (ODE) and in the second case, by stochastic differential equations (SDE).

In the present paper, we bridge these two viewpoints by starting from a discrete model which, after a suitable renormalization procedure, leads to a continuous model. This is a classical approach which allows to understand the extent to which a continuous model, usually more tractable, suitably approximates a discrete one. In addition to giving insight into how the epidemic develops over time, such schemes are also useful to understand the final outbreak size of the epidemic. In the SIR case for instance, this makes it possible to express the final outbreak size, properly renormalized, as the hitting time of 0 by a Brownian motion with a parabolic drift; this allows in turn for explicit computation, see for instance Martin-Löf [18]. Note that our forthcoming Theorem 2.3 generalizes this result to the multistage case $K \geq 1$.

For the SIR model it is well-known that different scaling limits may be obtained depending on the discrete model's parameters, see for instance von Bahr and Martin-Löf [23]. In particular, the limit may be deterministic or retain some stochasticity of the original discrete model, and the intermediate regime at which transition occurs is usually referred to as *critical regime*: this is the regime in which we are interested in in the present paper. In this regime, scaling limits of the SIR model have for instance been obtained by Martin-Löf et al. [17, 18, 23] and by Aldous [1] in the closely related context of the Erdös-Rényi random graph.

A popular method to establish such scaling limits relies on semi-group techniques, see for instance Ethier and Kurtz [8]. This approach has for instance recently been followed by Dolgoarshinnykh and Lalley [7] to study the related SIS model, who also sketch a proof for the SIR model. However, because of the different boundary conditions between the SIS and the SIR models, the authors only mention that different analytical results from the theory of parabolic partial differential equations are needed in the SIR case. In the present paper we use a different and more probabilistic approach, closer to Aldous' approach [1], that relies on semimartingale arguments. These arguments

rely on elementary calculations made on the infinitesimal generator of the Markov process, and require almost no analytical results. This is the reason why we believe that this method is robust and can be adapted to study other models such as the HIV/AIDS model of Hyman et al. [10].

Presentation and discussion of the main results. Let us present our results in the particular case where all individual transition rates are equal to one: in general these will be allowed to be *close* to one. For $n \geq 1$ we consider the $K+2$-dimensional Markov process $\mathbf{a}_n(t) = (a_{n,k}(t), k = 0, \ldots, K+1)$ that represents a multistage epidemic process in a population of size n. It is defined by an initial state $\mathbf{a}_n(0) = (a_{n,k}(0), k = 0, \ldots, K+1)$ satisfying the relation $a_{n,0}(0) + \cdots + a_{n,K+1}(0) = n$ and, for $a = (a_0, a_1, \ldots, a_{K+1}) \in \mathbb{N}^{K+2}$, by the following transition rates:

$$a \longrightarrow \begin{cases} a - e_k + e_{k+1} & \text{at rate } a_k, \ 1 \leq k \leq K, \\ a - e_0 + e_k & \text{at rate } a_k a_0 / n, \ 1 \leq k \leq K, \end{cases} \tag{1.1}$$

where $e_0 = (1, 0, \ldots, 0)$, $e_1 = (0, 1, 0, \ldots, 0)$, etc. For $t \geq 0$, $a_{n,k}(t)$ is the number of individuals in stage k at time t, i.e., $a_{n,0}(t)$ is the number of susceptible individuals; $a_{n,k}(t)$ for $k = 1, \ldots, K$ is the number of infected individuals in stage k; and $a_{n,K+1}(t)$ is the number of removed individuals. The transition rates (1.1) preserve the total population size, so that the relation $a_{n,0}(t) + \cdots + a_{n,K+1}(t) = n$ is satisfied for any $t \geq 0$.

These rates formalize the description of the dynamic given above: each infected individual progresses to the next stage of the epidemic at rate one; each infected individual makes an infection attempt at rate one, which is successful with probability proportional to the number of susceptible individuals (which gives the factor a_0 / n).

Consider the initial condition where $a_{n,1}(0) = n^\beta$ for some $\beta \in (0,1)$ and $a_{n,k}(0) = 0$ for $k = 2, \ldots, K+1$. The asymptotic behavior of the $K+1$-dimensional process $(a_{n,k}, k = 1, \ldots, K+1)$ as $n \to +\infty$ is then governed by the precise value of β, and there are three different regimes:

Small initial condition $\beta < 1/(K+2)$: on the time scale n^β, $a_{n,k}$ lives on the space scale $n^{k\beta}$ and the scaling limit is governed by the following SDE:

$$\begin{cases} \mathrm{d}A_1(t) = (2A_1(t))^{1/2} \mathrm{d}B(t), \\ \mathrm{d}A_k(t) = A_{k-1}(t)\mathrm{d}t \qquad \text{for } \ k = 2, \ldots, K+1; \end{cases} \tag{1.2}$$

Intermediate initial condition $\beta = 1/(K+2)$: on the time scale $n^{1/(K+2)}$, $a_{n,k}$ lives on the space scale $n^{k/(K+2)}$, the outbreak size is of the order of $n^{(K+1)/(K+2)}$ and the scaling limit is governed by the following SDE:

$$\begin{cases} \mathrm{d}A_1(t) = -A_{K+1}(t)A_1(t)\mathrm{d}t + (2A_1(t))^{1/2}\mathrm{d}B(t), \\ \mathrm{d}A_k(t) = (A_{k-1}(t) - A_{K+1}(t)A_k(t))\mathrm{d}t \qquad \text{for } \ k = 2, \ldots, K, \\ \mathrm{d}A_{K+1}(t) = A_K(t)\mathrm{d}t; \end{cases} \tag{1.3}$$

Large initial condition $\beta > 1/(K+2)$: on the time scale n^γ with $\gamma = (1-\beta)/(K+1)$, $a_{n,k}$ lives on the space scale $n^{\beta+(k-1)\gamma}$ and the scaling limit is governed by the following ODE:

$$\begin{cases} \mathrm{d}A_1(t) = -A_{K+1}(t)A_1(t)\mathrm{d}t, \\ \mathrm{d}A_k(t) = (A_{k-1}(t) - A_{K+1}(t)A_k(t))\mathrm{d}t \quad \text{for } \ k = 2, \ldots, K, \\ \mathrm{d}A_{K+1}(t) = A_K(t)\mathrm{d}t. \end{cases} \tag{1.4}$$

In the above, B is a standard Brownian motion. Moreover, since the total population size is fixed, the asymptotic behavior of $a_{n,0}$ is recovered from the behavior of the $a_{n,k}$'s: since $n - a_{n,0} = a_{n,1} + \cdots + a_{n,K+1}$ and since in all three regimes $a_{n,K+1} \gg a_{n,k}$ for $k = 1,\dots,K$, $n - a_{n,0}$ obeys the same scaling and has the same limit than $a_{n,K+1}$.

It is also interesting to note that the intermediate regime interpolates between the small and large ones in two ways: (1) the scaling at play there can be obtained by letting $\beta \uparrow 1/(K+2)$ in the small regime or $\beta \downarrow 1/(K+2)$ in the large regime, and (2) the evolution of A_1 in (1.3) is a mixture of (1.2) and (1.4). However, β does not appear in the asymptotic dynamics (1.2)–(1.4) and so (1.3) appears somehow discontinuously.

When specializing the results of the intermediate regime to $K = 1$ we recover the classical result on the SIR model [23]: starting with of the order of $n^{1/3}$ infected individuals, we end up with an outbreak size of the order of $n^{2/3}$. When $K = 2$, these results show that starting with $n^{1/4}$ individuals in the first stage, the number of individuals in the second stage during the outbreak is of the order $n^{1/2}$, while the total outbreak size is of the order of $n^{3/4}$.

Note that for the above comparison we have referred to the work by von Bahr and Martin-Löf [23] which actually considers a discrete-time version of the SIR model, the so-called Reed-Frost model. In this model, individuals stay infected for a deterministic amount of time and, at the end of this period, try to infect every other individual with a fixed probability. In the case $K = 1$ the two models are indeed equivalent, at least with respect to the final outbreak size: heuristically, it does not matter when and by whom a given individual gets infected. However, in the truly multistage case $K \geq 2$ this is no longer the case: the stage of an individual that spreads the disease matters. In that respect, individuals from different stages "compete" for the pool of susceptible individuals and it is natural to consider a continuous-time model. We suspect that, compared to the case $K = 1$, this additional difficulty is causing the discrepancy discussed in Section 6.

In order to get insight into the scalings at play, let us consider the continuous-time Markov branching process $\mathbf{a}^B = (a_k^B, k = 1,\dots,K+1)$ with $K+1$ types, whose transition rates for $a \in \mathbb{N}^{K+1}$ are given by

$$a \longrightarrow \begin{cases} a - e_k + e_{k+1} & \text{at rate } a_k,\ 1 \leq k \leq K, \\ a + e_k & \text{at rate } a_k,\ 1 \leq k \leq K. \end{cases} \tag{1.5}$$

These rates are obtained from (1.1) by setting $a_0 = n$: this corresponds to an infinite population limit, where an infection attempt is always successful. Then a_1^B is a continuous-time Markov branching process (also known as Bellman–Harris branching process), and it can be seen that if $a_1^B(0) = n^\beta$ and time is sped up by n^β, then it converges to an interesting object (Feller diffusion). Thus if we are to have an interesting limit for a_1^B, the correct time scale (for all processes) needs to be n^β.

But then, the total number individuals it ever begets is of the order of $\int_0^\infty a_1^B(t)\mathrm{d}t$ which is of the order of $n^{2\beta}$. These individuals form the initial condition for a_2^B, which can thus be seen as a Bellman–Harris branching process, started at $n^{2\beta}$ but for which time is sped up by n^β. However, we know from the above discussion that in order for a_2^B to evolve on the space scale $n^{2\beta}$, it needs to be sped up in time by $n^{2\beta}$: thus on the time scale n^β, a_2^B remains essentially constant. However, on this time scale it begets of the order of $\int_0^{n^\beta} a_2^B(t)\mathrm{d}t \propto n^{3\beta}$ individuals, which form the basis for a_3^B. Iterating this argument, we see the geometric progression observed in all three regimes appearing, where

$a_{n,k+1}$ is of the order of the time scale times $a_{n,k}$. But actually, more can be learned from this simple approximation.

First of all, we see that the approximation of $\mathbf{a}_n$ by $\mathbf{a}^B$ is asymptotically exact in the small regime, in the sense that (1.2) is also the scaling limit of $\mathbf{a}^B$ subject to the same scaling. This approximation begins to break down in the intermediate regime, which is therefore the regime where *finite-size population effects* begin to kick in. This phenomenon is also sometimes called the *depletion of points effect*: in the intermediate regime the epidemic begins to feel the decreasing number of susceptible individuals.

Moreover, we have just argued that we are considering a time scale which is suited for $a_{n,1}$ but not for $a_{n,2}$. More precisely, on the time scale where $a_{n,1}$ evolves, $a_{n,2}$ has not enough time to evolve on its own and all its randomness comes from $a_{n,1}$. This problem of time scales leads to a classical phenomenon in queueing theory, called *state space collapse*, but which is more surprising to find in the context of epidemic.

An unusual form of state space collapse. Let $\bar{K} = K-1$ and T_n be the first time at which $a_{n,1}$ hits zero, i.e., $T_n = \inf\{t \geq 0 : a_{n,1}(t) = 0\}$. It is then important to observe that the $\bar{K}$+2-dimensional process $\bar{\mathbf{a}}_n = (\bar{a}_{n,k}, k = 0, \dots, \bar{K}+1)$ obtained by keeping track of the individuals in stages $0, 2, \dots, K+1$ after time T_n, i.e., with $\bar{a}_{n,0}(t) = a_{n,0}(T_n + t)$ and $\bar{a}_{n,k}(t) = a_{n,k+1}(T_n + t)$ for $k = 1, \dots, \bar{K}+1$, is a multistage epidemic process with $\bar{K}$ stages. However, we know that in the intermediate regime, $a_{n,2}$, and thus $\bar{a}_{n,1}(0)$, is of the order of $n^{2/(K+2)} = n^{2/(\bar{K}+3)}$. As such, this corresponds to a multistage epidemic process with a large initial condition (with $\beta = 2/(\bar{K}+3) > 1/(\bar{K}+2)$), and we know from the above discussion that its scaling limit is deterministic. In summary, once the first stage has got extinct, the subsequent evolution of the process is deterministic, which supports the previous claim that all the randomness lies in the first stage. This property can actually be directly read off the SDE (1.3).

Indeed, a striking feature of this SDE is that the diffusion coefficients governing A_k for $k \neq 1$ are equal to zero: in particular, only the coordinate A_1 is truly random, the other ones being obtained deterministically from it via an ODE. More precisely, let $F : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}^K$ be the function such that, if $\bar{\mathbf{A}} = (A_2, \dots, A_{K+1})$, then (1.3) rewrites $\mathrm{d}\bar{\mathbf{A}}(t) = F(A_1(t), \bar{\mathbf{A}}(t))\mathrm{d}t$: then $\bar{\mathbf{A}}$ satisfies an ODE of the kind

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = F\big(y(t), x(t)\big), \ t \geq 0. \tag{1.6}$$

It will be seen that if $y : [0,\infty) \to [0,\infty)$ is continuous, then this ODE has a unique solution. This naturally defines a one-dimensional manifold $\mathscr{M}(y) \subset \mathbb{R}^K$, namely, if x_y is the unique solution to (1.6), the space $\mathscr{M}(y) = \big\{x_y(t) : t \geq 0\big\}$.

