SUBLINEAR MULTI-TYPE PREFERENTIAL ATTACHMENT

ANNA BEN-HAMOU, NICOLAS BROUTIN, AND CAMILLE CAZAUX

ABSTRACT. We consider a random growing graph model with two *types*, or *communities*, that belongs to the class of preferential attachment models. Starting from an initial graph G_0 in which each node has a type, the graph G_{n+1} is generated by addition of a new node, which attaches through an edge to a node v of G_n. The probability of attaching to v depends on the communities of both the new node and v, and is proportional to $\deg_n(v)^\gamma$, where $\gamma \in [0,1]$. We establish a convergence result for the limiting proportion of vertices of each degree and in each community in a certain regime of the parameter γ. To this end, we first use stochastic approximation techniques to reduce the convergence problem to the study of a nonlinear evolution equation. We then use a probabilistic coupling approach to investigate the long-time behavior of the unique solution to this evolution equation. We emphasize that the techniques previously used to handle the single-community case cannot be extended directly to our setting. Indeed, the associated continuous-time embedding is not a multitype branching process in the usual sense, since the transition rates at any given time t depend on the entire graph structure at that time.

1. INTRODUCTION

1.1. **Setting and main results.** We study a random growing graph model with two *types*, or *communities* (A and B), that belongs to the class of preferential attachment models. Starting from an initial graph G_0 in which each node has a type, the graph G_{n+1} is obtained by adding to G_n a new node v_{n+1}. This node first receives a type drawn according to a Bernoulli distribution. Then, given its type, it attaches by an edge to a randomly chosen node v — belonging to the node set $V(G_n)$ of G_n — with a probability that depends on the type of v and sublinearly on its degree. More precisely, we consider the following random graph model.

Definition 1 (MODEL 1)**.** *Let p, q_A, q_B and γ be four parameters in $[0,1]$. Start with a fixed finite connected graph G_0, whose $n_0 := \#V(G_0)$ nodes are of type A or B. Then, for $n \geq 0$, the graph G_{n+1} is obtained from G_n by adding a new node v_{n+1} endowed with an edge. The attachment rule is the following:*

(i) *the type $\sigma(v_{n+1})$ of v_{n+1} is randomly chosen: type A with probability p, type B with probability $1-p$;*

(ii) (a) *given its own community, the new node chooses whether it will attach to a node in the same community with probability $1 - q_{\sigma(v_{n+1})}$, or in the other community with probability $q_{\sigma(v_{n+1})}$;*

(b) *it attaches to a node u in the community chosen at step (a) with probability proportional to $\deg_n(u)^\gamma$, where $\deg_n(u)$ denotes the degree of u in G_n.*

One may also consider a "softer" variant, referred to as MODEL 2, where step (ii) is replaced by:

(ii') the new node attaches to a node u of G_n with probability proportional to

$$\begin{cases} (1-q_{\sigma(v_{n+1})})\deg_n(u)^\gamma & \text{if } \sigma(v_{n+1})=\sigma(u), \\ q_{\sigma(v_{n+1})}\deg_n(u)^\gamma & \text{if } \sigma(v_{n+1})\neq\sigma(u). \end{cases}$$

More precisely, to make the difference between Model 1 and Model 2 more explicit, let $\Lambda_{n+1} := \sigma(v_{n+1})$ denote the type of the new node arriving at time $n+1$. The random variables $(\Lambda_n)_{n\geq 1}$, are *i.i.d.* with distribution Bernoulli(p). We then define the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ by setting $\mathcal{F}_n := \sigma((G_i)_{0\leq i\leq n}, (\Lambda_n)_{1\leq i\leq n})$, for each $n\in\mathbb{N}$. Let us also define, for $\Lambda\in\{A,B\}$, the random variable

$$D_n(\Lambda) := \sum_{\substack{v\in V(G_n)\\ \sigma(v)=\Lambda}} \deg_n(v)^\gamma\,,$$

and for $\Lambda\in\{A,B\}$, let $\overline{\Lambda}$ denote the "other" type. In Model 1, the conditional probability that the new node v_{n+1} attaches to a given node $u\in V(G_n)$, given $\mathcal{F}_n$ and Λ_{n+1} is

$$(1-q_{\Lambda_{n+1}})\frac{\deg_n(u)^\gamma}{D_n\left(\Lambda_{n+1}\right)}\mathbb{1}_{\{\sigma(u)=\Lambda_{n+1}\}} + q_{\Lambda_{n+1}}\frac{\deg_n(u)^\gamma}{D_n\left(\overline{\Lambda}_{n+1}\right)}\mathbb{1}_{\{\sigma(u)\neq\Lambda_{n+1}\}}\,,$$

whereas in Model 2, this probability is

$$\frac{(1-q_{\Lambda_{n+1}})\deg_n(u)^\gamma\mathbb{1}_{\{\sigma(u)=\Lambda_{n+1}\}} + q_{\Lambda_{n+1}}\deg_n(u)^\gamma\mathbb{1}_{\{\sigma(u)\neq\Lambda_{n+1}\}}}{(1-q_{\Lambda_{n+1}})D_n\left(\Lambda_{n+1}\right) + q_{\Lambda_{n+1}}D_n\left(\overline{\Lambda}_{n+1}\right)}\,.$$

Remark 1. *In both models, the parameters q_A and q_B govern the inclination to attach to the "other" community. Let us mention here that we will sometimes write p_A and p_B instead of p and $1-p$. Finally, observe that one may recover the one-community case by setting $p=1$ and $q_A=0$.*

Our goal is to determine the limiting degree distribution in G_n when n tends to $+\infty$. We are thus led to consider the sequence of counting measures $(m_n)_{n\in\mathbb{N}}$ on $E := \mathbb{N}^*\times\{A,B\}$ given by[1]

$$\forall n\geq 0, \forall x\in E,\quad m_n(x) = \#\{v\in V(G_n) : (\deg_n(v),\sigma(v)) = x\}\,. \tag{1}$$

In other words, if $x=(k,\Lambda)$, then $m_n(x)$ counts the number of nodes in G_n with degree k and type Λ and $\frac{m_n(x)}{n+n_0}$ denotes the proportion of such nodes in G_n. The counting measure m_n will be identified to a sequence indexed by the lexicographic order on E: $(1,A),(1,B),(2,A),(2,B)...$

Our main contribution is to establish almost sure convergence in total variation distance of the renormalized counting measure $\frac{m_n}{n+n_0}$ (as n goes to infinity): in a certain parameter regime $\gamma\in[0,\gamma_0)$ (γ_0 being a strictly positive constant in $(0,1]$) in the two-community case (Theorem 1), and for any choice of $\gamma\in[0,1]$ satisfying a rather remarkable inequality in the one-community case (Theorem 2).

Theorem 1 (Limiting degree distribution in the two-community case)**.** *Let $(G_n)_{n\geq 0}$ be the sequence of random graphs generated according to* Model *$i\in\{1,2\}$ with parameters*

[1]Throughout this paper, we denote by $\mathbb{N}^* = \{1,2,3,\dots\}$ the set of positive integers and by $\mathbb{N} = \{0\}\cup\mathbb{N}^*$ the set of nonnegative integers.

p, q_A, q_B and γ all in $[0,1]$. There exists a probability measure $\mu^\star_{(i)}$ on E and a constant $0 < \gamma_0 \leq 1$ such that for $\gamma \in [0, \gamma_0)$, we have

$$\left\| \frac{m_n}{n_0 + n} - \mu^\star_{(i)} \right\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{\text{a.s.}} 0\,.$$

We denote by $\|\cdot\|_{\mathrm{TV}}$ the total variation distance, with the convention that $\|\cdot\|_{\mathrm{TV}} = \frac{1}{2}\|\cdot\|_1$, where $\|\cdot\|_1$ denotes the ℓ^1-norm.

Remark 2. *By the results of Jordan [12] and Bhamidi et al. [5], this was already known for $\gamma = 0$ and $\gamma = 1$ (they showed pointwise convergence to a limiting probability distribution, which is equivalent to convergence in total variation by Scheffé's lemma, since the state space is countable). We conjecture that the result of Theorem 1 holds for all $\gamma \in [0,1]$. The precise value of our γ_0 is implicitly defined in (28). Let us point out here that for $p = \frac{1}{2}$ and $q_A = q_B = q$, one can take $\gamma_0 = \frac{1}{7}$.*

Theorem 2 (Limiting degree distribution in the one-community case)**.** *Let $(G_n)_{n\in\mathbb{N}}$ be the sequence of random graphs generated according to Definition 1 with parameters $p = 1$, $q_A = 0$ and $\gamma \in [0,1]$, corresponding to the single-community case. Let $E = \mathbb{N}^*$ and, for $n \in \mathbb{N}$, define the counting measure m_n by*

$$m_n(x) = \#\{v \in V(G_n) : \deg_n(v) = x\}\,,$$

for each $x \in E$. There exists a probability measure $\mu^\star$ on E such that, for each $\gamma \in [0,1]$ satisfying the following inequality

$$\sum_{k\geq 1} \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1} > 1\,, \tag{2}$$

we have

$$\left\| \frac{m_n}{n_0 + n} - \mu^\star \right\|_{\mathrm{TV}} \xrightarrow[n\to\infty]{\text{a.s.}} 0\,.$$

1.2. **Related work.** Complex networks arise in a wide variety of settings, including technological, biological, social, and information systems. In many real-world networks, empirical degree distributions exhibit a power-law tail, that is, the probability that a randomly chosen node has degree k is approximately proportional to $k^{-\alpha}$, where α often lies between 2 and 3. Such networks typically contain many vertices with relatively few connections and a small fraction of highly connected vertices, known as hubs. This observation has motivated the study of growing network models, in which new vertices and edges are added over time. In 1999, Barabási and Albert introduced the *preferential attachment* mechanism, whereby new vertices are more likely to connect to vertices that already have high degrees (see [3]). This self-reinforcing mechanism can give rise to power-law decay of the degree distribution, a precise mathematical formulation and proof of this result were given in [6] and independently in [17]. The classical Barabási–Albert model is recovered as a special case of our model in the single-community setting when $\gamma = 1$. This model is now well understood. By contrast, in the classical Erdős–Rényi model $G(n,p)$, node degrees follow a binomial distribution and, in the usual regimes, concentrate around a characteristic scale. Power-law degree distributions, by comparison, are heavy-tailed and are commonly described as *scale-free*, reflecting the absence of a characteristic degree scale.

Several extensions of the model have been studied. The case of uniform attachment, corresponding to $\gamma = 0$ in our setting, is well understood and corresponds to the uniform recursive tree. When $\gamma > 1$, a condensation phenomenon occurs: in the limit, a single random node dominates the network by accumulating a disproportionate number of connections. This phenomenon was already known in the physics literature, notably through the work of Krapivsky, Redner, and Leyvraz (see [13]), before being established rigorously in subsequent mathematical studies (see, for example, [18], and more recently [21]). More generally, the probability governing the choice of the node or vertices to which a new node attaches may depend on the degrees in a variety of ways. Given a weight function $w : \mathbb{N} \to \mathbb{R}_+$, the attachment probability may be proportional to w evaluated at the degree of a node. In [18] and [21], the weight function w is superlinear, and the techniques developed therein rely on martingale arguments that are particularly well suited to this setting. The paper [2] studies preferential attachment graphs with specific sublinear weight functions $w(k) = ck^\gamma$ (with $\gamma \in (\frac{1}{2}, 1)$), linear functions $w(k) = ak + b$, and superlinear functions, showing that they lead to distinct asymptotic behaviors: slow degree growth, power-law degree distributions, and extreme concentration of connections, respectively. These results are established using probabilistic methods, particularly continuous-time embeddings and pure birth processes.

Using probabilistic asymptotic techniques based on embeddings of random tree growth processes into continuous-time general branching processes, Rudas [19] and Rudas et al. [20] determine the limiting degree distribution and the local limit in the sublinear regime. In [20], the main results rely on branching process limit theorems and random characteristics to establish the almost-sure convergence of the empirical distributions of node degrees and descendant subtrees. By contrast, to the best of our knowledge, the limiting degree distribution of the model in the sublinear regime with communities has not yet been investigated.