If $\mathbf{A} = (A_1, \dots, A_{K+1})$ is the solution to (1.3), then by definition $\bar{\mathbf{A}}$ is the unique solution to (1.6) with $y = A_1$, which justifies seeing $\bar{\mathbf{A}}$ as deterministically obtained from A_1. Thus we can see A_{K+1} as a function of A_1 and the initial condition $\bar{a} = \bar{\mathbf{A}}(0)$, and write $A_{K+1}(t) = \varphi(A_1, \bar{a}; t)$ for some deterministic map φ. With this notation, we see that A_1 on its own satisfies the following SDE

$$\mathrm{d}A_1(t) = \big(\gamma_1 - \varphi(A_1, \bar{a}; t)\big)\, A_1(t)\mathrm{d}t + (2A_1(t))^{1/2}\mathrm{d}B(t). \tag{1.7}$$

Such an SDE is sometimes referred to as an Itô process and is, in the terminology of Rogers and Williams [20], not of diffusion type since $\varphi(A_1, \bar{a}; t)$ depends on A_1 through the entire path $(A_1(s), 0 \leq s \leq t)$ and not only through the value of A_1 at time t.

In summary, we start with a sequence of K+2-dimensional stochastic processes (or K+1-dimensional, given that the total population size is fixed), but the diffusion $\mathbf{A}$ =

$(A_1,\dots,A_{K+1})$ appearing in the limit is truly one-dimensional: one coordinate, A_1, is random and given by (1.7), while the other coordinates $A_2,\dots,A_{K+1}$ are obtained deterministically from it via (1.6).

Such a phenomenon of reduction of the dimension of the state space in the critical (or near-critical) regime is well-known in queueing theory, where it is usually referred to as *state space collapse*. It has first been systematically investigated in [6, 24] in the context of multiclass queueing networks, and has since then been observed in various settings [5, 19, 21, 22]. In all these examples and similarly as what we observe in our model, the randomness of the limiting diffusion is contained in one coordinate (the workload process for queueing systems), while the other coordinates are obtained deterministically from it. In queueing theory typically, the workload process converges to a reflected Brownian motion W and the queue length processes $\mathbf{Q}$ are obtained via a deterministic map, e.g., $\mathbf{Q} = G \circ W$.

However, there are two main differences between these results from queueing theory and our present form of state space collapse. First of all, in all these queueing examples, the map G is linear, i.e., we can write $Q_k(t) = \alpha_k W(t)$ for some deterministic coefficients $\alpha_k > 0$. This makes the queue length processes $\mathbf{Q} = (Q_k)$ live in a *deterministic* one-dimensional manifold, namely, the space $\mathscr{M}_Q = \{(\alpha_k w) : w \geq 0\}$. In contrast, the coordinates $A_2,\dots,A_{K+1}$ live in our case in the *random* one-dimensional manifold $\mathscr{M}(A_1)$.

Moreover, the reasons leading to this state space collapse phenomenon are also distinct. As explained above, in our case this comes from a matter of time scales while in queueing theory, this essentially comes from the law of large numbers since the workload is the sum of the residual service times over the number of customers.

Organization of the paper. In the next section we introduce the general model considered in the present paper, which generalizes (1.1) to the near-critical case, and state the main results, which formalize the results described previously. In Section 3 we analyze the limiting SDE that generalizes (1.3), where we prove existence and uniqueness of solutions as well as a useful sample path property. These results rely on the analysis of the ODE that generalizes (1.6), and the deterministic results that are needed are proved in the Appendix A. Sections 4 and 5 are devoted to the proofs of the main results: Section 4 gives the proof for the scaling limits, i.e., the asymptotic behavior at the process level, and Section 5 is concerned with the proof of the asymptotic behavior of the outbreak size. Finally, we conclude the paper by discussing in Section 6 the conjecture formulated in [2], and how it relates to our results.

Acknowledgements. I am extremely grateful to Remco van der Hofstad and Johan van Leeuwaarden for introducing me to the topic of this paper and for very stimulating discussions on this particular model. I would also like to thank the two anonymous reviewers whose comments greatly helped to improve the exposition of the paper.

2. Main results

2.1. **Notation.** Let in the sequel $\mathscr{D}$ be the set of real-valued, càdlàg functions. For $f \in \mathscr{D}$ and $\varepsilon > 0$, we define $T_0(f) = \inf\{t \geq 0 : f(t) = 0\}$, $T_\varepsilon^{\uparrow}(f) = \inf\{t \geq 0 : f(t) \geq \varepsilon\}$ and $T_\varepsilon^{\downarrow}(f) = \inf\{t \geq 0 : f(t) \leq \varepsilon\}$, with the convention $\inf\emptyset = +\infty$. If $t \geq 0$ and f is càdlàg (real- or vector-valued) we consider θ_t the shift operator, defined by $\theta_t(f) = f(t + \cdot)$. The space of càdlàg functions with values in $\mathbb{R}^d$ is endowed with the topology of uniform convergence on compact sets.

In the sequel we fix some integer $K \geq 1$ and we consider the map $\bar{\pi} : \mathbb{R}^{K+2} \to \mathbb{R}^{K+1}$ defined by $\bar{\pi}(a_0, \ldots, a_{K+1}) = (a_0, a_2, \ldots, a_{K+1})$ (this unusual indexing of vectors will be convenient for our purposes).

2.2. **Model and main results.** We now present the full model investigated in the rest of the paper, which generalizes (1.1) by allowing individual transition rates to be close to one. For each $n \geq 1$, let $\mathbf{a}_n(t) = (a_{n,k}(t), k = 0, \ldots, K+1)$ be the $K+2$-dimensional Markov process corresponding to a finite population of size n, i.e., $a_{n,0}(0) + \cdots + a_{n,K+1}(0) = n$, and with non-zero transition rates given for $a = (a_0, a_1, \ldots, a_{K+1}) \in \mathbb{N}^{K+2}$ by

$$a \longrightarrow \begin{cases} a - e_k + e_{k+1} & \text{at rate } (1+\delta_{n,k})a_k, \ 1 \leq k \leq K, \\ a - e_0 + e_k & \text{at rate } (1+\varepsilon_{n,k})a_k a_0/n, \ 1 \leq k \leq K, \end{cases} \tag{2.1}$$

where $\delta_{n,k}, \varepsilon_{n,k} > -1$ for $k = 1, \ldots, K$, and we will always assume that $\delta_{n,k}, \varepsilon_{n,k} \to 0$ as $n \to +\infty$ for each $k = 1, \ldots, K$: the critical case (1.1) is recovered by setting $\delta_{n,k} = \varepsilon_{n,k} = 0$. Moreover, it will be convenient to define $\delta_{n,0} = \delta_{n,K+1} = \varepsilon_{n,K+1} = -1$. We first state the results concerning the intermediate regime, which constitute the main results of the paper.

Theorem 2.1 (Intermediate initial condition)**.** *Let $\beta = 1/(K+2)$ and assume that for each $k = 1, \ldots, K$, there exists $\gamma_k \in \mathbb{R}$ such that*

$$n^{\beta}\left(\varepsilon_{n,k} - \delta_{n,k}\right) \underset{n \to +\infty}{\longrightarrow} \gamma_k. \tag{2.2}$$

Let $\mathbf{A}_n = (A_{n,k}, k = 0, \ldots, K+1)$ be the $K+2$-dimensional process defined as follows:

$$A_{n,0}(t) = \frac{n - a_{n,0}(n^{\beta}t)}{n^{(K+1)\beta}} \ \text{ and } \ A_{n,k}(t) = \frac{a_{n,k}(n^{\beta}t)}{n^{k\beta}}, \ k = 1, \ldots, K+1, \ t \geq 0. \tag{2.3}$$

If $\mathbf{A}_n(0) \to a \in [0,\infty)^{K+2}$, then the sequence of processes $(\mathbf{A}_n, n \geq 1)$ converges weakly as $n \to +\infty$ to the unique solution $\mathbf{A} = (A_k, 0 \leq k \leq K+1)$ of the SDE

$$\mathbf{A}(t) = a + \int_0^t b(\mathbf{A}(u))\mathrm{d}u + \int_0^t \sigma(\mathbf{A}(u))\mathrm{d}B(u), \tag{2.4}$$

where B is a standard $K+2$-dimensional Brownian motion and for $a = (a_0, \ldots, a_{K+1}) \in [0,\infty)^{K+2}$ we have $\sigma(a) = \mathrm{diag}(0, (2a_1)^{1/2}, 0, \ldots, 0)$ and

$$\begin{cases} b_0(a) = b_{K+1}(a) = a_K, \\ b_1(a) = (\gamma_1 - a_0)a_1, \\ b_k(a) = a_{k-1} + (\gamma_k - a_0)a_k, \ k = 2, \ldots, K. \end{cases}$$

Since all the diffusion coefficients but one are equal to 0, when convenient we will identify the $(K+2) \times (K+2)$ matrix $\sigma(a)$ and its only non-zero entry $(2a_1)^{1/2}$. For the same reason, if we write $B = (B_0, \ldots, B_{K+1})$ then only the coordinate B_1 matters and we will identify B and B_1, i.e., whenever convenient we will consider that B is a one-dimensional standard Brownian motion; note that this is in line with the notation used in (1.2) and (1.3). We now state some properties of solutions to (2.4).

Proposition 2.2. *Consider the assumptions and notation of Theorem 2.1. Then $\mathbf{A}$ almost surely satisfies the following properties:*

- *(i) the two processes A_0 and A_{K+1} are equal;*
- *(ii) if $K \geq 2$, then A_0 is strictly increasing, and its terminal value $A_0(\infty)$ is finite and satisfies $A_0(\infty) > \gamma_k$ for every $k = 2, \ldots, K$;*
- *(iii) $T_0(A_1) < +\infty$;*
- *(iv) for every $k = 2, \ldots, K$, $A_k(t) \to 0$ as $t \to +\infty$ but $T_0(A_k) = +\infty$.*

Note that the first property of $\mathbf{A}$ above comes from the relation $\sum_{k=0}^{K+1} a_{n,k}(t) = n$, which translates after scaling to $\sum_{k=1}^{K+1} n^{k\beta} A_{n,k}(t) = n^{(K+1)\beta} A_{n,0}(t)$. Next, we turn to the asymptotic behavior of the outbreak size $A_{n,K+1}(\infty)$: the following result essentially states that we can interchange the limits $n \to +\infty$ and $t \to +\infty$.

Theorem 2.3. *Under the assumptions and notation of Theorem 2.1, the renormalized outbreak size $A_{n,K+1}(\infty)$ converges weakly as $n \to +\infty$ toward $A_{K+1}(\infty)$.*

Let us conclude the presentation of results in the intermediate regime by comparing with the SIR case $K = 1$. In the case $K = 1$ the natural notion of criticality is that every infected individual tries to infect in average (close to) one other individual. The notion of criticality that we consider here is at first sight different, since every infected individual tries to infect in average one other individual *in each stage*. If an individual starts in stage k, it will therefore make in average $K-k+1$ infection attempts. But the two notions actually coincide. Indeed, Theorem 2.1 shows that individuals in the last stage K of the epidemic dominate: there are $n^{K/(K+2)}$ such individuals and $n^{k/(K+2)} \ll n^{K/(K+2)}$ individuals at stage $k = 1, \dots, K-1$. Thus with overwhelming probability, an individual eventually infected has actually started the epidemic in stage K and has made in average one infection attempt. Beware however: this does not mean that only stage K matters, since $a_{n,K}$ becomes of the order of $n^{K/(K+2)}$ thanks to the help of the individuals in the previous stages!

Moreover, according to Proposition 2.2, the case $K = 1$ is the only case where $A_{K+1}(\infty)$ is reached in finite time. Indeed, for $K = 1$ the dynamic is frozen after the time $T_0(A_1)$ which is finite, while if $K \geq 2$ then A_{K+1} remains strictly increasing at all times. As will be discussed in Section 5 when proving Theorem 2.3, this difference creates an additional difficulty in order to control the asymptotic behavior of the outbreak size.

We now state the result for the large regime.

Proposition 2.4 (Large initial condition)**.** *Fix some sequence $n^{1/(K+2)} \ll \alpha_{n,1} \ll n$ and for $n \geq 1$ define*

$$\tau_n = \left(\frac{n}{\alpha_{n,1}}\right)^{1/(K+1)}, \ \alpha_{n,0} = \frac{n}{\tau_n} \ \text{and} \ \alpha_{n,k} = \tau_n^{k-1}\alpha_{n,1} \ \text{for} \ k = 2, \dots, K+1. \tag{2.5}$$

Assume that for each $k = 1, \dots, K$, there exists $\gamma_k \in \mathbb{R}$ such that

$$\tau_n \left(\varepsilon_{n,k} - \delta_{n,k}\right) \underset{n \to +\infty}{\longrightarrow} \gamma_k. \tag{2.6}$$

Let $\mathbf{A}_n = (A_{n,k}, k = 0, \dots, K+1)$ be the $K+2$-dimensional renormalized process defined as follows:

$$A_{n,0}(t) = \frac{n - a_{n,0}(\tau_n t)}{\alpha_{n,0}} \ \text{and} \ A_{n,k}(t) = \frac{a_{n,k}(\tau_n t)}{\alpha_{n,k}}, \ k = 1, \dots, K+1, \ t \geq 0. \tag{2.7}$$

If $\mathbf{A}_n(0) \to a \in [0,\infty)^{K+2}$, then the sequence of processes $(\mathbf{A}_n, n \geq 1)$ converges weakly as $n \to +\infty$ to the unique solution of the ODE

$$\mathbf{A}(t) = a + \int_0^t b(\mathbf{A}(u))\mathrm{d}u, \tag{2.8}$$

with b defined in Theorem 2.1.

Note that this result is coherent with Theorem 2.1, in the sense that the limit of $\theta_{T_0(A_{n,1})} \circ \bar{\pi} \circ \mathbf{A}_n = \bar{\pi} \circ \theta_{T_0(A_{n,1})} \circ \mathbf{A}_n$ is the same whether we consider this process as the

process $\bar{\pi} \circ \mathbf{A}_n$ shifted at time $T_0(A_{n,1})$ and then use Theorem 2.1, or whether we consider this process as a multistage epidemic process with $\bar{K} = K-1$ stages started from a large initial condition and then use Proposition 2.4.

We complete the results of Theorem 2.1 and Proposition 2.4 by studying the case of a small initial condition $a_{n,1}(0) \ll n^{1/(K+2)}$. When $a_{n,1}(0)$ converges to some finite number, the sequence of processes $(a_{n,k}, k = 1, \ldots, K+1)$ converges to the multitype branching process given by (1.5). As the next result shows, such a branching approximation continues to be valid in the small regime.

Proposition 2.5 (Small initial condition)**.** *Fix some sequence* $1 \ll \alpha_{n,1} \ll n^{1/(K+2)}$ *and assume that for each* $k = 1, \ldots, K$*, there exists* $\gamma_k \in \mathbb{R}$ *such that*

$$\alpha_{n,1}\left(\varepsilon_{n,k} - \delta_{n,k}\right) \underset{n \to +\infty}{\longrightarrow} \gamma_k. \tag{2.9}$$

Let $\mathbf{A}_n = (A_{n,k}, k = 0, \ldots, K+1)$ *for* $n \geq 1$ *be the* $K+2$*-dimensional renormalized process defined as follows:*

$$A_{n,0}(t) = \frac{n - a_{n,0}(\alpha_{n,1} t)}{\alpha_{n,1}^{K+1}} \text{ and } A_{n,k}(t) = \frac{a_{n,k}(\alpha_{n,1} t)}{\alpha_{n,1}^{k}},\ k = 1, \ldots, K+1,\ t \geq 0. \tag{2.10}$$

If $\mathbf{A}_n(0) \to a \in [0,\infty)^{K+2}$*, then the sequence of processes* $(\mathbf{A}_n, n \geq 1)$ *converges weakly as* $n \to +\infty$ *to the unique solution of the SDE*

$$\mathbf{A}(t) = a + \int_0^t b^S(\mathbf{A}(u))\mathrm{d}u + \int_0^t \sigma(\mathbf{A}(u))\mathrm{d}B(u), \tag{2.11}$$

where σ *is as in Theorem 2.1 and* b^S *is given by*

$$\begin{cases} b_0^S(a) = b_{K+1}^S(a) = a_K, \\ b_1^S(a) = \gamma_1 a_1, \\ b_k^S(a) = a_{k-1} + \gamma_k a_k, \ \ k = 2, \ldots, K. \end{cases}$$

As discussed in the introduction for the strictly critical case $\delta_{n,k} = \varepsilon_{n,k} = 0$, the limiting diffusion (2.11) obtained in the small regime is the same as the limit of the branching process corresponding to the infinite population setting, i.e., where $a_0/n = 1$ in (2.1) (this could be proved using the techniques of the present paper). Thus $a_{n,1}(0) \propto n^{1/(K+2)}$ is the threshold at which finite-size population effects (or depletion of points effects) begin to kick in: this is the threshold at which the branching approximation ceases to be valid.

Note also that for $\mathbf{A} = (A_k, k = 0, \ldots, K+1)$ satisfying (2.11), A_1 is Feller diffusion with drift γ_1, i.e., is the unique solution to the SDE $\mathrm{d}A_1 = \gamma_1 A_1 \mathrm{d}t + (2A_1)^{1/2}\mathrm{d}B$ (see (3.2) below), and then A_k for $k = 2, \ldots, K+1$ is given recursively by $A_k' = \gamma_k A_k + A_{k-1}$ (here and in the sequel, prime denotes differentiation with respect to the time variable). In particular, existence and uniqueness of solutions to (2.11) follow immediately.

The next three sections are devoted to proving these results. In the next section we prove uniqueness and existence to solutions of (2.4), and we also prove Proposition 2.2. The scaling limits results of Theorem 2.1 and Propositions 2.4 and 2.5 are proved in Section 4, and the proof of Theorem 2.3 on the asymptotic behavior of the outbreak size is given in Section 5.

3. Analysis of the SDE (2.4)

As mentioned in the introduction, the fact that $\sigma(a) = \mathrm{diag}(0, (2a_1)^{1/2}, 0, \dots, 0)$ is a manifestation of the state space collapse property, and it makes the process $\bar{\pi} \circ \mathbf{A}$ deterministically obtained from A_1 by an ODE. More precisely, we consider in the sequel $F : \mathbb{R} \times \mathbb{R}^{K+1} \to \mathbb{R}^{K+1}$ the function defined by $F(a_1, \bar{\pi}(a)) = \bar{\pi}(b(a))$ for $a \in \mathbb{R}^{K+2}$ and with b as in Theorem 2.1 (when all parameters $\gamma_k = 0$, this coincides with the function F in (1.6)). With this notation, we see that if $\mathbf{A}$ satisfies (2.4), then $\bar{\pi} \circ \mathbf{A}$ is a solution of the ODE

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = F(y(t), x(t)),\ x(0) = \bar{a}, \tag{3.1}$$

with $y = A_1$. The following properties of this ODE will be needed: their proof is postponed to the Appendix A.

Lemma 3.1. *For any $\bar{a} \in [0,\infty)^{K+1}$ and any continuous function $y : [0,\infty) \to [0,\infty)$, the ODE* (3.1) *has a unique solution defined on* $[0,\infty)$.

Lemma 3.2. *Fix $\bar{a} = (a_0, a_2, \dots, a_{K+1}) \in [0,\infty)^{K+1}$ and $y : [0,\infty) \to [0,\infty)$ a continuous function, and let $x = (x_0, x_2, \dots, x_{K+1})$ be the unique solution to* (3.1) *given by Lemma 3.1. Then the function x_0 is non-decreasing and its limit $x_0(\infty)$ as $t \to +\infty$ exists in* $[0,\infty]$. *Moreover:*

- *(i) if $y(0) > 0$ or $a_2 > 0$, then $x_k(t) > 0$ for every $k = 0, 2, \dots, K$ and $t > 0$;*
- *(ii) if $x_k(t) > 0$ for every $k = 2, \dots, K$ and $t > 0$ and x_0 is bounded, then $\int_0^\infty y < +\infty$ and moreover $\int_0^\infty x_k < +\infty$ and $x_0(\infty) > \gamma_k$ for every $k = 2, \dots, K$;*
- *(iii) if $y(t) = 0$ for all $t \geq 0$ and $a_2 > 0$, then x_0 is bounded and $x_k(t) \to 0$ as $t \to +\infty$ for every $k = 2, \dots, K$.*

We now turn to the study of the SDE (2.4). In the sequel we will call Feller diffusion with drift $\gamma \in \mathbb{R}$ the unique solution to the SDE

$$Z(t) = Z(0) + \gamma \int_0^t Z(u)\mathrm{d}u + \int_0^t (2Z(u))^{1/2}\mathrm{d}B(u). \tag{3.2}$$

It is well-known that there is a unique strong solution to (3.2). If Z is this solution, then it does not explode in finite time, $\mathbb{P}(T_0(Z) < +\infty) > 0$ and $\mathbb{P}(T_0(Z) < +\infty) = 1$ if $\gamma \leq 0$. Moreover, since A_1 satisfies

$$A_1(t) = A_1(0) + \int_0^t b_1(\mathbf{A}(u))\mathrm{d}u + \int_0^t (2A_1(u))^{1/2}\mathrm{d}B(u)$$

with $b_1(a) = (\gamma_1 - a_0)a_1 \leq \gamma_1 a_1$, Theorem V.43.1 in Rogers and Williams [20] implies that $A_1(t) \leq Z(t)$ for all $t \geq 0$ almost surely, where Z is Feller diffusion with drift γ_1 started at $Z(0) \geq A_1(0)$. This comparison argument will be used several times. The proof of the following result uses standard arguments, and we only sketch the proof.

Lemma 3.3. *Uniqueness in law holds for the SDE* (2.4).

Proof. The problem to invoke classical results is that some of the coefficients b_k grow quadratically. However, standard localization and change of drift arguments (to go back to the case of Feller diffusion, since $\sigma(a) = \mathrm{diag}(0, (2a_1)^{1/2}, 0, \dots, 0)$) show that for each $N \geq 1$, the law of a solution $\mathbf{A}$ to (2.4) stopped at $\inf\{t \geq 0 : \|\mathbf{A}(t)\| \geq N\}$ is uniquely determined (say, with $\|a\| = a_0 + \cdots + a_{K+1}$). By successively patching up these solutions, we obtain uniqueness to (2.4) until the time of explosion, and so it only remains to show that solutions to (2.4) do not explode. But A_1 cannot explode since it is dominated

by Feller diffusion, and since $\bar{\pi} \circ \mathbf{A}$ satisfies (3.1), no other coordinate can explode by Lemma 3.1. □

Lemma 3.4. *If* $\mathbf{A}$ *satisfies* (2.4)*, then* $T_0(A_1)$ *is almost surely finite.*

Before proving this result, we first explain briefly how it yields Proposition 2.2 in combination with Lemma 3.2.

Proof of Proposition 2.2. The fact that $T_0(A_1)$ is almost surely finite is precisely the content of Lemma 3.4. Then by shifting $\mathbf{A}$ at time $T_0(A_1)$, we see that $\theta_{T_0(A_1)} \circ \mathbf{A}$ satisfies the ODE (3.1) with $y = 0$, and so the results of Lemma 3.2 are precisely those that we need to prove for Proposition 2.2. □

Proof of Lemma 3.4. To prove $\mathbb{P}(T_0(A_1) < +\infty) = 1$, it is enough to prove that

$$\mathbb{P}\left(T_0(A_1) < +\infty, T^{\uparrow}_{\gamma_1}(A_0) < +\infty\right) = \mathbb{P}\left(T^{\uparrow}_{\gamma_1}(A_0) < +\infty\right) \tag{3.3}$$

and

$$\mathbb{P}\left(T_0(A_1) = +\infty, T^{\uparrow}_{\gamma_1}(A_0) = +\infty\right) = 0 \tag{3.4}$$

which we do now.

In the event $\{T^{\uparrow}_{\gamma_1}(A_0) < +\infty\}$, we have $\mathrm{d}\mathbf{A}^{\uparrow}(t) = b(\mathbf{A}^{\uparrow}(t))\mathrm{d}t + \sigma(\mathbf{A}^{\uparrow}(t))\mathrm{d}B^{\uparrow}(t)$, where $\mathbf{A}^{\uparrow}$ and $B^{\uparrow}$ are the processes $\mathbf{A}$ and B shifted at time $T^{\uparrow}_{\gamma_1}(A_0)$, so that, since $T^{\uparrow}_{\gamma_1}(A_0)$ is a stopping time, $B^{\uparrow}$ is a Brownian motion according to the strong Markov property. Moreover, we have by definition $b_1(\mathbf{A}^{\uparrow}(t)) = (\gamma_1 - A^{\uparrow}_0(t))A^{\uparrow}_1(t) \le 0$ and so $A^{\uparrow}_1$ is dominated by Feller diffusion without drift (see, e.g., Theorem V.43.1 in Rogers and Williams [20]), which almost surely hits 0 in finite time. This proves (3.3).

We now prove (3.4). Assume that $T^{\uparrow}_{\gamma_1}(A_0) = +\infty$: then A_0 is bounded (by γ_1, and in particular $\gamma_1 > 0$) and according to (i)+(ii) of Lemma 3.2, we obtain that $\int_0^{\infty} A_1$ is finite. In particular $L = 0$, where we define $L = \liminf_{t\to\infty} A_1(t)$, and so

$$\mathbb{P}\left(T_0(A_1) = +\infty, T^{\uparrow}_{\gamma_1}(A_0) = +\infty\right) \le \mathbb{P}\left(T_0(A_1) = +\infty, L = 0\right).$$

Thus to prove (3.4), we only have to show that $\mathbb{P}(T_0(A_1) = +\infty, L = 0) = 0$. Let ϑ_i be defined recursively by $\vartheta_0 = 0$ and $\vartheta_{i+1} = \inf\{t \ge 1 + \vartheta_i : A_1(t) \le 1\}$. Then in $\{L = 0\}$, ϑ_i is almost surely finite for every $i \ge 0$ and so we can define Z_i as the solution to the SDE $\mathrm{d}Z_i = \gamma_1 Z_i \mathrm{d}t + (2Z_i)^{1/2}\mathrm{d}B_i$ with initial condition $Z_i(0) = 1$, where B_i is the process B shifted at time ϑ_i. Note that conditionally on $\{\vartheta_i < +\infty\}$, Z_i is a Feller diffusion with drift $\gamma_1 > 0$ started at 1, and that the strong Markov property and the comparison theorem V.43.1 in Rogers and Williams [20] show that $A_1(t + \vartheta_i) \le Z_i(t)$ for $t \ge 0$ and $i \ge 1$. In particular,

$$\mathbb{P}\left(T_0(A_1) = +\infty, L = 0\right) \le \mathbb{P}\left(\forall i \ge 1 : \vartheta_i < +\infty \text{ and } T_0(Z_i) = +\infty\right)$$

and so we only have to show that this last probability is equal to 0. For $I \ge 1$ and in the event $\{\vartheta_I < +\infty\}$, the strong Markov property at time ϑ_I gives

$$\begin{aligned}
&\mathbb{P}\big(\vartheta_I < +\infty \text{ and } T_0(Z_i) = +\infty \text{ for } i = 1, \dots, I\big) \\
&\qquad = \mathbb{P}\big(T_0(Z_1) = +\infty\big)\mathbb{P}\big(\vartheta_I < +\infty \text{ and } T_0(Z_i) = +\infty \text{ for } i = 1, \dots, I-1\big) \\
&\qquad \le \mathbb{P}\big(T_0(Z_1) = +\infty\big)\mathbb{P}\big(\vartheta_{I-1} < +\infty \text{ and } T_0(Z_i) = +\infty \text{ for } i = 1, \dots, I-1\big)
\end{aligned}$$

and so we obtain by induction

$$\mathbb{P}\big(\vartheta_I < +\infty \text{ and } T_0(Z_i) = +\infty \text{ for } i = 1,\dots,I\big) \le \big[\mathbb{P}\big(T_0(Z_1) = +\infty\big)\big]^I.$$