Meanwhile, when $\gamma = 0$ or $\gamma = 1$, community-based variants of the Barabási–Albert model have also been investigated. In particular, [5] studies the case of our Model 1 with $q_A = q_B = q$ and $\gamma \in \{0, 1\}$. Since individual reproduction rates depend on the overall population composition, the resulting process is not a classical branching process. By combining continuous-time embedding techniques with stochastic analysis, the authors establish the convergence of certain macroscopic graph quantities and use these results to derive further asymptotic properties. In addition, Jordan considers the case of finitely many communities in [12], and even extends the framework to a compact metric space of communities, still under linear preferential attachment. More generally, he studies an extension of the model in which each new node arrives with a fixed number $m \in \mathbb{N}^*$ of edges. His approach consists in first establishing a convergence result for the analogues of the ratios $\frac{D_n(A)}{n}$ and $\frac{D_n(B)}{n}$ in our model, using an initial stochastic approximation result, as well as on the existence of a suitably chosen Lyapunov function. This convergence result is then used to derive a new recurrence relation for the proportion of vertices of given degrees and types in the sequence of graphs $(G_n)_{n\in\mathbb{N}}$. Finally, a second stochastic approximation result is applied to establish the desired convergence.

The difficulty in extending the approach of Jordan lies in the need to establish beforehand the convergence of the ratios $\frac{D_n(A)}{n}$ and $\frac{D_n(B)}{n}$. It is not clear to us whether a Lyapunov function satisfying the definition given in [12] can be found in our setting. This is why we

found it necessary to develop a new approach to address the sublinear multitype case, which consists in establishing a convergence result, as n tends to infinity, for the counting measure m_n describing the proportion of vertices associated with each pair $(k, \Lambda) \in \mathbb{N}^* \times \{A, B\}$.

Furthermore, the paper [1] studies a class of attribute-based preferential attachment models, in which new vertices join the network sequentially and connect to existing vertices according to their degrees and attributes, which play a role analogous to community labels in community-based network models. The authors establish asymptotic results on the local structure of these networks, the tail behavior of degree and PageRank score distributions, and the performance of sampling methods in capturing different attribute groups. Their analysis combines local weak convergence, continuous-time multitype branching processes, and stochastic approximation techniques (drawing in particular on the work of Jordan in [12]) to characterize the limiting network structure and assess sampling biases. However, most of this work focuses on the case where the weight function is linear.

Remark 3. *For the study of our two-community model, one could similarly introduce the following continuous-time multitype branching process. Without loss of generality, assume that G_0 consists of a single node, $\emptyset$. This initial node is assigned type A with probability p and type B with probability $1-p$. The initial node gives birth to vertices of types A and B at time-dependent rates to be specified just below. Similarly, each newly born node, in turn, gives birth to new vertices of types A and B at the following time-dependent rates. Let G_t denote the graph obtained at time t and let $V(G_t)$ be its node set. In* Model 1, *a node $v \in V(G_t)$ of type $\Lambda := \sigma(v)$ gives birth to a node of type Λ' at rate*

$$r^{(1)}_{v\to\Lambda'}(t) = \frac{p_{\Lambda'} \times [q_{\Lambda'}\mathbb{1}_{\Lambda\neq\Lambda'} + (1-q_{\Lambda'})\mathbb{1}_{\Lambda=\Lambda'}] \times \deg_t(v)^\gamma}{D_t(\Lambda)},$$

where $\deg_t(v)$ denotes the degree of v in G_t and $D_t(\Lambda)$ the following continuous-time analogue of $D_n(\Lambda)$

$$D_t(\Lambda) := \sum_{\substack{v\in V(G_t)\\ \sigma(v)=\Lambda}} \deg_t(v)^\gamma .$$

In Model 2, *we have*

$$r^{(2)}_{v\to\Lambda'}(t) = \frac{p_{\Lambda'} \times [q_{\Lambda'}\mathbb{1}_{\Lambda\neq\Lambda'} + (1-q_{\Lambda'})\mathbb{1}_{\Lambda=\Lambda'}] \times \deg_t(v)^\gamma}{(1-q_{\Lambda'})D_t(\Lambda') + q_{\Lambda'}D_t(\overline{\Lambda'})}.$$

We then define the sequence of stopping times $(T_n)_{n\geq 0}$, corresponding to the times at which a new node appears in the randomly growing tree

$$T_n := \inf\{t \geq 0 : |V(G_t)| = n+1\} .$$

With this notation, the sequence of growing graphs $(G^{\mathrm{d}}_n)_{n\geq 0}$ generated by the discrete-time model defined in Definition 1 is related to the continuous-time model $(G_t)_{t\geq 0}$ introduced above through the following equality in distribution:

$$G^{\mathrm{d}}_n \overset{(\mathrm{d})}{=} G_{T_n} .$$

Yet a key difference arises at this point: the point processes associated with each node of the continuous-time growing graph, which describe the birth times and types of its offspring, depend both on the type of the parent node and on the structure of the entire graph. To generalize the approach used in [20], one could rely on multi-type extensions of

the convergence results employed therein. In this regard, a paper by Jagers and Nerman seems particularly relevant (see [11]). However, to apply the result of this paper, as well as to extend the approach of [20] to the two-community setting, one would first need to establish a convergence result for the ratios $\frac{D_n(\Lambda)}{D_n(\Lambda)+D_n(\overline{\Lambda})}$ *or* $\frac{D_n(\Lambda)}{n}$*, for* $\Lambda \in \{A, B\}$*. The continuous-time embedding of our model cannot be interpreted as a branching process in the usual sense (as defined, for instance, in [11]), since the individual reproduction rates at any given time* t *depend on the composition of the entire population at that time. This obstacle is all the more interesting in that it also arises when attempting to generalize the approach used by Jordan in [12] where* $\gamma = 1$*. In addition, it is worth noting that, Theorem 1 is the key step in justifying the approximation of the continuous-time embedding introduced above by a branching process whose rates at time* t *no longer depend on the entire graph structure at that time. We emphasize that the rates of this approximating branching process are obtained by multiplying the original rates by* $D_t(\Lambda) + D_t(\overline{\Lambda})$*, which corresponds to a time change of the continuous-time process* (G_t)*, and then replacing the ratios* $\frac{D_t(\Lambda)}{D_t(\Lambda)+D_t(\overline{\Lambda})}$ *and* $\frac{D_t(\overline{\Lambda})}{D_t(\Lambda)+D_t(\overline{\Lambda})}$ *appearing in the resulting rates with their respective limits.*

1.3. **Proof strategy and outline of the paper.**

1.3.1. *Measure-valued Pólya processes.* The process $(m_n)_{n\in\mathbb{N}}$ can be described using the formalism of *measure-valued Pólya processes*, as studied in [14] and [15]. These processes provide a generalization of Pólya urn processes by allowing the number of colors to be infinite, with values in a Polish space E. In our setting, starting from a measure m_0 encoding the number of nodes in G_0 with a given degree and a given type, we obtain m_{n+1} from m_n in the following way. At time $n+1$, the new node v_{n+1} gets a type Λ_{n+1} which is equal to A with probability p, or to B with probability $1-p$. We let B_{n+1} be the Bernoulli random variable given by $B_{n+1} = \mathbb{1}_{\{\Lambda_{n+1}=A\}}$. Letting $X_{n+1} \in E$ the pair corresponding to the degree and type of the node to which v_{n+1} attaches to, we have, in Model 1, for all $x \in E$, and $\Lambda \in \{A, B\}$,

$$\mathbb{P}(X_{n+1} = x \mid m_n, \Lambda_{n+1} = \Lambda) = (1 - q_\Lambda)\frac{m_n D^\Lambda(x)}{m_n D^\Lambda(E)} + q_\Lambda \frac{m_n D^{\overline{\Lambda}}(x)}{m_n D^{\overline{\Lambda}}(E)},$$

where D^A is the infinite diagonal matrix with diagonal entries: $1, 0, 2^\gamma, 0, 3^\gamma, 0...$, and D^B is the infinite diagonal matrix with diagonal entries: $0, 1, 0, 2^\gamma, 0, 3^\gamma....$ We specify that if η is a measure on E and M is an infinite matrix in $\mathcal{M}_{E\times E}(\mathbb{R})$, then for $x \in E$, we have $\eta M(x) = \sum_{y\in E} \eta(y)M(y, x)$, and $\eta M(E) = \sum_{x\in E} \eta M(x)$.

In Model 2, we have for all $x \in E$ and $\Lambda \in \{A, B\}$,

$$\mathbb{P}(X_{n+1} = x \mid m_n, \Lambda_{n+1} = \Lambda) = \frac{m_n P^\Lambda(x)}{m_n P^\Lambda(E)},$$

where $P^\Lambda = (1 - q_\Lambda)D^\Lambda + q_\Lambda D^{\overline{\Lambda}}$.

Put differently, in Model 2, setting $P^{(n+1)} = B_{n+1}P^A + (1 - B_{n+1})P^B$, we have

$$\mathbb{P}(X_{n+1} = x \mid m_n, \Lambda_{n+1}) = \frac{m_n P^{(n+1)}(x)}{m_n P^{(n+1)}(E)}.$$

In both models, the replacement rule is then: $m_{n+1} = m_n + R^{(n+1)}_{X_{n+1}}$, with

$$R^{(n+1)}_{(k,\Lambda)} = \delta_{(1,\Lambda_{n+1})} - \delta_{(k,\Lambda)} + \delta_{(k+1,\Lambda)}\,.$$

The matrix $R^{(n+1)}$, whose rows are $(R^{(n+1)}_x)_{x\in E}$, is called the replacement matrix. As in [15], the replacement matrices are random: $R^{(n+1)}$ depends on the random type Λ_{n+1}.

Remark 4. *While we use the formalism of measure-valued Pólya processes presented in [15], it is important to note that none of the two models falls into the framework of this paper, in which the weight matrices are all equal to the same deterministic matrix P.* Model 2 *could be seen as a generalization where the weight matrices $P^{(n)}$ are random (with the same source of randomness as the one of the replacement matrices $R^{(n)}$).*

1.3.2. *Stochastic approximation.* Rather than studying the process (m_n) directly, it is convenient to first focus on the process $(\eta_n)_{n\geq 0}$, defined by $\eta_0 = 0$ and for $n \geq 1$, $\eta_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. By definition, we have

$$\eta_{n+1} = \left(1 - \frac{1}{n+1}\right)\eta_n + \frac{1}{n+1}\delta_{X_{n+1}}\,.$$

From now on, we focus on Model 1. Observe that

$$\mathbb{E}\left[\delta_{X_{n+1}} \,|\, \mathcal{F}_n\right] = r_A \frac{m_n D^A}{m_n D^A(E)} + r_B \frac{m_n D^B}{m_n D^B(E)}\,,$$

where $r_\Lambda = p_\Lambda(1-q_\Lambda) + p_{\overline{\Lambda}} q_{\overline{\Lambda}}$, and that

$$m_n = m_0 + \sum_{i=1}^n R^{(i)}_{X_i} = m_0 + \sum_{i=1}^n \delta_{X_i} R^{(i)}\,, \tag{3}$$

where

$$R^{(i)} = \begin{pmatrix} B_i & 1-B_i & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \\ B_i & 1-B_i & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \end{pmatrix} + \begin{pmatrix} -1 & 0 & 1 & 0 & & (0) \\ & \ddots & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 & 0 \\ & (0) & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

If in (3) each matrix $R^{(i)}$ could be approximated by its expectation R, where B_i is replaced by p, we would have

$$\frac{m_n}{n} \approx \eta_n R\,.$$

Defining $Q^A = RD^A$ and $Q^B = RD^B$, we would get

$$\mathbb{E}\left[\delta_{X_{n+1}} \,|\, \mathcal{F}_n\right] \approx r_A \frac{\eta_n Q^A}{\eta_n Q^A(E)} + r_B \frac{\eta_n Q^B}{\eta_n Q^B(E)}\,.$$

This heuristic leads us to write

$$\eta_{n+1} = \eta_n + \frac{1}{n+1}\left(r_A \frac{\eta_n Q^A}{\eta_n Q^A(E)} + r_B \frac{\eta_n Q^B}{\eta_n Q^B(E)} - \eta_n + \varepsilon_{n+1}\right),$$

where

$$\varepsilon_{n+1} = \delta_{X_{n+1}} - r_A \frac{\eta_n Q^A}{\eta_n Q^A(E)} - r_B \frac{\eta_n Q^B}{\eta_n Q^B(E)}\,.$$

In Section 2, we show, using stochastic approximation techniques developed in [4], that for all $\gamma \in [0,1)$, an affine interpolation of $(\eta_n)_{n\geq 0}$ with appropriate step sizes is an *asymptotic*

pseudo-trajectory of the semiflow on $\mathcal{P}(E)$ (the set of probability distributions on E) induced by

$$\partial_t \eta_t = F(\eta_t)\,,$$

for some explicit function F from $\mathcal{P}(E)$ to $\mathcal{M}_0(E)$, the set of signed measures on E with zero sum (Proposition 4).