Since $\mathbb{P}\big(T_0(Z_1) = +\infty\big) < 1$, letting $I \to +\infty$ achieves the proof of (3.4). □

4. Scaling limits

We now prove the convergence results of Theorem 2.1 and Propositions 2.4 and 2.5. The proofs of these three results can be cast into the same framework by defining $\mathbf{A}_n$ as in (2.7) with:

Intermediate initial condition (Theorem 2.1): $\alpha_{n,1} = n^{1/(K+2)}$, $\tau_n = \alpha_{n,1}$, $\alpha_{n,0} = \alpha_{n,K+1}$ and $\alpha_{n,k} = \alpha_{n,1}^k$ for $k = 2,\dots,K+1$;

Large initial condition (Proposition 2.4): $n^{1/(K+2)} \ll \alpha_{n,1} \ll n$ and τ_n and $\alpha_{n,k}$ for $k = 2,\dots,K+1$ are as in (2.5);

Small initial condition (Proposition 2.5): $1 \ll \alpha_{n,1} \ll n^{1/(K+2)}$, $\tau_n = \alpha_{n,1}$, $\alpha_{n,0} = \alpha_{n,K+1}$ and $\alpha_{n,k} = \alpha_{n,1}^k$ for $k = 2,\dots,K+1$;

and by assuming, with this notation, that $\tau_n(\varepsilon_{n,k} - \delta_{n,k}) \to \gamma_k$, which is consistent with the assumptions (2.2), (2.6) and (2.9).

In order to prove Theorem 2.1 and Propositions 2.4 and 2.5 with this notation, we have to prove that the sequence $(\mathbf{A}_n, n \ge 1)$ converges weakly toward: the solution of (2.4) in the intermediate regime; the solution of (2.8) in the large regime; and the solution of (2.11) in the small regime. In the sequel we will use the fact that in all three regimes, the algebraic relations $\alpha_{n,0} = \alpha_{n,K+1}$ and $\alpha_{n,k} = \tau_n \alpha_{n,k-1}$ for $k = 2,\dots,K+1$ hold.

The proof relies on the standard machinery: we first prove tightness and then identify accumulation points. Both steps rely on semimartingale arguments based on the explicit form of the generator of $\mathbf{A}_n$. Indeed, $\mathbf{A}_n$ is by definition a Markov process with generator Ω_n given, for any function $f : \mathbb{R}^{K+2} \to \mathbb{R}$ and any $a \in \mathbb{R}^{K+2}$, by

$$\begin{aligned}\Omega_n(f)(a) = \tau_n \sum_{k=1}^{K} \left[f\left(a - \frac{e_k}{\alpha_{n,k}} + \frac{e_{k+1}}{\alpha_{n,k+1}}\right) - f(a)\right](1+\delta_{n,k})\alpha_{n,k}a_k \\ + \tau_n\big(1 - \alpha_{n,0}a_0/n\big)\sum_{k=1}^{K}\left[f\left(a + \frac{e_0}{\alpha_{n,0}} + \frac{e_k}{\alpha_{n,k}}\right) - f(a)\right](1+\varepsilon_{n,k})\alpha_{n,k}a_k\end{aligned} \tag{4.1}$$

(note that $a_{n,0}/n = 1 - \alpha_{n,0}A_{n,0}/n$, which gives the factor $1 - \alpha_{n,0}a_0/n$ in the above expression). Since $\mathbf{A}_n$ lives on a finite state space, for any function f the process

$$M_n^f(t) = f(\mathbf{A}_n(t)) - f(\mathbf{A}_n(0)) - V_n^f(t) \text{ with } V_n^f(t) = \int_0^t \Omega_n(f)(\mathbf{A}_n(u))\mathrm{d}u$$

is a martingale whose quadratic variation process $\langle M_n^f\rangle$ is equal to

$$\langle M_n^f\rangle(t) = \int_0^t \Theta_n(f)(\mathbf{A}_n(u))\mathrm{d}u \text{ with } \Theta_n(f) = \Omega_n(f^2) - 2f\Omega_n(f),$$

see for instance Lemma VII.3.68 in Jacod and Shiryaev [11]. Recall that $\delta_{n,0} = \delta_{n,K+1} = \varepsilon_{n,K+1} = -1$, and let in the rest of this section π_k for $k = 0,\dots,K+1$ be the projection on the kth coordinate, i.e., $\pi_k(a) = a_k$ for any $a = (a_0,\dots,a_{K+1}) \in \mathbb{R}^{K+2}$.

Lemma 4.1. *For each $n \ge 1$ and $t \ge 0$,*

$$V_n^{\pi_0}(t) = \frac{\tau_n}{\alpha_{n,0}} \sum_{k=1}^{K} (1+\varepsilon_{n,k})\alpha_{n,k}\int_0^t (1 - \alpha_{n,0}A_{n,0}(u)/n)A_{n,k}(u)\mathrm{d}u, \tag{4.2}$$

$$V_n^{\pi_k}(t) = \int_0^t \Big[(1+\delta_{n,k-1}) A_{n,k-1}(u) + \tau_n(\varepsilon_{n,k}-\delta_{n,k}) A_{n,k}(u) - \frac{\tau_n \alpha_{n,0}}{n}(1+\varepsilon_{n,k}) A_{n,0}(u) A_{n,k}(u) \Big] \mathrm{d}u \tag{4.3}$$

for $k = 1, \dots, K+1$ *and*

$$\langle M_n^{\pi_k} \rangle(t) = \frac{1}{\alpha_{n,k}} V_n^{\pi_k}(t) + 2\frac{\tau_n}{\alpha_{n,k}}(1+\delta_{n,k}) \int_0^t A_{n,k}(u) \mathrm{d}u \tag{4.4}$$

for $k = 0, \dots, K+1$.

Proof. Recall that $V_n^{\pi_k}(t) = \int_0^t \Omega_n(\pi_k)(\mathbf{A}_n(u)) \mathrm{d}u$ and that $\langle M_n^{\pi_k} \rangle(t) = \int_0^t \Theta_n(\pi_k)(\mathbf{A}_n(u)) \mathrm{d}u$, so that we only have to compute $\Omega_n(\pi_k)$ and $\Theta_n(\pi_k)$ for $k = 0, \dots, K+1$. By writing $\Omega_n(f^2)(a) - 2f(a)\Omega_n(f)(a)$ in the form

$$\sum_j (x_j^2 - y^2)\beta_j - 2y \sum_j (x_j - y)\beta_j = \sum_j \Big((x_j^2 - y^2) - 2y(x_j - y) \Big) \beta_j = \sum_j (x_j - y)^2 \beta_j$$

for some x_j, β_j and y, we see that $\Theta_n(f)$ can alternatively be written as follows:

$$\begin{aligned} \Theta_n(f)(a) = \tau_n \sum_{k=1}^K \left[f\left(a - \frac{e_k}{\alpha_{n,k}} + \frac{e_{k+1}}{\alpha_{n,k+1}} \right) - f(a) \right]^2 (1+\delta_{n,k}) \alpha_{n,k} a_k \\ + \tau_n \big(1 - \alpha_{n,0} a_0 / n\big) \sum_{k=1}^K \left[f\left(a + \frac{e_0}{\alpha_{n,0}} + \frac{e_k}{\alpha_{n,k}} \right) - f(a) \right]^2 (1+\varepsilon_{n,k}) \alpha_{n,k} a_k. \end{aligned} \tag{4.5}$$

For $k = 0$,

$$\Omega_n(\pi_0)(a) = \frac{\tau_n}{\alpha_{n,0}}(1 - \alpha_{n,0} a_0 / n) \sum_{k=1}^K (1+\varepsilon_{n,k}) \alpha_{n,k} a_k \text{ and } \Theta_n(\pi_0)(a) = \frac{1}{\alpha_{n,0}} \Omega_n(\pi_0)(a)$$

which proves the result for $k = 0$. For $k = K+1$,

$$\Omega_n(\pi_{K+1})(a) = \frac{\tau_n}{\alpha_{n,K+1}}(1+\delta_{n,K}) \alpha_{n,K} a_K = (1+\delta_{n,K}) a_K$$

using $\alpha_{n,K+1} = \tau_n \alpha_{n,K}$ to get the last equality, and $\Theta_n(\pi_{K+1}) = \Omega_n(\pi_{K+1}) / \alpha_{n,K+1}$, which proves the result for $k = K+1$. Consider now $k = 1, \dots, K$:

$$\begin{aligned} \Omega_n(\pi_k)(a) &= \tau_n \left(-\frac{1}{\alpha_{n,k}}(1+\delta_{n,k}) \alpha_{n,k} a_k + \frac{1}{\alpha_{n,k}}(1+\delta_{n,k-1}) \alpha_{n,k-1} a_{k-1} \right) \\ &\qquad + \frac{\tau_n}{\alpha_{n,k}}(1 - \alpha_{n,0} a_0 / n)(1+\varepsilon_{n,k}) \alpha_{n,k} a_k \\ &= (1+\delta_{n,k-1}) a_{k-1} + \tau_n(\varepsilon_{n,k} - \delta_{n,k}) a_k - \frac{\tau_n \alpha_{n,0}}{n}(1+\varepsilon_{n,k}) a_0 a_k, \end{aligned}$$

using $\delta_{n,0} = -1$ for $k = 1$, and $\alpha_{n,k-1}\tau_n = \alpha_{n,k}$ for $k \geq 2$. This proves the result for $V_n^{\pi_k}$, while writing

$$\begin{aligned} \Theta_n(\pi_k)(a) &= \tau_n \left(\frac{1}{\alpha_{n,k}^2}(1+\delta_{n,k}) \alpha_{n,k} a_k + \frac{1}{\alpha_{n,k}^2}(1+\delta_{n,k-1}) \alpha_{n,k-1} a_{k-1} \right) \\ &\qquad + \frac{\tau_n}{\alpha_{n,k}^2}(1 - \alpha_{n,0} a_0 / n)(1+\varepsilon_{n,k}) \alpha_{n,k} a_k \\ &= \frac{1}{\alpha_{n,k}} \Omega_n(\pi_k)(a) + 2\frac{\tau_n}{\alpha_{n,k}}(1+\delta_{n,k}) a_k \end{aligned}$$

proves the result for $M_n^{\pi_k}$, which concludes the proof. □

4.1. **Tightness.** We now prove that the sequence $(\mathbf{A}_n, n \geq 1)$ is tight. Since $\mathbf{A}_n$ makes jumps of vanishing size (at most $1/\alpha_{n,1}$), in order to show that $(\mathbf{A}_n, n \geq 1)$ is tight it is sufficient to show that for every $T \geq 0$,

$$\lim_{\eta \to 0} \limsup_{n \to +\infty} \Delta_n(\eta) = 0 \text{ where } \Delta_n(\eta) = \sup_{\Upsilon} \sup_{0 \leq t \leq \eta} \mathbb{E}\left(\sum_{k=0}^{K+1} \left| A_{n,k}(\Upsilon + t) - A_{n,k}(\Upsilon) \right| \right) \tag{4.6}$$

and where the first supremum in the definition of $\Delta_n(\eta)$ is taken over all the random variables $\Upsilon \leq T$ that are stopping times relatively to the filtration generated by $\mathbf{A}_n$ (see Corollary on page 179 in Billingsley [4]). Fix in the rest of the proof some $T \geq 0$. Because of the strong Markov property and the fact that we consider stopping times $\Upsilon \leq T$, we have

$$\Delta_n(\eta) \leq \sup_{0 \leq t \leq \eta} \mathbb{E}\left[\sup_{0 \leq y \leq T} \Phi_n^t(\mathbf{A}_n(y)) \right] \tag{4.7}$$

where

$$\Phi_n^t(a) = \sum_{k=0}^{K+1} \mathbb{E}_a\left(\left| A_{n,k}(t) - A_{n,k}(0) \right| \right) \tag{4.8}$$

and the subscript refers to the initial state of the process $\mathbf{A}_n$ (when there is no subscript, this refers to an initial condition as in the statement of the theorem or the propositions). By definition we have $\mathbb{E}_a(|A_{n,k}(t) - A_{n,k}(0)|) = \mathbb{E}_a(|M_n^{\pi_k}(t) + V_n^{\pi_k}(t)|)$, and so combining the triangular inequality with Cauchy-Schwarz inequality and summing over $k = 0, \ldots, K+1$ leads to

$$\Phi_n^t(a) \leq \sum_{k=0}^{K+1} \sqrt{\mathbb{E}_a\left(\langle M_n^{\pi_k} \rangle(t) \right)} + \sum_{k=0}^{K+1} \mathbb{E}_a\left(\left| V_n^{\pi_k}(t) \right| \right). \tag{4.9}$$

In order to do control the first moment of $\langle M_n^{\pi_k} \rangle(t)$ and of $V_n^{\pi_k}(t)$, we introduce the functions $\psi = \pi_0 + \cdots + \pi_K$, i.e., $\psi(a) = a_0 + \cdots + a_K$ for $a = (a_0, \ldots, a_{K+1}) \in \mathbb{R}^{K+2}$, and $\Psi = \psi + \psi^2$. Since $\alpha_{n,1} \to +\infty$ as $n \to +\infty$, $\alpha_{n,k+1} = \tau_n \alpha_{n,k}$ for $k = 1, \ldots, K$ and $\tau_n \to +\infty$, we will assume for convenience that $1 \leq \alpha_{n,1} \leq \cdots \leq \alpha_{n,K+1}$. Moreover, it can be checked that the constant

$$C_0 = \sup_{n \geq 1, 1 \leq k \leq K} \left(\tau_n |\varepsilon_{n,k} - \delta_{n,k}|, 1 + |\delta_{n,k}|, 1 + |\varepsilon_{n,k}|, \frac{\tau_n \alpha_{n,0}}{n}, \frac{2\tau_n}{\alpha_{n,k}}, \frac{\tau_n \alpha_{n,k}}{\alpha_{n,0}} \right)$$

is finite. From these various definitions and (4.2)–(4.4), and also recalling that $\delta_{n,0} = \delta_{n,K+1} = \varepsilon_{n,K+1} = -1$, it follows that

$$\max\left(\left| V_n^{\pi_k}(t) \right|, \langle M_n^{\pi_k} \rangle(t) \right) \leq 2C_0^2 \int_0^t \Psi(\mathbf{A}_n(u)) \mathrm{d}u, \ k = 0, \ldots, K+1, t \geq 0. \tag{4.10}$$

Combined with (4.9) this leads to

$$\Phi_n^t(a) \leq C_1 \sqrt{\int_0^t \mathbb{E}_a\left[\Psi(\mathbf{A}_n(u)) \right] \mathrm{d}u} + C_1 \int_0^t \mathbb{E}_a\left[\Psi(\mathbf{A}_n(u)) \right] \mathrm{d}u$$

with $C_1 = 2C_0^2(K+2)$. For $t \leq \eta \leq 1$ we obtain

$$\Phi_n^t(a) \leq C_1 \eta^{1/2} \left(\sup_{0 \leq u \leq 1} \sqrt{\mathbb{E}_a\left[\Psi(\mathbf{A}_n(u)) \right]} + \sup_{0 \leq u \leq 1} \mathbb{E}_a\left[\Psi(\mathbf{A}_n(u)) \right] \right). \tag{4.11}$$

Lemma 4.2. *There exists a finite constant C_2 such that for every initial state a,*

$$\sup_{0\le u\le 1} \mathbb{E}_a\left[\psi^i(\mathbf{A}_n(u))\right] \le C_2 \sum_{j=0}^{i} \psi^j(a),\ i=1,2. \tag{4.12}$$

In particular,

$$\sup_{n\ge 1} \mathbb{E}\left[\sup_{0\le y\le T} \Psi(\mathbf{A}_n(y))\right] < +\infty. \tag{4.13}$$

Let us quickly finish the proof of the tightness of $(\mathbf{A}_n)$ based on this lemma: first, plugging in (4.12) into (4.11), we obtain the existence of a finite constant C_3 such that

$$\Phi_n^t(a) \le C_3\eta^{1/2}\left(\sqrt{\Psi(a)}+\Psi(a)\right),\ 0\le t\le\eta\le 1,$$

and so in view of (4.7), we see that for every $\eta\le 1$,

$$\Delta_n(\eta) \le C_3\eta^{1/2}\mathbb{E}\left[\sup_{0\le y\le T}\left(\sqrt{\Psi(\mathbf{A}_n(y))}+\Psi(\mathbf{A}_n(y))\right)\right].$$

Thus, (4.13) implies the existence of a finite constant C_4 such that $\Delta_n(\eta)\le C_4\eta^{1/2}$ which achieves to prove that $(\mathbf{A}_n)$ is tight. It remains to prove Lemma 4.2.

Proof of Lemma 4.2. Let us first prove (4.12). Defining $\eta_{n,k} = 1/\alpha_{n,k} - \mathbb{1}_{\{k\ne K\}}/\alpha_{n,k+1} \ge 0$ and writing

$$\psi^i(a) - \psi^i\left(a - \frac{e_k}{\alpha_{n,k}} + \frac{e_{k+1}}{\alpha_{n,k+1}}\right) = \big(\psi(a)\big)^i - \big(\psi(a)-\eta_{n,k}\big)^i = i\int_{\psi(a)-\eta_{n,k}}^{\psi(a)} x^{i-1}\mathrm{d}x,$$

we get

$$\psi^i(a) - \psi^i\left(a - \frac{e_k}{\alpha_{n,k}} + \frac{e_{k+1}}{\alpha_{n,k+1}}\right) \ge i\eta_{n,k}\big(\psi(a)-\eta_{n,k}\big)^{i-1}.$$

Defining $\mu_{n,k} = 1/\alpha_{n,0} + 1/\alpha_{n,k}$ similarly leads to

$$\psi^i\left(a + \frac{e_0}{\alpha_{n,0}} + \frac{e_k}{\alpha_{n,k}}\right) - \psi^i(a) \le i\mu_{n,k}\big(\psi(a)+\mu_{n,k}\big)^{i-1}.$$

Plugging these inequalities in the definition (4.1) of Ω_n, we obtain

$$\begin{aligned}\Omega_n(\psi^i)(a) \le i\tau_n(1-\alpha_{n,0}a_0/n)\sum_{k=1}^{K}\mu_{n,k}\big(\psi(a)+\mu_{n,k}\big)^{i-1}(1+\varepsilon_{n,k})\alpha_{n,k}a_k \\ - i\tau_n\sum_{k=1}^{K}\eta_{n,k}\big(\psi(a)-\eta_{n,k}\big)^{i-1}(1+\delta_{n,k})\alpha_{n,k}a_k.\end{aligned}$$

Using $1-\alpha_{n,0}a_0/n\le 1$, expanding the terms raised to the power $i-1$ and changing the order of summation, we end up with

$$\begin{aligned}\Omega_n(\psi^i)(a) \le i\sum_{j=0}^{i-1}\binom{i-1}{j}\big(\psi(a)\big)^{i-1-j} \\ \times\left\{\tau_n\sum_{k=1}^{K}\mu_{n,k}^{j+1}(1+\varepsilon_{n,k})\alpha_{n,k}a_k - (-1)^j\tau_n\sum_{k=1}^{K}\eta_{n,k}^{j+1}(1+\delta_{n,k})\alpha_{n,k}a_k\right\}.\end{aligned}$$

Plugging in the definitions of $\eta_{n,k}$ and $\mu_{n,k}$, we see that for $j=0$ the term between brackets in the previous display is given by

$$\tau_n \sum_{k=1}^{K} (1+\varepsilon_{n,k})(1+\alpha_{n,k}/\alpha_{n,0})a_k - \tau_n \sum_{k=1}^{K} (1+\delta_{n,k})(1-\mathbb{1}_{\{k\neq K\}}\alpha_{n,k}/\alpha_{n,k+1})\alpha_{n,k}a_k$$
$$= \tau_n \sum_{k=1}^{K} (\varepsilon_{n,k}-\delta_{n,k})a_k + \sum_{k=1}^{K} \frac{\tau_n \alpha_{n,k}}{\alpha_{n,0}}(1+\varepsilon_{n,k})a_k + \sum_{k=1}^{K-1}(1+\delta_{n,k})a_k \le 3C_0^2\psi(a).$$

For $j\ge 1$, we obtain similarly, using also $\mu_{n,k}\le 2/\alpha_{n,k}\le 2$ and $\eta_{n,k}\le 1/\alpha_{n,k}\le 1$, that the term between brackets is upper bounded by

$$\tau_n \sum_{k=1}^{K} \frac{4}{\alpha_{n,k}^2}(1+\varepsilon_{n,k})\alpha_{n,k}a_k + \tau_n \sum_{k=1}^{K} \frac{1}{\alpha_{n,k}^2}(1+\delta_{n,k})\alpha_{n,k}a_k \le 3C_0^2\psi(a).$$

We thus obtain, for some finite constant $C(i)$ and every a,

$$\Omega_n(\psi^i)(a) \le C(i)\sum_{j=1}^{i}\big(\psi(a)\big)^j. \tag{4.14}$$

In particular,

$$\mathbb{E}_a\Big(\psi^i(\mathbf{A}_n(t))\Big) = \psi^i(\mathbf{A}_n(0)) + \int_0^t \mathbb{E}_a\Big(\Omega_n(\psi^i)(\mathbf{A}_n(u))\Big)\,\mathrm{d}u$$
$$\le C(i)\int_0^t \mathbb{E}_a\left(\sum_{j=0}^{i-1}\psi^j(\mathbf{A}_n(u))\right)\mathrm{d}u + C(i)\int_0^t \mathbb{E}_a\Big(\psi^i(\mathbf{A}_n(u))\Big)\,\mathrm{d}u$$

assuming without loss of generality that $C(i)\ge 1$ for the last inequality. Thus Gronwall's lemma implies

$$\mathbb{E}_a\Big(\psi^i(\mathbf{A}_n(t))\Big) \le C(i)\int_0^t \sum_{j=0}^{i-1}\mathbb{E}_a\Big(\psi^j(\mathbf{A}_n(u))\Big)\,\mathrm{d}u \times e^{C(i)t}.$$

Then (4.12) follows from this inequality by induction on i. We now derive (4.13): since $\Psi=\psi+\psi^2$ it is enough to prove (4.13) with ψ^2 in place of Ψ. First of all, note that the previous reasoning shows the existence of a finite constant C_2' such that

$$\sup_{0\le u\le T}\mathbb{E}\Big[\psi^i(\mathbf{A}_n(u))\Big] \le C_2'\sum_{j=0}^{i}\psi^j(\mathbf{A}_n(0)),\ i=1,2,3,4. \tag{4.15}$$

By definition, $\psi^2\circ\mathbf{A}_n = \psi^2(\mathbf{A}_n(0)) + V_n + M_n$, defining $V_n = V_n^{\psi^2}$ and $M_n = M_n^{\psi^2}$, and so

$$\mathbb{E}\left[\sup_{0\le y\le T}\psi^2(\mathbf{A}_n(y))\right] \le \psi^2(\mathbf{A}_n(0)) + \mathbb{E}\left[\sup_{0\le y\le T}V_n(y)\right] + \mathbb{E}\left[\sup_{0\le y\le T}M_n(y)\right].$$

The first term of the above upper bound is bounded (in n) since the sequence $\mathbf{A}_n(0)$ converges by assumption. For the second term, we write

$$V_n(y) = \int_0^y \Omega_n(\psi^2)(\mathbf{A}_n(u))\mathrm{d}u \le C(2)\int_0^y \Psi(\mathbf{A}_n(u))\mathrm{d}u$$

where the inequality comes from (4.14). Thus

$$\mathbb{E}\left[\sup_{0\le y\le T}V_n(y)\right] \le C(2)\int_0^T \mathbb{E}\big[\Psi(\mathbf{A}_n(u))\big]\mathrm{d}u \le C(2)\,T\sup_{0\le u\le T}\mathbb{E}\big[\Psi(\mathbf{A}_n(u))\big]$$

which is finite by (4.15). We now control the last martingale term. For any real-valued random variable X we have $\mathbb{E}(X) \le 1 + \mathbb{E}(X^2)$ and so Doob's inequality gives

$$\mathbb{E}\left(\sup_{0\le t\le T} M_n(t)\right) \le 1 + 4\mathbb{E}\left[\langle M_n(T)\rangle\right] = 1 + 4\int_0^T \mathbb{E}\left[\Theta_n(\psi^2)(\mathbf{A}_n(u))\right] \mathrm{d}u.$$

Defining Θ_n' as in (4.5) but with the term $-\alpha_{n,0}a_0/n$ taken equal to 0, we have

$$\Theta_n(\psi^2)(a) \le \Theta_n'(\psi^2)(a) = \Omega_n'(\psi^4)(a) - 2\psi^2(a)\Omega_n'(\psi^2)(a)$$

defining Ω_n' similarly as Ω_n in (4.1) but with the term $-\alpha_{n,0}a_0/n$ taken equal to 0. Since the first step in the derivation of (4.14) was to use $1-\alpha_{n,0}a_0/n \le 1$, the reasoning leading to (4.14) also leads to an upper bound on $|\Omega_n'(\psi^2)|$ of the same kind, i.e., it leads to the existence of a finite constant $C'(i)$ such that $|\Omega_n'(\psi^i)(a)| \le C'(i)\sum_{j=1}^i \psi^j(a)$. This finally proves that $\Theta_n(\psi^2)(a) \le C'\sum_{j=1}^4 \psi^j(a)$ for some finite constant C', and in particular

$$\mathbb{E}\left(\sup_{0\le t\le T} M_n(t)\right) \le 1 + 4C'T\sum_{j=1}^4 \sup_{0\le u\le T} \mathbb{E}\left[\psi^j(\mathbf{A}_n(u))\right].$$

Since the supremum over $n \ge 1$ of the right-hand side is finite by (4.15), the proof is complete. □

4.2. **Characterization of accumulation points.** Let $\mathbf{A} = (A_k, k = 0,\dots,K+1)$ be any accumulation point of $(\mathbf{A}_n)$ and assume without loss of generality that $\mathbf{A}_n \Rightarrow \mathbf{A}$, i.e., $\mathbf{A}_n$ converges weakly to $\mathbf{A}$. Remember that we have to prove that $\mathbf{A}$ satisfies (2.4) in the intermediate regime, (2.8) in the large regime and (2.11) in the small regime. We treat the three regimes separately.