In Section 3, we show that there exists a unique fixed point to the continuous-time dynamics induced by F, *i.e.* a unique probability distribution $\eta^\star \in \mathcal{P}(E)$ satisfying $F(\eta^\star) = 0$ (Proposition 9). The next task is to show that η_t converges to $\eta^\star$ as t tends to infinity.

1.3.3. *Non-linear Markov processes and coupling.* Since the dynamics induced by F preserves probability distributions, we would like to view it as distribution of some stochastic process on E. However, no such process is immediately given from F itself. Indeed, we have $F(\eta) = \eta A_\eta$ but the matrix A_η is not an infinitesimal generator because the sum on each row is not equal to zero. Still, one may design a small perturbation L_η of A_η, which is such that $F(\eta) = \eta L_\eta$, which has negative entries on the diagonal and which has zero sum on each row. In the two-community case, to ensure that all the off-diagonal entries of L_η are positive, we could not do without a condition on γ (see Assumption 1), but such a condition is not needed in the one-community setting where the corresponding matrix L_η corresponds to an infinitesimal generator for all $\gamma \in [0,1]$. In Section 4, we first focus on this last setting, which allows us to introduce in a simpler setting the arguments that will be used for two communities. Having described our dynamics as the one of a non-linear Markov process on $\mathbb{N}^*$ (the non-linearity coming from the fact that the generator L_η depends on the measure η itself), we use coupling methods inspired by [16] to establish that, for any starting distribution, the equation $\partial_t \eta_t = \eta_t L_{\eta_t}$ has a unique solution on $\mathbb{R}_+$ (Proposition 12), and that, whatever the starting distribution, this solution converges to $\eta^\star$ in total variation (Proposition 13). This last result holds for all $\gamma \in [0,1]$ satisfying inequality (2). We strongly believe that this inequality holds for all $\gamma \in [0,1)$ (see Figure 3), but we only have a simple analytic proof for $\gamma \in \left[0, \frac{\ln 2}{\ln 3}\right]$. In Section 5, we come back to the two-community setting. We first use a fixed point theorem to establish existence and uniqueness of a solution to $\partial_t \eta_t = \eta_t L_{\eta_t}$ (Proposition 15). Then we adapt the coupling arguments that will be introduced in Section 4 to the case of two communities and show that, for any starting distribution, the unique solution to $\partial_t \eta_t = \eta_t L_{\eta_t}$ converges to $\eta^\star$. Once this is done, the proof of Theorem 1 (and of Theorem 2) easily follows. Section 6 highlights the few changes that have to be made to adapt the proof to Model 2.

2. Stochastic approximation

Recall from Section 1.3.2 that the approximation $\frac{m_n}{n} \approx \eta_n R$ led us to write

$$\eta_{n+1} = \eta_n + \frac{1}{n+1}\left(r_A \frac{\eta_n Q^A}{\eta_n Q^A(E)} + r_B \frac{\eta_n Q^B}{\eta_n Q^B(E)} - \eta_n + \varepsilon_{n+1} \right)\,, \tag{4}$$

where

$$\varepsilon_{n+1} = \delta_{X_{n+1}} - r_A \frac{\eta_n Q^A}{\eta_n Q^A(E)} - r_B \frac{\eta_n Q^B}{\eta_n Q^B(E)}\,.$$

The next lemma, which quantifies the approximation $\frac{m_n}{n} \approx \eta_n R$, will be used repeatedly.

Lemma 3. *For all $n \geq 1$, we have*

$$\frac{m_n}{n} - \eta_n R = \frac{m_0}{n} + \frac{1}{n}\sum_{i=1}^{n}(B_i - p, p - B_i, 0, 0, \dots)\,.$$

Proof. Using (3), we have

$$\begin{aligned}\frac{m_n}{n} - \eta_n R &= \frac{m_0}{n} + \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}(R^{(i)} - R)\\ &= \frac{m_0}{n} + \frac{1}{n}\sum_{i=1}^{n}\delta_{X_i}\begin{pmatrix} B_j - p & p - B_j & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \\ B_j - p & p - B_j & 0 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & & \vdots & \end{pmatrix}\\ &= \frac{m_0}{n} + \frac{1}{n}\sum_{i=1}^{n}(B_i - p, p - B_i \dots, 0, \dots)\,,\end{aligned}$$

which concludes the proof. □

It will be convenient to slightly change the step sizes and to rewrite Equation (4) in the form

$$\eta_{n+1} = \eta_n + \gamma_{n+1}\left\{F(\eta_n) + \varepsilon'_{n+1}\right\}\,,$$

where

$$\gamma_{n+1} = \frac{1}{n+1}\left(\frac{r_A}{\eta_n Q^A(E)} + \frac{r_B}{\eta_n Q^B(E)}\right)\,,$$

$$\varepsilon'_{n+1} = \frac{\varepsilon_{n+1}}{\dfrac{r_A}{\eta_n Q^A(E)} + \dfrac{r_B}{\eta_n Q^B(E)}},$$

and where

$$F(\eta) = s(\eta)\left(\eta Q^A - \eta Q^A(E)\eta\right) + (1 - s(\eta))\left(\eta Q^B - \eta Q^B(E)\eta\right)\,, \tag{5}$$

with

$$s(\eta) = \frac{r_A \eta Q^B(E)}{r_B \eta Q^A(E) + r_A \eta Q^B(E)}\,. \tag{6}$$

We introduce the following notations for later use:

$$s^A(\eta) = s(\eta) \quad \text{and} \quad s^B(\eta) = 1 - s(\eta).$$

Note that γ_{n+1} is of the order $\frac{1}{n+1}$. Indeed, for all $\eta \in \mathcal{P}(E)$ and for $\Lambda \in \{A, B\}$, we have

$$p_\Lambda \leq \eta Q^\Lambda(E) = p_\Lambda + \sum_{k\geq 1}\eta(k,\Lambda)\left((k+1)^\gamma - k^\gamma\right) \leq p_\Lambda + 2^\gamma - 1\,.$$

We then consider the increasing sequence $(\tau_n)_{n\geq 0}$ defined by $\tau_0 = 0$ and for all $n \geq 1$,

$$\tau_n = \gamma_1 + \gamma_2 + \dots + \gamma_n\,,$$

and we introduce the linearized version $(X(t))_{t\geq 0}$ of $(\eta_n)_{n\in\mathbb{N}}$ defined, for all $n \geq 1$ and all $t \in [\tau_n, \tau_{n+1}]$, by

$$X(t) = \eta_n + \frac{t-\tau_n}{\tau_{n+1}-\tau_n}(\eta_{n+1} - \eta_n)\,.$$

Now, let Φ be the semiflow induced by the dynamical system

$$\partial_t \eta_t = F(\eta_t)\,, \tag{7}$$

with F defined in (5), that is, the semiflow given by

$$\begin{aligned} \Phi\colon \mathbb{R}_+ \times \mathcal{P}(E) &\to \mathcal{P}(E) \\ (t,\eta) &\mapsto \Phi(t,\eta) = \Phi_t(\eta) = \eta_t\,, \end{aligned}$$

where η_t is the state at time t of the system given by $\partial_t\eta_t = F(\eta_t)$, with initial condition η. The fact that this semiflow is well-defined will be established in Proposition 15. The rest of this section is devoted to the proof of the following result.

Proposition 4. *For all $\gamma \in [0,1)$, the interpolated process $(X(t))_{t\geq 0}$ is an asymptotic pseudo-trajectory of the semiflow Φ induced by F, that is, for all $T > 0$, we have*

$$\lim_{t\to+\infty} \sup_{0\leq h\leq T} \|X(t+h) - \Phi_h(X(t))\|_1 = 0\,.$$

To prove Proposition 4, we apply [4][Theorem 3.2] which reduces the problem to showing three points:

(1) uniform continuity: the function $X : \mathbb{R}_+ \to \mathcal{P}(E)$ is uniformly continuous;
(2) compact closure: the image of $X : \mathbb{R}_+ \to \mathcal{P}(E)$ has compact closure in $\mathcal{P}(E)$;
(3) every limit point of the translation flow[2] applied to X belongs to the set

$$\{t \mapsto \Phi_t(\eta), \eta \in \mathcal{P}(E)\}\,.$$

Point (1) is established in Section 2.1, Point (2) in Section 2.2, Point (3) in Section 2.3.

2.1. Uniform continuity.

Lemma 5. *The function $X : \mathbb{R}_+ \to \mathcal{P}(E)$ is uniformly continuous.*

Proof. For $t \geq 0$, let us define

$$m(t) := \sup\{k \geq 0, t \geq \tau_k\}\,.$$

Using a slight reformulation of [4][Equation (9)], we have, for all $s, t \geq 0$,

$$X(t+s) - X(t) = \int_t^{t+s} \frac{1}{\gamma_{m(u)+1}(m(u)+1)} \left(\delta_{X_{m(u)+1}} - \eta_{m(u)}\right) ds\,.$$

Using that $\gamma_{n+1}(n+1)$ is uniformly bounded away from 0, and that both $\delta_{X_{n+1}}$ and η_n are probability distributions, we obtain that there exists $c > 0$ such that for all $s, t \geq 0$,

$$\|X(t+s) - X(t)\|_1 \leq cs\,.$$

In particular, the function X is uniformly continuous. □

[2]The definition of the translation flow will be given in Section 2.3.

2.2. **Compact closure.** To prove that $\{X(t), t \in \mathbb{R}_+\}$ has compact closure, let us show the following.

Lemma 6. *For all $\gamma \in [0,1[$, there exists $\delta > 0$ small enough, and a random value $C \in \mathbb{R}_+$, such that, defining $W : E \to \mathbb{N}^*$ by*

$$\forall (k, \Lambda) \in E\,,\ W(k, \Lambda) = k^\delta\,.$$

we have

$$\forall n \geq 1\,, \langle \eta_n, W\rangle := \sum_{x \in E} \eta_n(x) W(x) \leq C\,.$$

Proof. By (4), letting $u_n = \langle \eta_n, W\rangle$, we have, for all $n \geq 1$,

$$u_{n+1} = u_n + \frac{1}{n+1}\left(r_A \left\langle \frac{\eta_n Q^A}{\eta_n Q^A(E)}, W \right\rangle + r_B \left\langle \frac{\eta_n Q^B}{\eta_n Q^B(E)}, W \right\rangle - u_n + \langle \varepsilon_{n+1}, W\rangle\right)\,.$$

On the one hand, if $\delta > 0$ is chosen small enough so that $\gamma + \delta < 1$, then for $\Lambda \in \{A, B\}$, we have

$$\begin{aligned}
\left\langle \frac{\eta_n Q^\Lambda}{\eta_n Q^\Lambda(E)}, W \right\rangle &= \frac{1}{\eta_n Q^\Lambda(E)} \langle \eta_n, Q^\Lambda W\rangle \\
&= \frac{1}{\eta_n Q^\Lambda(E)} \left(p_\Lambda + \sum_{k \geq 1} \eta_n(k, \Lambda) \left((k+1)^{\gamma+\delta} - k^{\gamma+\delta} \right) \right) \\
&\leq \frac{p_\Lambda + 1}{p_\Lambda}\,.
\end{aligned}$$

Hence there exists a real-valued constant $a > 0$ such that

$$r_A \left\langle \frac{\eta_n Q^A}{\eta_n Q^A(E)}, W \right\rangle + r_B \left\langle \frac{\eta_n Q^B}{\eta_n Q^B(E)}, W \right\rangle \leq a\,.$$

We thus have

$$u_{n+1} \leq u_n + \frac{1}{n+1}\left(a - u_n + \langle \varepsilon_{n+1}, W\rangle\right)\,.$$

If we could show that

$$L := \sup_{1 \leq n \leq N} \left| \sum_{i=n}^{N} \frac{1}{i+1} \langle \varepsilon_{i+1}, W\rangle \right|$$

is finite almost surely, then we would get that for all $n \geq 1$, $u_n \leq 2(L + c)$, with $c = a \vee u_1$. Indeed, this is immediate for $n = 1$ and, assuming that it is true for some $n \geq 1$:

- if $u_n \leq c$, then $u_n + 1 \leq c + a + L$ and the result holds for $n + 1$;
- if $u_n > c$, then define n_0 by

  $$n_0 = \sup\{k \in \{1, \ldots, n\} : u_k > c, u_{k-1} \leq c\}\,,$$

  which is well defined because $u_1 \leq c$. By the above, we have $u_{n_0} \leq c + a + L$, and, since $u_{n_0}, \ldots, u_n > c$, we have

  $$u_n \leq u_{n_0} + \sum_{i=n_0}^{n} \frac{1}{i+1} \langle \varepsilon_{i+1}, W\rangle \leq c + a + 2L\,,$$

  and the result holds for $n + 1$.