In each regime, if $f \in \mathscr{D}$ we denote by $\int f$ the function $(\int_0^t f(u)\mathrm{d}u, t \ge 0)$, and we will use the following result in conjunction with the continuous mapping theorem. If (f_n), (g_n) are two sequences of functions with $f_n \to f$ and $g_n \to g$ for some continuous functions f and g (where $\to$ denotes uniform convergence on compact sets), then $f_n g_n \to fg$ and $\int f_n \to \int f$.

In particular, since in all three regimes it holds that $\tau_n\alpha_{n,k}/\alpha_{n,0} = \alpha_{n,k+1}/\alpha_{n,K+1} \to \mathbb{1}_{\{k=K\}}$ and $\alpha_{n,0}/n \to 0$, we have from (4.2) that $V_n^{\pi_0} \Rightarrow \int A_K$. Since $\alpha_{n,0} \to +\infty$ and $\delta_{n,0} = -1$ we have $\langle M_n^{\pi_0}\rangle \Rightarrow 0$ by (4.4), which implies by Doob's inequality that $M_n^{\pi_0} \Rightarrow 0$. Similarly, recalling that $\delta_{n,K+1} = \varepsilon_{n,K+1} = -1$ we have $V_n^{\pi_{K+1}} \Rightarrow \int A_K$ and $M_n^{\pi_{K+1}} \Rightarrow 0$.

4.2.1. *Large initial condition.* In this regime, we have $\tau_n\alpha_{n,0} = n$. Since moreover $\delta_{n,0} = \delta_{n,K+1} = \varepsilon_{n,K+1} = -1$, we obtain from (4.3) that $V_n^{\pi_1} \Rightarrow \int(\gamma_1 - A_0)A_1$ and $V_n^{\pi_k} \Rightarrow \int(A_{k-1} + (\gamma_k - A_0)A_k)$ for $k = 2,\dots,K$. In other words, $V_n^{\pi_k} \Rightarrow \int b_k \circ \mathbf{A}$ for $k = 0,\dots,K+1$.

Moreover, since $\alpha_{n,k} \to 0$ and $\tau_n/\alpha_{n,k} \to 0$ for $k = 1,\dots,K$, we obtain from (4.4) that $\langle M_n^{\pi_k}\rangle \Rightarrow 0$ for any $k = 0,\dots,K+1$. By Doob's inequality, this implies $M_n^{\pi_k} \Rightarrow 0$ and since $A_{n,k} = A_{n,k}(0) + M_n^{\pi_k} + V_n^{\pi_k}$ we finally get $A_{n,k} \Rightarrow a_k + \int b_k \circ \mathbf{A}$. Since all the above convergences hold jointly as a consequence of the continuous mapping theorem, we get on the one hand that $\mathbf{A}_n \Rightarrow a + \int b \circ \mathbf{A}$, while on the other hand, since $\mathbf{A}_n \Rightarrow \mathbf{A}$ by assumption, we get $\mathbf{A} = a + \int b \circ \mathbf{A}$. Thus $\mathbf{A}$ solves (2.8) as desired, and since uniqueness of solutions to this ODE is guaranteed by Lemma 3.1, this uniquely characterizes $\mathbf{A}$.

4.2.2. *Intermediate initial condition.* We still have $\tau_n\alpha_{n,0} = n$, so that as in the large regime we have $V_n^{\pi_k} \Rightarrow \int b_k \circ \mathbf{A}$. Moreover, in this regime we have $\tau_n/\alpha_{n,k} \to 0$ for $k = 2,\dots,K$ which implies as in the previous regime $M_n^{\pi_k} \Rightarrow 0$ for $k = 0, 2, \dots, K+1$.

The difference with the large regime is that since $\tau_n = \alpha_{n,1}$, we have $\langle M_n^{\pi_1}\rangle \Rightarrow 2\int A_1$. Note that $2\int A_1 = \langle M\rangle$ where $M(t) = \int_0^t (2A_1(u))^{1/2}\mathrm{d}B(u)$. Moreover, since $M_n^{\pi_k} \Rightarrow 0$ for

$k \neq 1$ while $\langle M_n^{\pi_1} \rangle \Rightarrow \langle M \rangle$, we get that the quadratic co-variation processes $\langle M_n^{\pi_k}, M_n^{\pi_\ell} \rangle$ vanish for any $k \neq \ell$ by polarization. Since all these convergences hold jointly, Theorem IX.2.4 in Jacod and Shiryaev [11] shows that $\mathbf{A}$ is the semimartingale with characteristics $(b \circ \mathbf{A}, M)$ in the sense that $\mathbf{A} = a + \int b \circ \mathbf{A} + M$, i.e., $\mathbf{A}$ solves (2.4) (and thus is uniquely determined by Lemma 3.3).

4.2.3. *Small initial condition.* It follows similarly as in the two previous regimes, noting that in this regime we have $\tau_n \alpha_{n,0}/n \to 0$ (which leads to the drift term b^S instead of b) and $\tau_n/\alpha_{n,k} \to \mathbb{1}_{\{k=1\}}$ (which leads to the non-vanishing diffusion term as in the intermediate regime).

5. Asymptotic behavior of the outbreak size

This section is devoted to proving Theorem 2.3. For $f \in \mathscr{D}$ we define the operator $\bar{T}_0(f) = \sup\{t \geq 0 : f(t) > 0\}$; recall also the various operators defined in Section 2.1. Then we have $A_{n,0}(\infty) = A_{n,0}(\bar{T}_0(A_{n,K}))$ and, in order to compute $\bar{T}_0(A_{n,K})$, we will use the relation $\bar{T}_0(A_{n,2}) = T_0(A_{n,1}) + T_0 \circ \theta_{T_0(A_{n,1})}(A_{n,2})$ which, iterated, leads to

$$\bar{T}_0(A_{n,K}) = \sum_{k=1}^{K-1} T_0 \circ \theta_{T_0(A_{n,k-1})} \circ \cdots \circ \theta_{T_0(A_{n,1})}(A_{n,k}).$$

There are two difficulties to solve in order to prove Theorem 2.3: the first one is that hitting times are in general not continuous functional, i.e., we may have $f_n \to f$ but $T_0(f_n) \not\to T_0(f)$. The second difficulty is that, as Proposition 2.1 shows, $\bar{T}_0(A_K) = +\infty$ for $K \geq 2$ while the convergence $\mathbf{A}_n \Rightarrow \mathbf{A}$ holds uniformly on *compact* sets (see the discussion following Theorem 2.3). We address these two difficulties in two steps.

5.1. **First step.** The goal of this first step is to prove that $(\mathbf{A}_n, T_0(A_{n,1})) \Rightarrow (\mathbf{A}, T_0(A_1))$. Assume first that for every $\delta > 0$,

$$\limsup_{n \to +\infty} \mathbb{P}\left(T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) \underset{\varepsilon \to 0}{\longrightarrow} 0. \tag{5.1}$$

We now argue that this implies $(\mathbf{A}_n, T_0(A_{n,1})) \Rightarrow (\mathbf{A}, T_0(A_1))$. First of all, note that if $T_0(A_{n,1}) \Rightarrow T_0(A_1)$ then the joint convergence automatically holds, see for instance Corollary 2.2 in Lambert et al. [15].

To see that (5.1) implies $T_0(A_{n,1}) \Rightarrow T_0(A_1)$, let us say that A_1 *goes across* ε if for every $\eta > 0$ we have $\inf_{0 \leq t \leq \eta} A_1(T_\varepsilon^{\downarrow}(A_1) + t) < \varepsilon$, and define the random set $\mathscr{G}$ through $\mathscr{G} = \{\varepsilon > 0 : A_1 \text{ goes across } \varepsilon\}$. Then it is known (and actually easy to show) that if $T_\varepsilon^{\downarrow}(A_1)$ is almost surely finite and $\mathbb{P}(\varepsilon \in \mathscr{G}) = 1$, then $T_\varepsilon^{\downarrow}(A_{n,1}) \Rightarrow T_\varepsilon^{\downarrow}(A_1)$, see for instance Proposition VI.2.11 in Jacod and Shiryaev [11] or Lemma 3.1 in Lambert and Simatos [14]. Note that in our case, $T_\varepsilon^{\downarrow}(A_1)$ is finite by Proposition 2.2.

On the other hand, the complement $\mathscr{G}^c$ of $\mathscr{G}$ is precisely the set of discontinuities of the process $(T_\varepsilon^{\downarrow}(A_1), \varepsilon > 0)$. Since $(T_\varepsilon^{\downarrow}(A), \varepsilon > 0)$ is càglàd, as the left-continuous inverse of the process $(\inf_{[0,t]} A_1, t \geq 0)$, the set $\{\varepsilon > 0 : \mathbb{P}(\varepsilon \in \mathscr{G}^c) > 0\}$ (sometimes called set of fixed times of discontinuities) is at most countable, see for instance Billingsley [4, Section 13]. Gathering these two observations, we see that $T_\varepsilon^{\downarrow}(A_{n,1}) \Rightarrow T_\varepsilon^{\downarrow}(A_1)$ for all $\varepsilon > 0$ outside a countable set. Then, writing

$$\begin{aligned}\mathbb{P}\left(T_0(A_{n,1}) \geq x\right) = \mathbb{P}\left(T_0(A_{n,1}) \geq x, T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) \\ + \mathbb{P}\left(T_0(A_{n,1}) \geq x, T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) < \delta\right),\end{aligned}$$

using (5.1) and playing with quantifiers gives the convergence of $T_0(A_{n,1})$ toward $T_0(A_1)$. Indeed, we can for instance write

$$\mathbb{P}\left(T_0(A_{n,1}) \geq x\right) \leq \mathbb{P}\left(T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) + \mathbb{P}\left(T_\varepsilon^{\downarrow}(A_{n,1}) \geq x - \delta\right),$$

then choose ε such that $T_\varepsilon^{\downarrow}(A_{n,1}) \to T_\varepsilon^{\downarrow}(A_1)$ to get by the portmanteau theorem, for any $\delta > 0$,

$$\limsup_{n\to+\infty} \mathbb{P}\left(T_0(A_{n,1}) \geq x\right) \leq \limsup_{n\to+\infty} \mathbb{P}\left(T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) + \mathbb{P}\left(T_\varepsilon^{\downarrow}(A_1) \geq x - \delta\right).$$

Since $T_\varepsilon^{\downarrow}(A_1) \to T_0(A_1)$ as $\varepsilon \to 0$, we get by letting first $\varepsilon \to 0$ and then $\delta \to 0$, and using (5.1),

$$\limsup_{n\to+\infty} \mathbb{P}\left(T_0(A_{n,1}) \geq x\right) \leq \mathbb{P}\left(T_0(A_1) \geq x\right).$$

Since x was arbitrary, this shows that $T_0(A_{n,1}) \Rightarrow T_0(A_1)$ by the portmanteau theorem. In conclusion, (5.1) indeed implies $T_0(A_{n,1}) \Rightarrow T_0(A_1)$.

The proof of (5.1) relies on a simple coupling between $A_{n,1}$ and a continuous-time branching process (more precisely, a Bellman-Harris branching process). Looking at the transition rates of the process $a_{n,1}$ in (2.1), we see that $A_{n,1}$ decreases by $1/\alpha_{n,1}$ at rate $(1+\delta_{n,1})A_{n,1}\alpha_{n,1}^2$ and increases by $1/\alpha_{n,1}$ at rate

$$(1+\varepsilon_{n,1})A_{n,1}(1 - A_{n,0}/\alpha_{n,1})\alpha_{n,1}^2 \leq (1+\varepsilon_{n,1})A_{n,1}\alpha_{n,1}^2.$$

In particular, shifting the origin of time at $T_\varepsilon^{\downarrow}(A_{n,1})$ and using the strong Markov property, one sees that $A_{n,1}$ can be coupled with a Markov process $Z_{n,1}$ in such a way that $Z_{n,1}(0) = \alpha_{n,1}\lfloor \varepsilon/\alpha_{n,1} \rfloor$, $A_{n,1}(T_\varepsilon^{\downarrow}(A_1) + t) \leq Z_{n,1}(t)$ for $t \geq 0$, and $Z_{n,1}$ decreases by $1/\alpha_{n,1}$ at rate $(1+\delta_{n,1})Z_{n,1}\alpha_{n,1}^2$ and increases by $1/\alpha_{n,1}$ at rate $(1+\varepsilon_{n,1})Z_{n,1}\alpha_{n,1}^2$ (note that the law of $Z_{n,1}$ depends on ε, but we omit this dependency in order to ease the notation). More concretely, this coupling can for instance be realized by adding a "ghost" individual in the population each time an individual in stage one makes an unsuccessful infection attempt (alternatively, we could also invoke the comparison result of Rogers and Williams [20, Theorem V.43.1]). Using the strong Markov property at time $T_\varepsilon^{\downarrow}(A_{n,1})$ and this coupling, we get

$$\mathbb{P}\left(T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) \leq \mathbb{P}\left(T_0(Z_{n,1}) \geq \delta\right).$$

It is well-known that $Z_{n,1} \Rightarrow Z_1$, where Z_1 is Feller diffusion with drift γ_1 started at ε (see for instance Ethier and Kurtz [8, Chapter 9]). Moreover, standard arguments can be used to show that $T_0(Z_{n,1}) \Rightarrow T_0(Z_1)$, for instance by using the fact that $Z_{n,1}$ and Z_1 are time-change of Lévy processes killed at 0, say Y_n and Y, so that $T_0(Z_{n,1}) = \int_0^\infty Y_n \Rightarrow \int_0^\infty Y = T_0(Z_1)$ (this time-change transformation is usually called Lamperti transformation, see for instance Lamperti [16]). Thus we have (making clear the role of the initial condition)

$$\limsup_{n\to+\infty} \mathbb{P}\left(T_0(A_{n,1}) - T_\varepsilon^{\downarrow}(A_{n,1}) \geq \delta\right) \leq \mathbb{P}\left(T_0(Z_1) \geq \delta \mid Z_1(0) = \varepsilon\right).$$

Since $T_0(Z_1) \Rightarrow 0$ as $Z_1(0) \to 0$, we have finally proved (5.1), which concludes the first step.