To prove that L is finite almost surely, let us show that it is bounded in L^1. We have

$$\mathbb{E}[L] \leq \underbrace{\mathbb{E}\left[\sup_{1\leq n\leq N}\left|\sum_{i=n}^{N}\frac{1}{i+1}\langle \delta_{X_{i+1}} - \mathbb{E}[\delta_{X_{i+1}}|\mathcal{F}_i], W\rangle\right|\right]}_{(\mathtt{I})} + \underbrace{\sum_{\Lambda\in\{A,B\}} r_\Lambda \sum_{i=1}^{+\infty}\frac{1}{i+1}\mathbb{E}\left[\left|\left\langle \frac{m_i D^\Lambda}{m_i D^\Lambda(E)} - \frac{\eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}, W\right\rangle\right|\right]}_{(\mathtt{II})}.$$

For term (I), using the Jensen Inequality and the Burkholder–Davis–Gundy Inequality, we see that there exists $C_2 > 0$ such that

$$(\mathtt{I}) \leq \sqrt{C_2 \sum_{i=1}^{+\infty}\frac{1}{(i+1)^2}\mathbb{E}\left[(W(X_{i+1}) - \mathbb{E}[W(X_{i+1})|\mathcal{F}_i])^2\right]} \lesssim \sqrt{\sum_{i=1}^{+\infty}\frac{1}{(i+1)^{2-2\delta}}}, \tag{8}$$

where we used that $W(X_{i+1})$ is deterministically smaller, up to constants[3], than i^δ. Hence, taking $\delta < 1/2$ ensures that (I) is bounded. Moving to term (II), we set $\widetilde{m}_i = \frac{m_i}{i}$. Clearly, $\frac{m_i D^\Lambda}{m_i D^\Lambda(E)} = \frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)}$, and by the triangle inequality, we have, for all $i \geq 1$ and $\Lambda \in \{A, B\}$,

$$\mathbb{E}\left[\left|\left\langle \frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)} - \frac{\eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}, W\right\rangle\right|\right] \leq \underbrace{\mathbb{E}\left[\left|\left\langle \frac{\widetilde{m}_i D^\Lambda - \eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}, W\right\rangle\right|\right]}_{(\mathtt{II.a})} + \underbrace{\mathbb{E}\left[\left|\left\langle \frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)} - \frac{\widetilde{m}_i D^\Lambda}{\eta_i Q^\Lambda(E)}, W\right\rangle\right|\right]}_{(\mathtt{II.b})}.$$

For term (II.a), using the facts that $\eta_i Q^\Lambda(E)$ is lower bounded by p_Λ, that $\widetilde{m}_i D^\Lambda - \eta_i Q^\Lambda = (\widetilde{m}_i - \eta_i R)D^\Lambda$, and applying Lemma 3, we get

$$(\mathtt{II.a}) \leq \frac{W(1,\Lambda)}{p_\Lambda}\left(\frac{m_0(1,\Lambda)}{i} + \frac{1}{i}\mathbb{E}\left[\left|\sum_{j=1}^{i}(B_j - p)\right|\right]\right) \lesssim \frac{1}{\sqrt{i}}, \tag{9}$$

where we used that $\mathbb{E}\left[\left|\sum_{j=1}^{i}(B_j - p)\right|\right] \leq \sqrt{p(1-p)i}$. For term (II.b), we write

$$(\mathtt{II.b}) = \mathbb{E}\left[\left|1 - \frac{\widetilde{m}_i D^\Lambda(E)}{\eta_i Q^\Lambda(E)}\right|\left\langle \frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)}, W\right\rangle\right] \lesssim \mathbb{E}\left[\left|\eta_i Q^\Lambda(E) - \widetilde{m}_i D^\Lambda(E)\right|\right] i^\delta,$$

where we used that $\eta_i Q^\Lambda(E)$ is larger than p_Λ and that the probability measure $\frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)}$ is supported on degrees at most $i + d_{\max}$, where $d_{\max}$ is the maximum degree in G_0. Hence, another application of Lemma 3 yields

$$(\mathtt{II.b}) \lesssim \frac{1}{i^{1/2-\delta}}. \tag{10}$$

[3]Here and in the rest of the paper, for two functions f and g from $\mathbb{N}^*$ to $\mathbb{R}_+$, we write $f(n) \lesssim g(n)$ to say that there exists a constant $C \in (0, +\infty)$ such that for all $n \in \mathbb{N}^*$, we have $f(n) \leq Cg(n)$.

Combining (8), (9) and (10), we obtain

$$\mathbb{E}[L] \lesssim \sqrt{\sum_{i=1}^{+\infty} \frac{1}{i^{2-2\delta}} + \sum_{i=1}^{+\infty} \frac{1}{i^{3/2}} + \frac{1}{i^{3/2-\delta}}}\,,$$

which is finite for $\delta < 1/2$, and thus completes the proof. □

2.3. **Limit points of the translation flow are solutions.** The translation flow $\Theta : C^0(\mathbb{R}_+, \mathcal{P}(E)) \times \mathbb{R}_+ \to C^0(\mathbb{R}_+, \mathcal{P}(E))$ is defined by

$$\Theta^t(X)(s) \mapsto X(t+s)\,.$$

Proposition 7. *For all $\gamma \in [0,1)$, every limit point of the sequence $\{\Theta^t(X)\}_{t\geq 0}$ belongs to the set $\{t \mapsto \Phi_t(\eta), \eta \in \mathcal{P}(E)\}$.*

Before proving Proposition 7, let us introduce the same notation as in [4]. For $u \in [\tau_n, \tau_{n+1}[$, we define

$$\overline{X}(u) = \eta_n \quad \text{and} \quad \overline{\varepsilon}(u) = \varepsilon'_{n+1}\,.$$

Proof of Proposition 7. We proceed as in the proof of [4][Proposition 4.1] and decompose $\Theta^t(X)$ as

$$\Theta^t(X) = L_F(\Theta^t(X)) + A_t + B_t\,,$$

where $L_F : C^0(\mathbb{R}_+, \mathbb{R}^E) \to C^0(\mathbb{R}_+, \mathbb{R}^E)$ is given by

$$L_F(X)(s) = X(0) + \int_0^s F(X(u))\mathrm{d}u\,,$$

and where

$$\begin{aligned} A_t(s) &= \int_t^{t+s} \{F(\overline{X}(u)) - F(X(u))\}\mathrm{d}u\,, \\ B_t(s) &= \int_t^{t+s} \overline{\varepsilon}(u)\mathrm{d}u\,. \end{aligned}$$

We will show that[4]

$$\sup_{s\geq 0} \|A_t(s)\|_1 \underset{t\to+\infty}{\longrightarrow} 0 \quad \text{and} \quad \sup_{s\geq 0} \|B_t(s)\|_1 \underset{t\to+\infty}{\longrightarrow} 0\,. \tag{11}$$

Let us first control the term A_t. Note that for some measure $\eta \in \mathcal{P}(E)$, we can write

$$F(\eta) = s(\eta)\eta Q^A + (1 - s(\eta))\eta Q^B - \lambda(\eta)\eta\,,$$

where

$$\lambda(\eta) = \frac{\eta Q^A(E)\eta Q^B(E)}{r_A \eta Q^B(E) + r_B \eta Q^A(E)}\,.$$

With this writing at hand, repeated applications of the triangle inequality gives

$$\begin{aligned} \|F(\eta) - F(\nu)\|_1 \leq\ & s(\eta)\|(\eta-\nu)Q^A\|_1 + (1-s(\eta))\|(\eta-\nu)Q^B\|_1 \\ & + |s(\eta) - s(\nu)|\left(\|\nu Q^A\|_1 + \|\nu Q^B\|_1\right) + \lambda(\eta)\|\eta - \nu\|_1 + |\lambda(\eta) - \lambda(\nu)|\,. \end{aligned}$$

We claim that

$$|s(\eta) - s(\nu)| \leq \frac{2^\gamma(2^\gamma - 1)}{8 p_A p_B}\|\eta - \nu\|_1\,, \tag{12}$$

[4]This is more than enough: it would suffice to have it for the supremum over $0 \leq s \leq T$ for any fixed $T > 0$.

and that

$$|\lambda(\eta) - \lambda(\nu)| \leq \frac{1}{2}\left(1 + \frac{1}{2p_A p_B}\right)\|\eta - \nu\|_1 \,. \tag{13}$$

The proof of those two inequalities can be found in Section 7.1 (for (12), the precise constant factor is not useful here but it will be used later in the paper). Omitting the precise constant factors, we have

$$\|F(\eta) - F(\nu)\|_1 \lesssim \|\eta - \nu\|_1 + \max_{\Lambda\in\{A,B\}} \left\{\|(\eta-\nu)Q^\Lambda\|_1 + \|\eta-\nu\|_1\|\nu Q^\Lambda\|_1\right\} .$$

Choosing $\eta = X(u)$ and $\nu = \overline{X}(u)$ with $u \in [\tau_n,\tau_{n+1}[$ and using that $\eta_{n+1} - \eta_n = \frac{1}{n+1}(\delta_{X_{n+1}} - \eta_n)$, this yields

$$\begin{aligned}\|F(X(u)) - F(\overline{X}(u))\|_1 &\lesssim \|\eta_{n+1} - \eta_n\|_1 \\ &\qquad + \max_{\Lambda\in\{A,B\}} \left\{\|(\eta_{n+1}-\eta_n)Q^\Lambda\|_1 + \|\eta_{n+1}-\eta_n\|_1\|\eta_n Q^\Lambda\|_1\right\} \\ &\lesssim \frac{1}{n+1}\left(1 + \max_{\Lambda\in\{A,B\}} \left\{\|\delta_{X_{n+1}}Q^\Lambda\|_1 + \|\eta_n Q^\Lambda\|_1\right\}\right) .\end{aligned}$$

Now observe that both measures $\delta_{X_{n+1}}$ and η_n are supported on degrees at most $d_{\max} + n$, where $d_{\max}$ is the maximum degree in the initial graph G_0. Hence both $\|\delta_{X_{n+1}}Q^\Lambda\|_1$ and $\|\eta_n Q^\Lambda\|_1$ are smaller (up to constant factors) than $(n+1)^\gamma$. Finally, we obtain that for $t \in [\tau_n,\tau_{n+1}[$,

$$\begin{aligned}\sup_{s\geq 0}\|A_t(s)\|_1 &\leq \int_t^{+\infty} \|F(X(u)) - F(\overline{X}(u))\|_1 \mathrm{d}u \\ &\lesssim \sum_{i\geq n} \gamma_{i+1}(i+1)^{\gamma-1} \lesssim \sum_{i\geq n}(i+1)^{\gamma-2} \lesssim n^{\gamma-1} \underset{n\to+\infty}{\longrightarrow} 0\,,\end{aligned}$$

establishing the first part of (11). Moving on to B_t, if $t \in [\tau_n,\tau_{n+1})$ and $t+s \in [\tau_m,\tau_{m+1})$ with $m \geq n$, then

$$B_t(s) = (\tau_{n+1} - t)\varepsilon'_{n+1} + \sum_{i=n+1}^{m-1} \gamma_{i+1}\varepsilon'_{i+1} + (t+s-\tau_m)\varepsilon'_{m+1}\,.$$

Noticing that $\gamma_{i+1}\varepsilon'_{i+1} = \frac{1}{i+1}\varepsilon_{i+1}$, we have

$$\|B_t(s)\|_1 \leq \frac{1}{n+1}\|\varepsilon_{n+1}\|_1 + \frac{1}{m+1}\|\varepsilon_{m+1}\|_1 + \left\|\sum_{i=n+1}^{m-1}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1 .$$

For the first two terms, we can argue as above:

$$\|\varepsilon_{n+1}\|_1 \lesssim 1 + \max_{\Lambda\in\{A,B\}}\|\eta_n Q^\Lambda\|_1 \lesssim (n+1)^\gamma\,,$$

so that $\frac{1}{n+1}\|\varepsilon_{n+1}\|_1 \to 0$ as $n \to +\infty$. For the last term, we show that

$$\sup_{m\geq n}\left\|\sum_{i=n}^{m}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1 \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{a.s.} \tag{14}$$

The arguments are quite similar to the ones that were used in the proof of Lemma 6 to prove that the quantity L was finite almost surely. The proof can be found in Section 7.2. This yields $\sup_{s\geq 0}\|B_t(s)\|_1 \to 0$ as $t \to +\infty$, establishing (11).

Hence, if $X^\star$ is a limit point of $\{\Theta^t(X)\}_{t\geq 0}$, then it satisfies

$$X^\star = L_F(X^\star)\,,$$

that is, $X^\star$ is solution to (7). □

3. Existence and uniqueness of the fixed point

For $u \in [0,1]$, let $Q_u = uQ^A + (1-u)Q^B$. Then, for $\eta \in \mathcal{P}(E)$, we have $F(\eta) = \eta Q_{s(\eta)} - \eta Q_{s(\eta)}(E)\eta$, with $s(\eta)$ defined in (6). In this section, we prove that there exists a unique probability measure $\eta^\star$ on E such that $F(\eta^\star) = 0$. To this end, we show that for all $u \in [0,1]$, there exists a unique probability measure π_u on E such that π_u is a left eigenvector of Q_u, then it follows from a straightforward calculation that the associated eigenvalue $\lambda(u) > 0$. Furthermore, we observe that if a probability measure η on E satisfies $F(\eta) = 0$, then η is a left eigenvector of $Q_{s(\eta)}$ associated with the eigenvalue $\lambda_\eta := \eta Q_{s(\eta)}(E) > 0$. Since $s(\eta) \in [0,1]$, there exists some u such that $\eta = \pi_u$, and this parameter satisfies $s(\pi_u) = u$. The uniqueness of a solution to the equation $F(\eta) = 0$ follows from the uniqueness of a solution to the equation $s(\pi_u) = u$. We denote this unique solution by $u^\star$ (or $s^\star$ since $s(\pi_{u^\star}) = u^\star$), so that $\eta^\star = \pi_{u^\star}$. Finally, we present an alternative proof of this result based on Schauder's fixed-point theorem, which yields a different, yet still implicit, characterization of the solution $\eta^\star$.

3.1. **Existence and uniqueness of a eigenmeasure of Q_u.** Let us first recall the definition of Q_u. For every $u \in [0,1]$, Q_u is the convex combination of Q^A and Q^B defined as $Q_u := uQ^A + (1-u)Q^B$. More explicitly, Q_u can be written as the sum of the following two infinite matrices:

$$\begin{pmatrix} up & (1-u)(1-p) & 0 & 0 & \cdots \\ up & (1-u)(1-p) & 0 & 0 & \cdots \\ up & (1-u)(1-p) & 0 & 0 & \cdots \\ up & (1-u)(1-p) & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and

$$\begin{array}{c} \\ \\ (1,A)\to \\ (1,B)\to \\ (2,A)\to \\ (2,B)\to \\ \vdots \\ (k-1,A)\to \\ (k-1,B)\to \\ (k,A)\to \\ (k,B)\to \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccccc} (1,A) & (1,B) & (2,A) & (2,B) & \cdots & (k,A) & (k,B) & (k{+}1,A) & (k{+}1,B) & \cdots \end{array} \\ \begin{pmatrix} -u & 0 & u2^\gamma & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & -(1-u) & 0 & (1-u)2^\gamma & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & -u2^\gamma & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & -(1-u)2^\gamma & \cdots & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & 0 & \cdots & uk^\gamma & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & (1-u)k^\gamma & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & -uk^\gamma & 0 & u(k{+}1)^\gamma & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & -(1-u)k^\gamma & 0 & (1-u)(k+1)^\gamma & \cdots \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array}.$$

Lemma 8. *For every $u \in [0,1]$, there exists a unique probability measure on E that is a eigenmeasure of Q_u. We denote this measure by π_u and we have*

$$\pi_u(k,A) = p\prod_{j=1}^{k}\frac{uj^\gamma}{\lambda(u)+uj^\gamma} \quad \text{and} \quad \pi_u(k,B) = (1-p)\prod_{j=1}^{k}\frac{(1-u)j^\gamma}{\lambda(u)+(1-u)j^\gamma},$$

where $\lambda(u)$ denotes the unique real number satisfying

$$p\sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{uj^\gamma}{\lambda(u)+uj^\gamma} + (1-p)\sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{(1-u)j^\gamma}{\lambda(u)+(1-u)j^\gamma} = 1\,.$$

In other words, $(\lambda(u),\pi_u)$ is the unique eigenpair of Q_u such that $\lambda(u)\in\mathbb{R}$ and $\pi_u \in \mathcal{P}(E)$. Moreover, it turns out that $\lambda(u)$ is equal to $\pi_u Q_u(E)$ and is strictly positive.

Proof. Suppose that there exist $\eta \in \mathcal{P}(E)$ and $\lambda \in \mathbb{R}$ such that $\eta Q_u = \lambda\eta$. The equality $\eta Q_u = \lambda\eta$ is equivalent, componentwise, to the following system:

$$\begin{cases} \eta Q_u(1,A) = up - u\eta(1,A) & = \lambda\eta(1,A) \\ \eta Q_u(1,B) = (1-u)(1-p) - (1-u)\eta(1,B) & = \lambda\eta(1,B) \\ \eta Q_u(k,A) = uk^\gamma(\eta(k-1,A) - \eta(k,A)) & = \lambda\eta(k,A), \quad \forall k \geq 2 \\ \eta Q_u(k,B) = (1-u)k^\gamma(\eta(k-1,B) - \eta(k,B)) & = \lambda\eta(k,B), \quad \forall k \geq 2\,. \end{cases}$$

Thus, (λ,η) satisfies the following relations:

$$\begin{cases} \eta(1,A) = p\,\dfrac{u}{\lambda+u} \\ \eta(1,B) = (1-p)\,\dfrac{1-u}{\lambda+1-u} \end{cases}$$

and

$$\forall k \geq 2, \quad \begin{cases} \eta(k,A) = \dfrac{uk^\gamma}{\lambda+uk^\gamma}\,\eta(k-1,A) \\ \eta(k,B) = \dfrac{(1-u)k^\gamma}{\lambda+(1-u)k^\gamma}\,\eta(k-1,B)\,. \end{cases}$$

By induction, we have

$$\forall k \geq 1, \quad \begin{cases} \eta(k,A) = p\displaystyle\prod_{j=1}^{k}\frac{uj^\gamma}{\lambda+uj^\gamma} \\ \eta(k,B) = (1-p)\displaystyle\prod_{j=1}^{k}\frac{(1-u)j^\gamma}{\lambda+(1-u)j^\gamma}\,. \end{cases} \tag{15}$$

Note that, for η to be a probability measure, all the terms appearing in (15) must be positive and sum to 1. It follows that λ must be strictly positive. Let χ be the function defined on $(0,\infty)$ by

$$\chi(x) = p\sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{uj^\gamma}{x+uj^\gamma} + (1-p)\sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{(1-u)j^\gamma}{x+(1-u)j^\gamma}\,.$$

Since η must be a probability measure, such a pair (η,λ) exists if and only if the equation $\chi(x) = 1$ admits a solution $x > 0$. The function χ is clearly strictly decreasing on $(0,\infty)$.

Moreover, by standard arguments, we can show that $\chi : (0,\infty) \to (0,\infty)$ is continuous and thus bijective. Hence, the equation $\chi(x) = 1$ admits a unique solution, which we denote by $\lambda(u)$. Defining π_u as in (15) with $\lambda = \lambda(u)$, we have thus established the existence and uniqueness of a left eigenpair $(\lambda(u), \pi_u) \in \mathbb{R} \times \mathcal{P}(E)$ of Q_u.
The fact that $\pi_u Q_u = \lambda(u)\pi_u$, combined with the fact that π_u is a probability measure, immediately implies that $\lambda(u) = \pi_u Q_u(E)$. □

3.2. **Existence and uniqueness of a solution to $F(\eta) = 0$.**

Proposition 9. *Let F be the function defined on $\mathcal{P}(E)$ by $F(\eta) = \eta Q_{s(\eta)} - \eta Q_{s(\eta)}(E)\eta$. There exists a unique probability measure $\eta \in \mathcal{P}(E)$ such that $F(\eta) = 0$.*

To prove this proposition, we use the following lemma, whose proof is provided at the end of the paper.

Lemma 10. *Let s be the function defined on $\mathcal{P}(E)$ by*

$$s(\eta) = \frac{r_A \eta Q^B(E)}{r_B \eta Q^A(E) + r_A \eta Q^B(E)},$$

where we recall the notation $r_\Lambda = p_\Lambda(1 - q_\Lambda) + p_{\overline{\Lambda}} q_{\overline{\Lambda}}$. For every $u \in [0,1]$, recall that $(\lambda(u),\pi_u)$ is the unique eigenpair in $\mathbb{R} \times \mathcal{P}(E)$ of Q_u . There exists a unique $u \in [0,1]$ such that $s(\pi_u) = u$. We denote by $s^\star$ this unique solution and we set $s^\star_A = s^\star$ and $s^\star_B = 1 - s^\star$.

Proof of Proposition 9. Suppose that there exists a probability measure η on E such that $F(\eta) = 0$. Thus, η is a eigenmeasure of $Q_{s(\eta)}$ associated with the eigenvalue $\lambda_\eta := \eta Q_{s(\eta)}(E)$. Since $u := s(\eta) \in [0,1]$, Lemma 8 ensures that $\eta = \pi_{s(\eta)} = \pi_u$. Moreover, $\pi_{s(\eta)}$ satisfies $\pi_{s(\eta)} Q_{s(\eta)} = \lambda_\eta \pi_{s(\eta)}$. And it follows from this equality that $s(\eta) = s(\pi_{s(\eta)})$, which, recalling the notation $u := s(\eta)$, can be rewritten as $u = s(\pi_u)$. The existence and uniqueness of a solution to this equation, established by Lemma 10, completes the proof. □

Remark 5. *Note that, in the single-community case, the stationary measure $\eta^\star$ is given by, for all $k \in \mathbb{N}^*$,*

$$\eta^\star(k) = \prod_{j=1}^{k} \frac{j^\gamma}{\lambda^\star + j^\gamma},$$

where $\lambda^\star$ is the unique solution λ to the following equation

$$\sum_{k=1}^{\infty} \prod_{j=1}^{k} \frac{j^\gamma}{\lambda + j^\gamma} = 1\,.$$

This parameter $\lambda^\star$ corresponds to the Malthusian parameter defined in [20].

3.3. **Alternative proof.** Let us give an alternative proof of the existence and uniqueness of $\eta^\star \in \mathcal{P}(E)$ such that $F(\eta^\star) = 0$.

Alternative proof of Proposition 9. Consider the operator R defined by: for all $\mu \in \mathcal{P}(E)$,

$$R_\mu = s(\mu)(Q^A - \mu Q^A(E) I) + (1 - s(\mu))(Q^B - \mu Q^B(E) I)\,,$$

and define the application

$$\begin{aligned} \Psi : \quad & \mathcal{P}(E) \longrightarrow \mathcal{P}(E) \\ & \mu \longmapsto \Psi(\mu)\,, \end{aligned}$$

where $\Psi(\mu)$ is the unique probability measure π such that $\pi R_\mu = 0$. One can check that $\Psi(\mu)$ is given by: for all $k \geq 1$ and $\Lambda \in \{A, B\}$,

$$\Psi(\mu)(k, \Lambda) = C^{-1} p_\Lambda \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + \frac{\alpha_\mu^\Lambda}{r_\Lambda}},$$

where

$$\alpha_\mu^\Lambda = \mu Q^\Lambda(E) = p_\Lambda + \sum_{k \geq 1} \mu(k, \Lambda) \left((k+1)^\gamma - k^\gamma\right),$$

and where $C > 0$ is the normalizing constant:

$$C = \sum_{\Lambda \in \{A,B\}} \sum_{k \geq 1} p_\Lambda \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + \frac{\alpha_\mu^\Lambda}{r_\Lambda}}.$$

We observe that Ψ is continuous and that, since α_μ^Λ is always between p_Λ and $p_\Lambda + 2^\gamma - 1$, all the probability measures $\Psi(\mu)$ admit a first moment that is uniformly bounded by some constant $M > 0$. Hence Ψ takes values in a compact convex subset of the Banach space $\ell^1(E)$ and by Schauder's fixed point theorem, it admits a fixed point. In other words, there exists a probability measure $\eta^\star \in \mathcal{P}(E)$ such that $\eta^\star R(\eta^\star) = F(\eta^\star) = 0$.