5.2. **Second step.** The first step shows, by using the strong Markov property at time $T_0(A_{n,1})$, that we only need to prove Theorem 2.3 when $A_{n,1}(0) = 0$. In this case, Theorem 2.1 shows that $\mathbf{A}_n \Rightarrow \mathbf{A}$, where $\mathbf{A}$ is a solution to the ODE (3.1) with $y = 0$.

With this initial condition, we have $A_{n,1}(t) = 0$ for all $t \geq 0$, and $\bar{\pi} \circ \mathbf{A}_n$ is a Markov process: actually, it is a multistage epidemic process with $K-1$ stages. The problem to iterate the arguments of the first step is that Proposition 2.2 shows that, although $T_\varepsilon^\downarrow(A_2) < +\infty$ for every $\varepsilon > 0$, we have $T_0(A_2) = +\infty$. To get round this problem, we will use a time-change argument. Such an idea is classical in the SIR case $K = 1$, see for instance von Bahr and Martin-Löf [23].

Let us set up a similar coupling as in the first step. The process $A_{n,2}$ decreases by $1/\alpha_{n,1}^2$ at rate $(1+\delta_{n,2})\alpha_{n,1}^3 A_{n,2}$ and increases by $1/\alpha_{n,1}^2$ at rate

$$(1+\varepsilon_{n,2})\alpha_{n,1}^3 A_{n,2}(1 - A_{n,0}/\alpha_{n,1}) \leq (1+\varepsilon_{n,2})\alpha_{n,1}^3 A_{n,2}(1 - A_{n,0}(0)/\alpha_{n,1}),$$

where the inequality follows from the monotonicity of $A_{n,0}$. Thus similarly as in the first step, we can couple $A_{n,2}$ with a continuous-time Markovian branching process $Z_{n,2}$ such that $Z_{n,2}(0) = A_{n,2}(0)$, $A_{n,2}(t) \leq Z_{n,2}(t)$ for $t \geq 0$, and $Z_{n,2}$ decreases by $1/\alpha_{n,1}^2$ at rate $(1+\delta_{n,2})\alpha_{n,1}^3 Z_{n,2}$ and increases by $1/\alpha_{n,1}^2$ at rate $(1+\varepsilon_{n,2})\alpha_{n,1}^3 Z_{n,2}(1 - A_{n,0}(0)/\alpha_{n,1})$. In particular, $Z_{n,2} \Rightarrow z_2$ with $z_2(t) = z_2(0)\exp(-(A_0(0) - \gamma_2)t)$.

Since by Proposition 2.2, A_0 is strictly increasing with $A_0(\infty) > \gamma_2$, we can assume without loss of generality by shifting the processes at time $T_{\gamma_2}^\uparrow(A_{n,0}) + 1$ that $A_0(0) > \gamma_2$, so that each z_2 vanishes exponentially fast. The problem, as mentioned earlier, is that $T_0(z_2) = +\infty$: we now introduce the time-change argument.

Let $C_{n,2}$ be the right-continuous inverse of $t \mapsto \int_0^t Z_{n,2}$ and c_2 be the right-continuous inverse of $t \mapsto \int_0^t z_2$, in the sense that $\int_0^{C_{n,2}(t)} Z_{n,2} = t$ for $t < \int_0^\infty Z_{n,2}$ and $\int_0^{c_2(t)} z_2 = t$ for $t < \int_0^\infty z_2$. Since $\int_0^\infty z_2 < +\infty$, c_2 blows up at time $\int_0^\infty z_2$. Moreover, such random time-change transformations induce continuous mappings, so that $\mathbf{A}_n \circ C_{n,2} \Rightarrow \mathbf{A} \circ c_2$ and $Z_{n,2} \circ C_{n,2} \Rightarrow z_2 \circ c_2$, see for instance Helland [9].

Time-changing $Z_{n,2}$ with $C_{n,2}$ actually corresponds to the Lamperti transformation mentioned above: $Z_{n,2} \circ C_{n,2}$ is a continuous-time random walk (killed at 0), $z_2 \circ c_2$ starts at $A_2(0)$ and decays linearly at rate $A_0(0) - \gamma_2$, i.e., $z_2(c_2(t)) = A_2(0) - (A_0(0) - \gamma_2)t$ for $t \leq \int_0^\infty z_2$, and $T_0(Z_{n,2} \circ C_{n,2}) \Rightarrow T_0(z_2 \circ c_2)$. In particular, since

$$\mathbb{P}\left(T_0(A_{n,2} \circ C_{n,2}) - T_\varepsilon^\downarrow(A_{n,2} \circ C_{n,2}) \geq \delta\right) \leq \mathbb{P}\left(T_0(Z_{n,2} \circ C_{n,2}) \geq \delta \mid Z_{n,2}(0) = \varepsilon\right)$$

we obtain

$$\limsup_{n \to +\infty} \mathbb{P}\left(T_0(A_{n,2} \circ C_{n,2}) - T_\varepsilon^\downarrow(A_{n,2} \circ C_{n,2}) \geq \delta\right) \underset{\varepsilon \to 0}{\longrightarrow} 0$$

and the arguments of the first step imply that

$$\left(\mathbf{A}_n \circ C_{n,2}, T_0(A_{n,2} \circ C_{n,2})\right) \Rightarrow \left(\mathbf{A} \circ c_2, T_0(A_2 \circ c_2)\right).$$

In particular, the strong Markov property at time $T_0(A_{n,2} \circ C_{n,2})$ shows that

$$(\mathbf{A}_n \circ C_{n,2})\left(T_0(A_{n,2} \circ C_{n,2})\right) = \mathbf{A}_n(T_0(A_{n,2})) \Rightarrow (x \circ c_2)\left(T_0(A_2 \circ c_2)\right) = x(\infty) = 0.$$

By using the strong Markov property at time $T_0(A_{n,2})$ and iterating this argument, we finally end up with the desired result that $A_{n,0}(\infty) \Rightarrow A_0(\infty)$, which concludes the proof of Theorem 2.3.

6. AN INTRIGUING CONJECTURE

We conclude this paper by discussing a conjecture formulated in [2] which initially motivated the present work.

Conjecture (Antal and Krapivsky [2])**.** *Assume that $\varepsilon_{n,k} = \delta_{n,k} = 0$ and let $N_{n,k}$ be the number of individuals ever being of type $k = 1,\dots,K$ over the course of the epidemic, starting from the initial condition $a_{n,1}(0) = 1$ and $a_{n,k}(0) = 0$ for $k = 2,\dots,K+1$. Then $\mathbb{E}(N_{n,k})$ grows as $n \to +\infty$ like $n^{k\lambda_K}$, where*

$$\lambda_K = \frac{2^K - 1}{(K+1)2^K - 1}.$$

Note that, with our notation, $N_{n,K} = a_{n,K+1}(\infty)$, since any individual ever removed must have been in stage K of the epidemic at some point (and vice-versa). On the other hand, Theorem 2.3 shows that starting with of the order of $n^{1/(K+2)}$ individuals in stage one (instead of just one as in the above conjecture), $a_{n,K+1}(\infty)$ is, in distribution, of the order of $n^{(K+1)/(K+2)}$. When $K = 1$, there is a well-known argument that links these two objects: the connection goes through a random partitioning of $\{1,\dots,n\}$.

Consider indeed the following model, where n individuals are assigned a unique label from the set $\{1,\dots,n\}$ and which results in a random partition $\Pi_1,\dots,\Pi_S$ of the set $\{1,\dots,n\}$. Imagine that $\Pi_1,\dots,\Pi_s$ have been generated and that the set $F_s = \{1,\dots,n\} \setminus (\Pi_1 \cup \cdots \cup \Pi_s)$ is not empty: then the iteration proceeds as follows. Choose an individual v uniformly at random from F_s, and run the epidemic with the following initial condition: at time 0 the individuals in $F_s \setminus \{v\}$ are susceptible and the individuals in $\Pi_1 \cup \cdots \cup \Pi_s$ are removed, so that only v is infected (and, in the case $K \geq 2$, is in the first stage of the epidemic). Eventually, this epidemic will die out and we define Π_{s+1} as the set of individuals infected over the course of this epidemic.

This connection between random partition and epidemic processes is well-known, see for instance Barbour and Mollison [3]. In the case K=1 this links clusters of the Erdös-Rényi random graph to the SIR process. This construction leads to several interesting by-products, one of them being that it makes it possible to compute the mean size of a typical cluster in terms of the mean size of the largest ones. More precisely, choose an individual v uniformly at random in $\{1,\dots,n\}$ and let Π^* be the cluster to which it belongs. Let moreover $(\Pi_{(i)}, i \geq 1)$ be the clusters ordered in decreasing size, i.e., $\{\Pi_{(i)}\} = \{\Pi_i\}$ and $|\Pi_{(1)}| \geq |\Pi_{(2)}| \geq \cdots$ with $|E|$ the size of a set $E \subset \mathbb{N}$. Then the probability that v belongs to $\Pi_{(i)}$ is exactly $|\Pi_{(i)}|/n$, in which case $\Pi^* = \Pi_{(i)}$ and so

$$\mathbb{E}(|\Pi^*|) = \sum_{i \geq 1} \mathbb{E}\left(\frac{|\Pi_{(i)}|}{n} |\Pi_{(i)}| \right).$$

Assuming that the largest term in this sum dominates, we get the approximation

$$\mathbb{E}\left(|\Pi^*|\right) \approx \frac{1}{n} \mathbb{E}\left(|\Pi^2_{(1)}|\right). \tag{6.1}$$

When $K = 1$, $|\Pi^*|$ is, by exchangeability, equal in distribution to $N_{n,1}$ and so $\mathbb{E}(|\Pi^*|) = \mathbb{E}(N_{n,1})$. Moreover, $|\Pi_{(1)}|$ is of the same order than an outbreak size started from an *intermediate* initial condition. Roughly speaking, this comes from the fact that the intermediate regime is precisely the one where the finite-size population effects begin to kick in. In particular, $|\Pi_{(1)}|$ is of the order of $n^{(K+1)/(K+2)} = n^{2/3}$. Gathering these two observations, we end up thanks to (6.1) with the relation $\mathbb{E}(N_{n,1}) \approx n^{4/3-1} = n^{1/3}$, when $K = 1$. This answer coincides with Antal and Krapivsky's conjecture and the above reasoning

through random partitioning can be made rigorous, see, e.g., Janson and Luczak [12].

It is tempting to also use this reasoning for $K \geq 2$, and this initially motivated us to identify the intermediate regime and the scalings at play there. However, this reasoning would lead to the estimate

$$\mathbb{E}(N_{n,K}) \approx \frac{1}{n} n^{2(K+1)/(K+2)} = n^{K/(K+2)}$$

which is different (for $K \geq 2$) from the $n^{K\lambda_K}$ predicted by Antal and Krapivsky [2]. We find this discrepancy, and the related fact that the above reasoning through random partitioning seems to fail, very intriguing. We believe that the temporal aspects intrinsic to this multistage epidemic (see the discussion of the main results in the introduction) play a major role in this discrepancy, although it is challenging to obtain rigorous results in that direction.

Appendix A. Proof of Lemmas 3.1 and 3.2

Let x be any solution to (3.1) defined on the interval $J = [0, t^*)$ for some $t^* \in (0, \infty]$. It will be convenient to define $x_1 = y$ and to index the $\mathbb{R}^{K+1}$-valued function x by the set $\{0, 2, \dots, K+1\}$, i.e., to write $x = (x_0, x_2, \dots, x_{K+1})$. Let also in the sequel

$$I_k(t) = \int_0^t \big(x_0(u) - \gamma_k\big)\,\mathrm{d}u,\ t \in J, k = 1, \dots, K.$$

Then $(e^{I_k} x_k)' e^{-I_k} = (x_0 - \gamma_k) x_k + x_k'$ is equal to x_{k-1} for $k = 2, \dots, K$ by (3.1). Thus for these k we have $(e^{I_k} x_k)' e^{-I_k} = x_{k-1}$, which can be rewritten as

$$x_k(t) = \left(x_k(0) + \int_0^t x_{k-1}(u) e^{I_k(u)} \mathrm{d}u\right) e^{-I_k(t)},\ k = 2, \dots, K,\ t \in J. \tag{A.1}$$

A.1. **Proof of Lemma 3.1.** Note first that the representation (A.1) implies that $x(t) \in [0, \infty)^{K+1}$ for every $t \in J$. Indeed, since $x_1(t) = y(t) \geq 0$, this implies that $x_2(t) \geq 0$ and by induction on k, this implies that $x_k(t) \geq 0$ for every $k = 2, \dots, K$ and $t \in J$. Since finally $x_0(0), x_{K+1}(0) \in [0, \infty)$ and $x_0' = x_{K+1}' = x_K$ which has just been showed to stay non-negative, we obtain that x_0 and x_{K+1} also stay non-negative.