Let us now show that this measure is unique. A solution $\eta^\star$ satisfies, for all $k \geq 1$, $\Lambda \in \{A, B\}$,

$$\eta^\star(k, \Lambda) = p_\Lambda \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + \frac{\alpha_{\eta^\star}^\Lambda}{r_\Lambda}}.$$

In particular, we have

$$\alpha_{\eta^\star}^\Lambda = p_\Lambda + \sum_{k \geq 1} p_\Lambda \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + \frac{\alpha_{\eta^\star}^\Lambda}{r_\Lambda}} \left((k+1)^\gamma - k^\gamma\right).$$

Since the function $x \mapsto \sum_{k \geq 1} \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + \frac{x}{r_\Lambda}} \left((k+1)^\gamma - k^\gamma\right)$ is strictly decreasing on $\mathbb{R}_+^*$, the values $\alpha_{\eta^\star}^A$ and $\alpha_{\eta^\star}^B$ are uniquely determined, hence $\eta^\star$ is unique. □

Remark 6. *The two proofs yield two expressions for the stationary measure $\eta^\star$, one parametrized by $\lambda^\star$ and $s^\star = s(\eta^\star)$ and the other by $\alpha_{\eta^\star}^\Lambda$, for $\Lambda \in \{A, B\}$. There is, of course, a natural relationship between these three parameters $\lambda^\star$, $s(\eta^\star)$ and $\alpha_{\eta^\star}^\Lambda$. Since we have*

$$\begin{aligned}
\lambda^\star &= \eta^\star Q_{s(\eta^\star)}(E) = s_A^\star \alpha_{\eta^\star}^A + s_B^\star \alpha_{\eta^\star}^B \\
&= \frac{r_A \alpha_{\eta^\star}^B \alpha_{\eta^\star}^A}{r_B \alpha_{\eta^\star}^A + r_A \alpha_{\eta^\star}^B} + \frac{r_B \alpha_{\eta^\star}^A \alpha_{\eta^\star}^B}{r_B \alpha_{\eta^\star}^A + r_A \alpha_{\eta^\star}^B} \\
&= \frac{\alpha_{\eta^\star}^A \alpha_{\eta^\star}^B}{r_B \alpha_{\eta^\star}^A + r_A \alpha_{\eta^\star}^B} \qquad (\textit{since } r_A + r_B = 1),
\end{aligned}$$

it follows that

$$\frac{\lambda^\star}{s_\Lambda^\star} = \frac{\alpha_{\eta^\star}^\Lambda}{r_\Lambda},$$

for $\Lambda \in \{A, B\}$.

4. Convergence in the one-community case

In the case of one single community, the operator F that drives the equation $\partial_t \eta_t = F(\eta_t)$ is given by

$$F(\eta) = \eta Q - \alpha_\eta \eta \,, \tag{16}$$

where Q is given by

$$Q = \begin{pmatrix} 1 & 0 & 0 & \dots & \dots \\ 1 & 0 & 0 & \dots & \dots \\ \vdots & \vdots & \vdots & \dots & \dots \\ 1 & 0 & 0 & \dots & \dots \\ \vdots & \vdots & \vdots & \dots & \dots \end{pmatrix} + \begin{pmatrix} -1^\gamma & 2^\gamma & 0 & \dots & \dots & \\ 0 & -2^\gamma & 3^\gamma & 0 & \dots & \\ 0 & 0 & \ddots & \ddots & & \\ 0 & \dots & & -k^\gamma & (k+1)^\gamma & \\ 0 & \dots & \dots & & \ddots & \ddots \end{pmatrix}$$

and where

$$\alpha_\eta = \eta Q(E) := 1 + \sum_{k \geq 1} \eta_k \left((k+1)^\gamma - k^\gamma \right) .$$

One approach to study Equation (16) is to follow [15] and to view it as the equation describing the law of a conditioned Markov process. Noticing that the sum on each row of the matrix Q is always less that 2^γ, the matrix $Q - 2^\gamma I$ is a sub-Markovian infinitesimal generator: upon adding an additional state ∂ with the appropriate jump rates, it becomes Markovian. The function $F(\eta) = \eta(Q - 2^\gamma I) - \eta(Q - 2^\gamma I)(E)\eta$ is then interpreted as the one that governs the distribution of a Markov process on $E \cup \{\partial\}$ conditioned to stay on E, and one may use results such as [9, 10] to show convergence to a unique *quasi-stationary distribution*. In our setting, the problem of this approach is that it is not clear how to extend it to two communities. Instead, our approach consists in trying to write $F(\eta)$ as $F(\eta) = \eta L_\eta$ with L_η having negative entries on the diagonal, positive entries off the diagonal, and zero sum on each row. Note that $Q - \alpha_\eta I$ is not far from doing the job but the sum on each row does not equal zero. One may rather consider the following small perturbation of $Q - \alpha_\eta I$ and define L_η as

$$L_\eta = \begin{pmatrix} -2^\gamma & 2^\gamma & & & & & & \\ \alpha_\eta + 2^\gamma - 3^\gamma & -(\alpha_\eta + 2^\gamma) & 3^\gamma & & & & & \\ \vdots & \ddots & \ddots & & & (0) & & \\ \vdots & & \ddots & \ddots & & & & \\ \vdots & & & \ddots & \ddots & & & \\ \alpha_\eta + k^\gamma - (k+1)^\gamma & (0) & & & -(\alpha_\eta + k^\gamma) & (k+1)^\gamma & \cdots \\ \vdots & & & & & \ddots & \ddots & \end{pmatrix} .$$

In words, the first column of the matrix L_η is given by the sequence $(\alpha_\eta + k^\gamma - (k+1)^\gamma)$ except for the first entry which is -2^γ, the diagonal is given by the sequence $(-\alpha_\eta - k^\gamma)$ (except for the first entry which is -2^γ), and the top-diagonal is given by the sequence $((k+1)^\gamma)$. Note that, for all $k \geq 2$, we have

$$\alpha_\eta + k^\gamma - (k+1)^\gamma \geq \theta := 1 + 2^\gamma - 3^\gamma > 0 \,.$$

The operator L_η corresponds to the infinitesimal generator of a non-linear Markov jump process[5] on the positive integers which jumps from k to $k+1$ at rate $(k+1)^\gamma$, and from k to 1 at rate $\alpha_\eta + k^\gamma - (k+1)^\gamma$. We observe that only the jump rates associated with returns to state 1 depend on the law, whereas the jump rates from k to $k+1$ depend only on k. For $\gamma \in \{0,1\}$, the non-linearity completely disappears from L_η: whatever the current state $k \geq 2$, the process jumps to 1 at rate 1. Convergence can be easily established in those two extreme cases, *e.g.* by standard coupling arguments for (linear) Markov processes.

4.1. **Construction of the process $(X_t)_{t\geq 0}$.** Let $(\mu_t)_{t\geq 0}$ be a continuous trajectory taking values in $\mathcal{P}(E)$. The equation

$$\partial \eta_t = \eta_t L_{\mu_t} \tag{17}$$

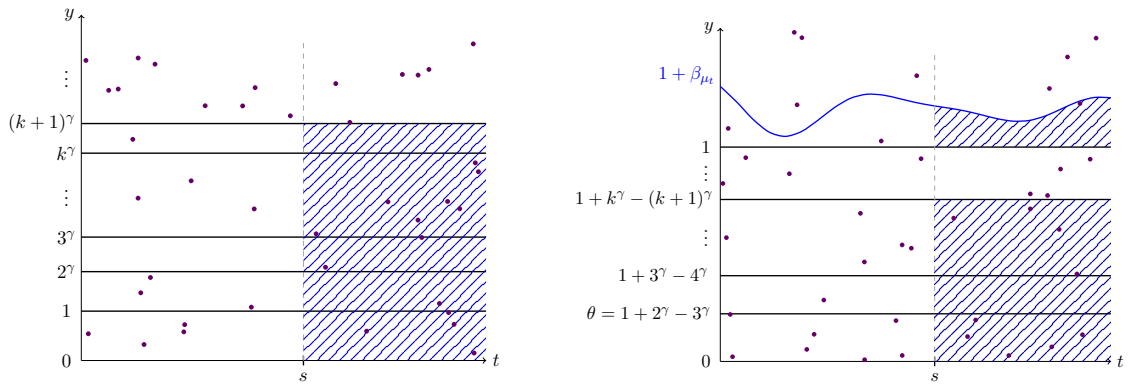
is then the evolution equation for the law of a time-inhomogeneous Markov process whose infinitesimal generator at time t is given by L_{μ_t}. In this section, we construct a process $(X_t)_{t\geq 0}$ whose law $(\eta_t)_{t\geq 0}$ is a solution to Equation (17). To this end, we show that, for every bounded measurable function f_0 on E, the function $f_t := P^\mu_{0,t} f_0$ associated with the semi-group $(P^\mu_{0,t})_{t\geq 0}$ is a solution to $\partial_t f_t = L_{\mu_t} f_t$. Note that we use the same notation L for the action of an operator on functions and on measures, the dual action being distinguished only by whether the operator acts on the left (for functions) or on the right (for measures). Then in Section 4.2, we construct a coupling that allows us to establish, in Section 4.3, the existence and uniqueness of a solution to the equation $\partial_t \eta_t = \eta_t L_{\eta_t}$.

Let us consider two independent Poisson point processes Π_{up} and Π_{down} on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $\mathrm{d}t \times \mathrm{d}y$. The first process governs the upward jumps, while the second governs the returns to 1. Let us first focus on the upward jump rates. As illustrated in Figure 1a, which corresponds to the point cloud Π_{up}, for each $k \in \mathbb{N}$, we draw the half-line defined by the equation $y = k^\gamma$ on $\mathbb{R}_+$. We thus partition the cone $\mathbb{R}_+ \times \mathbb{R}_+$ into horizontal strips of respective heights $1-0$, $2^\gamma - 1$, $3^\gamma - 2^\gamma$, ..., $(k+1)^\gamma - k^\gamma$, *etc.* For the point cloud Π_{down}, as illustrated in Figure 1b, we partition the cone in the same way, by drawing the horizontal lines $y = 1 + k^\gamma - (k+1)^\gamma$, for each $k \in \mathbb{N}^*$, and then the curve $t \mapsto \beta_{\eta_t}$, for $t \in \mathbb{R}_+$, where

$$\beta_\eta := \sum_{k\geq 1} \eta_k \left((k+1)^\gamma - k^\gamma\right) = \alpha_\eta - 1\,. \tag{18}$$

Now, let $\eta_0 \in \mathcal{P}(E)$ be the initial condition for Equation (17). We then define the random variable X_0 to be distributed according to η_0. We define the stopping time τ_1 as the random variable corresponding to the first jump time of the process. More precisely, if X_0 is at state k for some $k \in \mathbb{N}^*$, τ_1 is given by the point of minimal abscissa $t > 0$ among the points of $\Pi = \Pi_{\text{up}} \cup \Pi_{\text{down}}$ that lie, for the point cloud Π_{up}, between the lines $y = 0$ and $y = (k+1)^\gamma$, and, for the point cloud Π_{down}, either between the lines $y = 0$ and $y = 1 + k^\gamma - (k+1)^\gamma$, or between the line $y = 1$ and the curve $y = 1 + \beta_{\eta_t}$. Note that τ_1 is well defined, since these regions have finite area and, almost surely, contain only finitely many points, with no two points having the same abscissa. If the point belongs to Π_{up}, then $X_{\tau_1} = X_0 + 1 = k+1$, whereas if it belongs to Π_{down}, then $X_{\tau_1} = 1$. Moreover, for every $t \in [0, \tau_1)$ — note that τ_1 is strictly positive almost surely —, we have $X_t = X_0 = k$. We proceed in the same way for the second jump, and so on. Figure 1 shows the regions in which to look for the point with minimal abscissa when the process X is in state k at

[5] See [7][Section A.4.3] for a nice introduction to non-linear Markov processes.



(A) Point cloud Π_{up} (B) Point cloud Π_{down}

FIGURE 1. Representation of the two point clouds Π_{up} and Π_{down}. The hatched regions indicate where to look for the point corresponding to the next jump when X is in state k at time s.

time s, for the two point clouds. Thus, by induction, we define a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}^*}$ that is strictly increasing almost surely and hence the process X for all $t \in [0, \sup_{n\geq 0} \tau_n)$ — where we set $\tau_0 = 0$. In fact, this construction allows us to define the process X on the whole of $\mathbb{R}_+$, since X is non-explosive, as established in the following lemma.

Lemma 11. *The process X constructed above is non-explosive, that is,*

$$\tau_\infty := \sup_{n\geq 0} \tau_n = \infty \quad \text{a.s.}$$

Proof. To begin with, observe that the process $(X_t)_{t\geq 0}$ is stochastically dominated by the homogeneous Markov process $(Y_t)_{t\geq 0}$, with $Y_0 := X_0$, that jumps from k to $k+1$ at rate $(k+1)^\gamma$, for each $k \in \mathbb{N}^*$, and from $k \geq 2$ to 1 at rate c_γ, where the constant c_γ is defined by

$$c_\gamma = \begin{cases} 1 + 2^\gamma - 3^\gamma & \text{if } \ 0 \leq \gamma < 1 \\ 1 & \text{if } \gamma = 1\,. \end{cases}$$

Note that this stochastic domination implies that we can construct a coupling $(X_t, Y_t)_{t\geq 0}$ with $X_0 = Y_0$ (we keep the same notation X and Y for the coupled processes) such that whenever X jumps upward, Y does so as well, and whenever Y returns to state 1, so does X. In particular, this coupling ensures that Y is always greater than or equal to X, which is an equivalent characterization of the fact that the law of Y stochastically dominates that of X. For each $n \geq 1$, let τ_n^X and τ_n^Y denote the n-th jump times of the processes X and Y, respectively. The previous coupling ensures that $\tau_n^X \geq \tau_n^Y$ for every $n \geq 1$, almost surely. We therefore have almost surely

$$\sup_{n\geq 1} \tau_n^X \geq \sup_{n\geq 1} \tau_n^Y\,.$$

Thus, it suffices to show that $\sup_{n\geq 1} \tau_n^Y = +\infty$ almost surely to prove that the process X is non-explosive. However, by standard results on continuous-time Markov jump processes on a countable state space, showing that the process Y is recurrent implies that

$\lim_{n\to\infty} \tau_n^Y = +\infty$ almost surely. Moreover, for the model associated with the process Y, it is actually sufficient to show that state 1 is recurrent, since the probability of reaching state $k \geq 2$ starting from state 1 is strictly positive for every $k \geq 2$. Let us now show that state 1 is recurrent for the process Y. Conditional on $Y_n = k$ ($k \geq 2$), the probability that $Y_{n+1} = k+1$ is equal to

$$\frac{(k+1)^\gamma}{(k+1)^\gamma + c_\gamma}\,.$$

To show that state 1 is recurrent, it suffices to consider the process started from state 1 and prove that its first return time to state 1, denoted by T_1^+, is almost surely finite. We have

$$\mathbb{P}_1(T_1^+ = \infty) = \lim_{n\to\infty} \mathbb{P}_1(T_1^+ > n)\,,$$

where we denote by $\mathbb{P}_1$ the probability measure under which Y is a continuous-time Markov jump process with the transition rates specified above and initial condition $Y_0 = 1$ almost surely. Hence, we have

$$\begin{aligned}\mathbb{P}_1(T_1^+ > n) &= \mathbb{P}_1(Y_1 = 2, Y_2 = 3, ..., Y_n = n+1)\\ &= \prod_{k=1}^{n} \frac{(k+1)^\gamma}{(k+1)^\gamma + c_\gamma}\\ &\leq \exp\left(-\sum_{k=1}^{n} \frac{c_\gamma}{(k+1)^\gamma + c_\gamma}\right) \underset{n\to+\infty}{\longrightarrow} 0\,,\end{aligned}$$

as $\gamma \in [0,1]$ and $c_\gamma > 0$. Thus, T_1^+ is finite almost surely, which concludes the proof. □

Finally, it is straightforward to verify that the law of the process X constructed above is a solution to Equation (17).

4.2. **The coupling.** Now, consider two trajectories $(\mu_t^1)_{t\geq 0}$ and $(\mu_t^2)_{t\geq 0}$ taking values in $\mathcal{P}(E)$. In this section, we construct a coupling of two stochastic processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ whose laws at time t are respectively given by η_t^1, where $(\eta_t^1)_{t\geq 0}$ solves

$$\partial \eta_t = \eta_t L_{\mu_t^1}\,, \tag{19}$$

and η_t^2, where $(\eta_t^2)_{t\geq 0}$ solves

$$\partial \eta_t = \eta_t L_{\mu_t^2}\,. \tag{20}$$

Let η_0^1 and η_0^2 be two probability measures on E, corresponding to the initial conditions for Equations (19) and (20), respectively. Let (X_0, Y_0) be an optimal coupling of η_0^1 and η_0^2. As before, let Π_{up} and Π_{down} be two independent Poisson point processes on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $\mathrm{d}t \times \mathrm{d}y$. For both processes X and Y, we use the same Poisson point processes: the first governs the upward jumps, while the second governs the returns to 1. The only difference is that, for the jumps back to state 1, the curve $y = 1 + \beta_{\mu_t^1}$ is used to construct X, whereas the curve $y = 1 + \beta_{\mu_t^2}$ is used to construct Y, as illustrated in Figure 2, with β_η defined in (18). As a consequence, when the two processes are in the same state $k \in \mathbb{N}^*$, they can separate only when the first point appears in the "red region" (see Figure 2 for the definitions of the "red" and "green regions"). We will exploit this key property of the coupling both to establish the existence and uniqueness of a solution to the dynamical system $\partial_t \eta_t = \eta_t L_{\eta_t}$ and to prove that this solution converges to the stationary state as t tends to infinity.

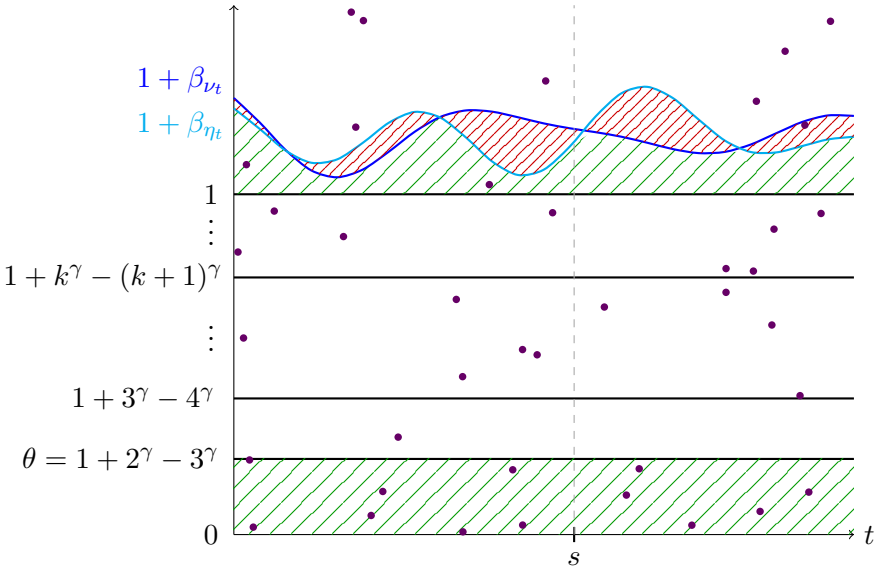


FIGURE 2. Point cloud Π_{down} associated to downward jumps to 1. We refer to the finely hatched region as the "red region" and to the more sparsely hatched region as the "green region".

4.3. **Existence and uniqueness of the solution.** We can now establish the existence and uniqueness result stated earlier by combining the coupling constructed above with a contraction argument.

Proposition 12. *For all measure $\eta_0 \in \mathcal{P}(E)$, there exists a unique solution to $\partial_t \eta_t = \eta_t L_{\eta_t}$.*

Proof. Let $\mu_1 = (\mu_t^1)_{t\geq 0}$ and $\mu_2 = (\mu_t^2)_{t\geq 0}$ be two trajectories of measures in $\mathcal{P}(E)$, and let ν_0 be a measure in $\mathcal{P}(E)$. Consider the process $(X_t, Y_t)_{t\geq 0}$ as defined in Section 4.2. Since $X_0 = Y_0$, we have

$$\|\nu_0 P_{0,t}^{\mu_1} - \nu_0 P_{0,t}^{\mu_2}\|_{\text{TV}} \leq \mathbb{P}(\tau < t)\,,$$

where τ corresponds to the arrival time of the first point in the "red zone" (see Figure 2 for the definition of the "red region"). Indeed, when the two processes are in the same state at time $t = 0$, the coupling is constructed so that they first separate when one process jumps back to state 1 while the other remains unchanged. This corresponds precisely to a jump triggered by a point in the "red region". By setting $\beta_s^i := \beta_{\mu_s^i}$, we have

$$\begin{aligned}\mathbb{P}(\tau < t) &\leq 1 - \exp\left(-\int_0^t |\beta_s^1 - \beta_s^2| \mathrm{d}s\right) \\ &\leq 1 - \exp\left(-(2^\gamma - 1)\int_0^t \|\mu_s^1 - \mu_s^2\|_{\text{TV}} \mathrm{d}s\right)\,,\end{aligned}$$

where we used the characterization of the total-variation distance

$$\|\mu - \nu\|_{\text{TV}} = \sup_{f,\, 0\leq f\leq 1} |\mathbb{E}_\mu f - \mathbb{E}_\nu f|\,,$$

and the fact that $\beta_\mu = \mathbb{E}_\mu f$, with, for all $k \geq 1$, $f(k) = (k+1)^\gamma - k^\gamma \leq 2^\gamma - 1$. Using the inequality $1 - \mathrm{e}^{-x} \leq x$, we obtain

$$\|\nu_0 P_{0,t}^{\mu_1} - \nu_0 P_{0,t}^{\mu_2}\|_{\mathrm{TV}} \leq (2^\gamma - 1) \int_0^t \|\mu_s^1 - \mu_s^2\|_{\mathrm{TV}} \mathrm{d}s \,.$$

Now, setting $t_0 = \frac{1}{2(2^\gamma - 1)}$, we see that

$$\sup_{s \in [0,t_0]} \|\nu_0 P_{0,s}^{\mu_1} - \nu_0 P_{0,s}^{\mu_2}\|_{\mathrm{TV}} \leq \frac{1}{2} \sup_{s \in [0,t_0]} \|\mu_s^1 - \mu_s^2\|_{\mathrm{TV}} \,.$$

In particular, the application $\Psi : (\mu_t)_{t \in [0,t_0]} \mapsto (\nu_0 P_{0,t}^{\mu})_{t \in [0,t_0]}$ is a contraction of the Banach space $L^\infty([0,t_0], \mathcal{P}(E))$. Since this space is complete, the Banach fixed-point theorem ensures that Ψ admits a unique fixed point, which is a solution of $\partial_t \eta_t = \eta_t L_{\eta_t}$ on $[0,t_0]$. Then there exists a unique solution on $[nt_0, (n+1)t_0]$ for all $n \geq 0$, and thus on $\mathbb{R}_+$. □

Remark 7. *Rewriting $F(\eta)$ as ηL_η allows us to interpret the equation $\partial_t \eta_t = F(\eta_t)$ as a nonlinear Fokker–Planck-type evolution equation (on a countable state space), that is, as the deterministic equation governing the evolution of the probability distribution. The construction of the process $(X_t)_{t \geq 0}$ we have presented naturally leads to a McKean–Vlasov-type equation, which describes the evolution of a stochastic process whose dynamics depend on its own probability law, as governed by the associated nonlinear Fokker–Planck equation. The two-part survey by Chaintron and Diez provides a very accessible and engaging introduction to topics related to propagation of chaos and, consequently, to McKean–Vlasov equations (see [7] and [8]).*

4.4. **Convergence of the solution.**

Proposition 13. *Let $\gamma \in [0,1]$ satisfy $\beta^\star > 3^\gamma - 2$ or $\gamma = 1$. Let $(\eta_t)_{t \geq 0}$ be the unique solution to $\partial_t \eta_t = \eta_t L_{\eta_t}$ on $\mathbb{R}_+$, and $\eta^\star$ be unique solution to $\eta L_\eta = 0$. We have*

$$\|\eta_t - \eta^\star\|_{\mathrm{TV}} \underset{t \to \infty}{\longrightarrow} 0 \,.$$

Proof. Let $(X_t, Y_t)_{t \geq 0}$ be a coupling of $(\eta_t)_{t \geq 0}$ and $\eta^\star$, generated as in Section 4.2, with (X_0, Y_0) distributed according to an optimal coupling of η_0 and $\eta^\star$. Let

$$\tau = \inf\{s \geq 0, X_s = Y_s\} \,.$$

This random variable has an atom at 0 since $\mathbb{P}(\tau = 0) = 1 - \|\eta_0 - \eta^\star\|_{\mathrm{TV}}$, and is absolutely continuous with respect to the Lebesgue measure λ on $\mathbb{R}_+^*$. Letting f_τ be the density of τ with respect to $\mu = \delta_0 + \lambda$, we have

$$\mathbb{P}(X_t \neq Y_t) = \mathbb{P}(\forall s \in [0,t], X_s \neq Y_s) + \int_0^t \mathbb{P}(X_t \neq Y_t \,|\, \tau = s) f_\tau(s) \mathrm{d}\mu(s) \,.$$