Let us now prove Lemma 3.1, i.e., existence and uniqueness of solutions to (3.1). Since F is locally Lipschitz, the Picard-Lindelhöf theorem implies local existence and uniqueness to (3.1). To show this globally, we only have to show that local solutions do not explode. Since $x_2' = y + (\gamma_2 - x_0) x_2$ and $x_2, x_0 \geq 0$, we obtain $x_2' \leq y + \gamma_2^+ x_2$ where $\gamma^+ = \max(0, \gamma)$ for any $\gamma \in \mathbb{R}$. Gronwall's lemma thus shows that x_2 does not explode. Similarly, for $k = 3, \dots, K$ we have $x_k' \leq x_{k-1} + \gamma_k^+ x_k$ and so x_k also does not explode. Finally, since $x_0' = x_{K+1}' = x_K$, solutions stay locally bounded which proves the global existence and uniqueness on $[0, \infty)$.

A.2. **Proof of Lemma 3.2.** We prove each property separately. As mentioned earlier, in the rest of the proof we define $x_1 = y$.

Proof of (i). Assume that $y(0) > 0$ or $a_2 > 0$: then it is clear from (A.1) that $x_2(t) > 0$ for all $t > 0$. By induction, we see that $x_k(t) > 0$ for all $k = 2, \dots, K$ and $t > 0$.

Proof of (ii). Assume that $x_k(t) > 0$ for $k = 2, \dots, K$ and $t > 0$, and that x_0 is bounded: we prove by backwards induction on k that $\int_0^\infty x_k$ is finite for $k = 1, \dots, K$ and that $x_0(\infty) > \gamma_k$ for $k = 2, \dots, K$. For $k = K$, the fact that $\int_0^\infty x_K$ is finite comes from the fact that x_0 is

bounded and non-decreasing, and so its derivative x_K is integrable on $[0,\infty)$. Consider now any $2 \le k \le K$ and assume that $\int_0^\infty x_k$ is finite: we prove that $\int_0^\infty x_{k-1}$ is finite and that $x_0(\infty) > \gamma_k$.

Since $\int_0^\infty x_k$ is finite, there must exist a sequence $t_n \to +\infty$ such that $x_k(t_n) \to 0$. Moreover, we have by definition $x_k' = x_{k-1} + (\gamma_k - x_0)x_k$ and so integrating between times 0 and t_n we obtain

$$\int_0^{t_n} x_{k-1} = x_k(t_n) - x_k(0) + \int_0^{t_n} (x_0(u) - \gamma_k) x_k(u) \mathrm{d}u \le x_k(t_n) + (x_0(\infty) - \gamma_k) \int_0^{t_n} x_k.$$

Letting $n \to +\infty$, we obtain the inequality $\int_0^\infty x_{k-1} \le (x_0(\infty) - \gamma_k) \int_0^\infty x_k$ which shows, by induction, that $\int_0^\infty x_{k-1}$ is finite and also that $x_0(\infty) > \gamma_k$ (since $\int_0^\infty x_{k-1} > 0$).

Proof of (iii). Assume that $y(t) = 0$ for all $t \ge 0$ and that $a_2 > 0$: in particular, according to (i), we see that $x_k(t) > 0$ for every $t > 0$ and $k = 0, 2, \dots, K$, a fact that will repeatedly be used in the sequel. We begin by proving the following formula:

$$x_k(t) = \left(\sum_{i=0}^{k-2} x_{k-i}(0) \phi_{k,i}(t) \right) \times \exp\left(- \int_0^t (x_0(u) - \gamma_k) \mathrm{d}u \right), \ k = 2, \dots, K, \ t \ge 0, \tag{A.2}$$

where the functions $\phi_{k,i}$ for $k = 2, \dots, K$ and $i = 0, \dots, k-2$ are defined recursively by $\phi_{k,0}(t) = 1$ and for $i = 1, \dots, k-2$,

$$\phi_{k,i}(t) = \int_0^t \phi_{k-1,i-1}(u) e^{\eta_k u} \mathrm{d}u, \ \text{ with } \ \eta_k = \gamma_{k-1} - \gamma_k \ \ (k = 3, \dots, K). \tag{A.3}$$

We prove (A.2) by induction on k. For $k = 2$, we have by (A.1), and since $x_1 = 0$, $x_2(t) = x_2(0) e^{-I_2(t)}$ which is precisely (A.2). Assume now that (A.2) holds for $k \ge 2$ and let us prove it for $k+1$: plugging in (A.2) into (A.1), we obtain

$$x_{k+1}(t) = \left(x_{k+1}(0) + \int_0^t \left(\sum_{i=0}^{k-2} x_{k-i}(0) \phi_{k,i}(u) \right) e^{-I_k(u)} \times e^{I_{k+1}(u)} \mathrm{d}u \right) e^{-I_{k+1}(t)}.$$

Using $I_{k+1}(u) - I_k(u) = (\gamma_k - \gamma_{k+1})u = \eta_{k+1} u$, exchanging the integral and the sum and changing variables in the sum, we obtain

$$x_{k+1}(t) = \left(x_{k+1}(0) + \sum_{i=1}^{k-1} x_{k+1-i}(0) \int_0^t \phi_{k,i-1}(u) e^{\eta_{k+1} u} \mathrm{d}u \right) e^{-I_{k+1}(t)}$$

from which we get (A.2) by (A.3). We now prove that $x_0(\infty)$ is finite by contradiction, so assume that $x_0(\infty) = +\infty$. Starting from the definition (A.3) of the $\phi_{k,i}$'s, we get by induction that $\phi_{k,i}(t) \le e^{i\eta^* t}$ with $\eta^* = 1 + \max_{3 \le j \le K} |\eta_j|$. In particular, we get from (A.2) for $k = K$ that

$$x_K(t) \le \left(\sum_{i=0}^{k-2} x_{k-i}(0) \right) e^{K\eta^* t} \times \exp\left(- \int_0^t (x_0(u) - \gamma_K) \mathrm{d}u \right).$$

Since $x_0(t) - \gamma_K > K\eta^*$ for t large enough (since we are assuming $x_0(\infty) = +\infty$), the last display implies that x_K converges to 0 exponentially fast, and in particular $\int_0^\infty x_K < +\infty$. Since $x_0' = x_K$, $x_0(\infty)$ is finite, which yields the contradiction. Thus $x_0(\infty)$ must be finite and so the conclusions of (ii) apply, in particular $x_0(\infty) > \gamma_k$ for every $2 \le k \le K$.

We now complete the proof and show that $x_k(t) \to 0$ for every $k = 2, \dots, K$: in view of (A.2) we only have to show that

$$\lim_{t \to +\infty} \left(\phi_{k,i}(t) e^{-\int_0^t (x_0 - \gamma_k)} \right) = 0 \tag{A.4}$$

for every $i = 0, \dots, K-2$ and every $k = i+2, \dots, K$. We prove this by induction on i: for $i = 0$ this comes immediately from the facts that $\phi_{k,0}(t) = 1$ and $x_0(\infty) > \gamma_k$. So assume that (A.4) holds for some $i = 0, \dots, K-2$ and every $k = i+2, \dots, K$: we show that it also holds for $i+1$ and $k = i+3, \dots, K$. By definition (A.3) we have

$$\phi_{k,i+1}(t) e^{-\int_0^t (x_0 - \gamma_k)} = \left(\int_0^t \phi_{k-1,i}(u) e^{\eta_k u} \mathrm{d}u \right) e^{-\int_0^t (x_0 - \gamma_k)}.$$

Let $\varepsilon > 0$ and, by induction hypothesis, t^* such that $\phi_{k-1,i}(t) \le \varepsilon e^{\int_0^t (x_0 - \gamma_{k-1})}$ for every $t \ge t^*$. Then for such t,

$$\begin{aligned} \phi_{k,i+1}(t) e^{-\int_0^t (x_0 - \gamma_k)} \le & \left(\int_0^{t^*} \phi_{k-1,i}(u) e^{\eta_k u} \mathrm{d}u \right) e^{-\int_0^t (x_0 - \gamma_k)} \\ & + \varepsilon \left(\int_0^t e^{\int_0^u (x_0 - \gamma_{k-1})} e^{\eta_k u} \mathrm{d}u \right) e^{-\int_0^t (x_0 - \gamma_k)}. \end{aligned}$$

Since the first term of the above upper bound vanishes as $t \to +\infty$ and we can rewrite the second term as

$$\left(\int_0^t e^{\int_0^u (x_0 - \gamma_{k-1})} e^{\eta_k u} \mathrm{d}u \right) e^{-\int_0^t (x_0 - \gamma_k)} = \left(\int_0^t e^{\int_0^u (x_0 - \gamma_k)} \mathrm{d}u \right) e^{-\int_0^t (x_0 - \gamma_k)} = \int_0^t e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u,$$

we obtain

$$\limsup_{t \to +\infty} \left(\phi_{k,i+1}(t) e^{-\int_0^t (x_0 - \gamma_k)} \right) \le \varepsilon \sup_{t \ge 0} \left(\int_0^t e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u \right).$$

Thus to achieve the proof we only have to show that this last supremum is finite. Let $\kappa > 0$ and $s^* < +\infty$ be such that $x_0(t) - \gamma_k \ge \kappa$ for $t \ge s^*$ and $k = 2, \dots, K$. Then for $t \ge s^*$

$$\int_{s^*}^t e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u \le \int_{s^*}^t e^{-\kappa(t-u)} \mathrm{d}u \le \int_0^\infty e^{-\kappa u} \mathrm{d}u$$

and $\int_0^{s^*} e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u \le \int_0^{s^*} e^{-\int_u^{s^*} (x_0 - \gamma_k)} \mathrm{d}u$ so that writing

$$\int_0^t e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u = \int_0^{s^*} e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u + \int_{s^*}^t e^{-\int_u^t (x_0 - \gamma_k)} \mathrm{d}u$$

achieves the proof.

REFERENCES

[1] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.

[2] Tibor Antal and P. L. Krapivsky. Outbreak size distributions in epidemics with multiple stages. *J. Stat. Mech. Theory Exp.*, (7):P07018, 2012.

[3] A. D. Barbour and D. Mollison. *Stochastic Processes in Epidemic Theory*, volume 86 of *Lecture Notes in Biomaths.*, chapter Epidemics and random graphs, pages 86–89. Springer, 1990.

[4] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.

[5] Sem Borst and Florian Simatos. A stochastic network with mobile users in heavy traffic. *Queueing Systems Theory Appl.*, 74(1):1–40, 2013.

[6] Maury Bramson. State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Systems Theory Appl.*, 30(1-2):89–148, 1998.

[7] R. G. Dolgoarshinnykh and Steven P. Lalley. Critical scaling for the SIS stochastic epidemic. *J. Appl. Probab.*, 43(3):892–898, 2006.

[8] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[9] Inge S. Helland. Continuity of a class of random time transformations. *Stochastic Processes Appl.*, 7(1):79–99, 1978.

[10] James M. Hyman, Jia Li, and E. Ann Stanley. The differential infectivity and staged progression models for the transmission of HIV. *Math. Biosci.*, 155(2):77–109, 1999.

[11] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.

[12] Svante Janson and Malwina J. Luczak. Susceptibility in subcritical random graphs. *J. Math. Phys.*, 49(12):125207, 23, 2008.

[13] David G. Kendall. Deterministic and stochastic epidemics in closed populations. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. IV*, pages 149–165, Berkeley and Los Angeles, 1956. University of California Press.

[14] Amaury Lambert and Florian Simatos. Asymptotic behavior of local times of compound Poisson processes with drift in the infinite variance case. To appear in *Journal of Theoretical Probability*.

[15] Amaury Lambert, Florian Simatos, and Bert Zwart. Scaling limits via excursion theory: Interplay between Crump-Mode-Jagers branching processes and Processor-Sharing queues. *Ann. Appl. Probab.*, 23(6):2357–2381, 2013.

[16] John Lamperti. Continuous state branching processes. *Bull. Amer. Math. Soc.*, 73:382–386, 1967.

[17] Anders Martin-Löf. Symmetric sampling procedures, general epidemic processes and their threshold limit theorems. *J. Appl. Probab.*, 23(2):265–282, 1986.

[18] Anders Martin-Löf. The final size of a nearly critical epidemic, and the first passage time of a Wiener process to a parabolic barrier. *J. Appl. Probab.*, 35(3):671–682, 1998.

[19] Martin I. Reiman. Some diffusion approximations with state space collapse. In *Modelling and performance evaluation methodology (Paris, 1983)*, volume 60 of *Lecture Notes in Control and Inform. Sci.*, pages 209–240. Springer, Berlin, 1984.

[20] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2: Itô calculus*. John Wiley & Sons Inc., New York, 1987.

[21] Alexander L. Stolyar. Maxweight scheduling in a generalized switch: state space collapse and workload minimization in heavy traffic. *Ann. Appl. Probab.*, 14(1):1–53, 2004.

[22] I. M. Verloop, U. Ayesta, and R. Núñez-Queija. Heavy-traffic analysis of a multiple-phase network with discriminatory processor sharing. *Oper. Res.*, 59(3):648–660, 2011.

[23] Bengt von Bahr and Anders Martin-Löf. Threshold limit theorems for some epidemic processes. *Adv. in Appl. Probab.*, 12(2):319–349, 1980.

[24] R. J. Williams. Diffusion approximations for open multiclass queueing networks: sufficient conditions involving state space collapse. *Queueing Systems Theory Appl.*, 30(1-2):27–88, 1998.