On the one hand, by setting $\theta := 1 + 2^\gamma - 3^\gamma$, we have

$$\mathbb{P}(\forall s \in [0,t], X_s \neq Y_s) \leq \mathrm{e}^{-\theta t} \|\eta_0 - \eta^\star\|_{\mathrm{TV}} \,,$$

because the presence of a point at time s between the lines $y = 0$ and $y = \theta$ brings both processes back to state 1, so that they coincide at time s. On the other hand, conditionally on $\{\tau = s\}$, if $X_t \neq Y_t$, then there is a point in the "red zone" between s and t which is not followed by a point in the "green zone" (see Figure 2 for the definitions of the "green" and "red regions"). Let τ' be the random variable corresponding to the time of the last point in the "red zone" between s and t if there is one, and to $-\infty$ otherwise. To compute

the density of τ' observe that, for $u \in [s,t]$, denoting by $\mathcal{P}(\lambda)$ a Poisson random variable with parameter λ,

$$\mathbb{P}(\tau' \le u) = \mathbb{P}\left(\mathcal{P}\left(\int_u^t |\beta_x - \beta^\star| \mathrm{d}x\right) = 0\right) = \exp\left(-\int_u^t |\beta_x - \beta^\star| \mathrm{d}x\right),$$

where $\beta_x = \beta_{\eta_x}$. Hence τ' has a density on $[s,t]$ given by, for all $u \in [s,t]$,

$$f_{\tau'}(u) = |\beta_u - \beta^\star| \exp\left(-\int_u^t |\beta_x - \beta^\star| \mathrm{d}x\right).$$

Noticing that the probability that there is no point in the "green zone" between u and t is equal to $\mathrm{e}^{-\int_u^t (\theta + \min\{\beta_x, \beta^\star\}) \mathrm{d}x}$, this yields

$$\begin{aligned}
\mathbb{P}(X_t \neq Y_t \,|\, \tau = s) &\le \int_s^t f_{\tau'}(u)\, \mathrm{e}^{-\int_u^t (\theta + \min\{\beta_x, \beta^\star\}) \mathrm{d}x} \,\mathrm{d}u \\
&\le \int_s^t |\beta_u - \beta^\star|\, \mathrm{e}^{-\int_u^t (\theta + \max\{\beta_x, \beta^\star\}) \mathrm{d}x} \,\mathrm{d}u \\
&\le \int_s^t |\beta_u - \beta^\star|\, \mathrm{e}^{-\int_u^t (\theta + \beta^\star) \mathrm{d}x} \,\mathrm{d}u \\
&\le \int_s^t |\beta_u - \beta^\star|\, \mathrm{e}^{-(\theta+\beta^\star)(t-u)} \,\mathrm{d}u\,.
\end{aligned}$$

Notice that the right-hand side above is maximized when $s = 0$, so that we get

$$\mathbb{P}(X_t \neq Y_t) \le \mathrm{e}^{-\theta t} \|\eta_0 - \eta^\star\|_{\mathrm{TV}} + \int_0^t |\beta_u - \beta^\star|\, \mathrm{e}^{-(\theta+\beta^\star)(t-u)} \,\mathrm{d}u\,.$$

Using that $|\beta_u - \beta^\star| \le (2^\gamma - 1)\|\eta_u - \eta^\star\|_{\mathrm{TV}}$, and setting $g(t) = \mathrm{e}^{(\theta+\beta^\star)t} \|\eta_t - \eta^\star\|_{\mathrm{TV}}$, we have

$$g(t) \le \mathrm{e}^{\beta^\star t}\, g(0) + (2^\gamma - 1) \int_0^t g(u) \mathrm{d}u\,.$$

By Gronwall's lemma, we have

$$\begin{aligned}
g(t) &\le g(0) \left(\mathrm{e}^{\beta^\star t} + (2^\gamma - 1) \int_0^t \mathrm{e}^{\beta^\star u}\, \mathrm{e}^{(2^\gamma - 1)(t-u)} \,\mathrm{d}u\right) \\
&= g(0) \left(\mathrm{e}^{\beta^\star t} + (2^\gamma - 1) \frac{\mathrm{e}^{(2^\gamma - 1)t} - \mathrm{e}^{\beta^\star t}}{2^\gamma - 1 - \beta^\star}\right).
\end{aligned}$$

Therefore,

$$\|\eta_t - \eta^\star\|_{\mathrm{TV}} \le \|\eta_0 - \eta^\star\|_{\mathrm{TV}} \left(\mathrm{e}^{-\theta t} + \frac{2^\gamma - 1}{2^\gamma - 1 - \beta^\star} \left(\mathrm{e}^{-(\beta^\star + 2 - 3^\gamma)t} - \mathrm{e}^{-\theta t}\right)\right). \tag{21}$$

When $\gamma < 1$, then $\theta > 0$. In this case, if $\beta^\star + 2 - 3^\gamma > 0$, then the claimed convergence follows. When $\gamma = 1$, the "red region" in Figure 2 is empty, as $\alpha_\eta = 1 + \beta_\eta = 2$ for every $\eta \in \mathcal{P}(E)$ and thus we have

$$\|\eta_t - \eta^\star\|_{\mathrm{TV}} \le \|\eta_0 - \eta^\star\|_{\mathrm{TV}}\, \mathrm{e}^{-t}\,. \qquad \square$$

Lemma 14. *For all $\gamma \in [0,1]$, the following equivalence holds*

$$\beta^\star > 3^\gamma - 2 \iff \sum_{k \ge 1} \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1} > 1\,.$$

Proof. First, notice that $\eta^\star L_{\eta^\star} = 0$ is equivalent to

$$\begin{cases} \eta^\star(1) = \frac{1}{1+\alpha^\star} \\ \forall k \geq 2,\, \eta^\star(k) = \frac{k^\gamma}{k^\gamma+\alpha^\star}\eta^\star(k-1)\,, \end{cases}$$

where $\alpha^\star = 1 + \beta^\star$. Hence, for all $k \geq 1$, we have

$$\eta^\star(k) = \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 1 + \beta^\star}\,,$$

and $\beta^\star$ can be defined as the solution of

$$\beta = \sum_{k\geq 1}\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 1 + \beta}\left((k+1)^\gamma - k^\gamma\right)\,.$$

Since the function $\beta \mapsto \sum_{k\geq 1}\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma+1+\beta}\left((k+1)^\gamma - k^\gamma\right)$ is decreasing, showing that $\beta^\star > 3^\gamma - 2$ amounts to showing that

$$\sum_{k\geq 1}\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1}\left((k+1)^\gamma - k^\gamma\right) > 3^\gamma - 2\,.$$

After simplifications, this is equivalent to

$$\sum_{k\geq 1}\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1} > 1\,, \tag{22}$$

which concludes the proof. □

The preceding lemma allows us to obtain an equivalent reformulation of Proposition 13. Since proving Inequality (22) is equivalent to showing that $\beta^\star > 3^\gamma - 2$, and since $\beta_\eta > 0$ for all $\eta \in \mathcal{P}(E)$, the inequality $\beta^\star > 3^\gamma - 2$ holds for every $\gamma \in [0,1]$ such that $3^\gamma - 2 \leq 0$, that is, for every $\gamma \in [0, \frac{\ln(2)}{\ln(3)}]$. We have not been able to find an analytical proof of this inequality for all $\gamma \in [0,1]$. However, the graph of the function φ (see Figure 3), together with the fact that its curve lies locally above the line $y = 1$ in a left-hand neighborhood of $\gamma = 1$ (see Remark 8), strongly suggests that the inequality holds throughout $[0,1]$. Nevertheless, we strongly believe that a numerical proof of this inequality can be obtained by using the mean value inequality to construct a sufficiently fine mesh $(x_i)_{1\leq i\leq N}$ of $[0,1]$, consisting of N points, such that the inequality holds at each mesh point x_i. This can be established by numerically approximating the infinite sum by a finite sum and obtaining a suitable bound on the remainder. The conclusion then follows by controlling the variation of φ between consecutive mesh points using the mean value inequality. Thus, this approach also requires establishing an upper bound on $|\varphi'|$, either globally on $[0,1]$ or locally on each subinterval $[x_i, x_{i+1}]$. Since such a proof would provide little insight and be of limited interest, we have chosen not to include pages of calculations and numerical approximations. We therefore leave the search for an elegant proof as an open question.

Remark 8. *Let φ be the function defined on $[0,1]$ by*

$$\varphi(\gamma) = \sum_{k=1}^{\infty}\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1}\,.$$

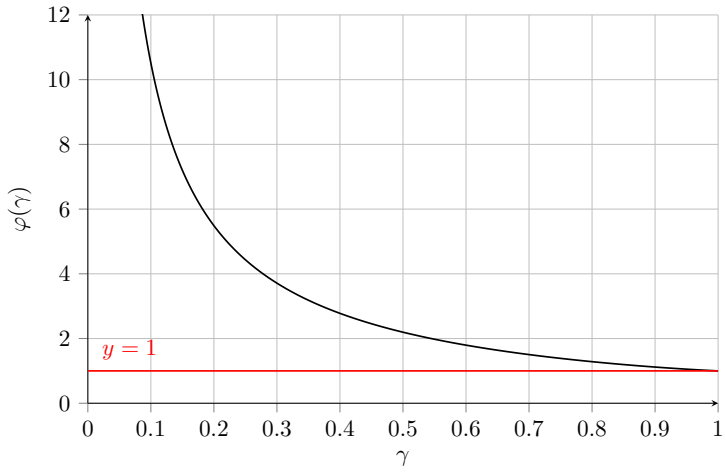


FIGURE 3. Graph of $\varphi(\gamma) = \sum_{k=1}^{+\infty} \prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1}$ on $(0, 1]$.

Note that $\varphi(0) = \infty$ *and* $\varphi(1) = 1$. *Furthermore, a direct computation shows that the derivative of the function* φ *is equal to*

$$\varphi'(\gamma) = \sum_{k=1}^{\infty} \left(\prod_{j=1}^{k} \frac{j^\gamma}{j^\gamma + 3^\gamma - 1} \right) \cdot \sum_{i=1}^{k} \frac{3^\gamma(\ln i - \ln 3) - \ln i}{i^\gamma + 3^\gamma - 1},$$

and that

$$\begin{aligned} \varphi'(1) &= \sum_{k=1}^{\infty} \left(\prod_{j=1}^{k} \frac{j}{j+2} \right) \sum_{i=1}^{k} \frac{2\ln i - 3\ln 3}{i+2} \\ &= \sum_{i=1}^{\infty} \frac{2\ln i - 3\ln 3}{i+2} \sum_{k=i}^{\infty} 2\left(\frac{1}{k+1} - \frac{1}{k+2} \right) \\ &= 4 \sum_{i=1}^{\infty} \frac{\ln i}{(i+1)(i+2)} - 3\ln 3\,. \end{aligned}$$

A comparison between the series and the corresponding integral shows that

$$\sum_{i=1}^{\infty} \frac{\ln i}{(i+1)(i+2)} \leq \int_1^{\infty} \frac{\ln x}{x(x+1)} \mathrm{d}x\,.$$

Finally, standard integration techniques show that the latter integral is equal to $\frac{\pi^2}{12}$. *Hence, we obtain*

$$\varphi'(1) \leq \frac{\pi^2}{3} - 3\ln 3 < -0{,}00597\,.$$

The above numerical approximation allows us to conclude that the derivative of the function φ *at* $\gamma = 1$ *is strictly negative. Consequently, there exists* $\varepsilon > 0$ *such that Inequality* (22) *holds for all* $\gamma \in [1 - \varepsilon, 1]$.

5. Convergence in the two-community case

In the same way as in the one-community, we would like to write the dynamics as the one of a non-linear Markov process. To do so, we use the fact that, for all $\eta \in \mathcal{P}(E)$,

$$\eta \begin{pmatrix} p_A \\ p_A \\ \vdots \\ p_A \\ p_A \\ \vdots \end{pmatrix} = \eta \begin{pmatrix} \alpha_\eta^A - 2^\gamma + 1 \\ \alpha_\eta^A \\ \vdots \\ \alpha_\eta^A - (k+1)^\gamma + k^\gamma \\ \alpha_\eta^A \\ \vdots \end{pmatrix} \quad \text{and} \quad \eta \begin{pmatrix} p_B \\ p_B \\ \vdots \\ p_B \\ p_B \\ \vdots \end{pmatrix} = \eta \begin{pmatrix} \alpha_\eta^B \\ \alpha_\eta^B - 2^\gamma + 1 \\ \vdots \\ \alpha_\eta^B \\ \alpha_\eta^B - (k+1)^\gamma + k^\gamma \\ \vdots \end{pmatrix},$$

where for $\Lambda \in \{A, B\}$, we set

$$\alpha_\eta^\Lambda = \eta Q^\Lambda(E) = p_\Lambda + \sum_{k \geq 1} \eta(k, \Lambda) \left((k+1)^\gamma - k^\gamma \right) .$$

This leads us to define L_η^A as

$$L_\eta^A = \begin{pmatrix} \alpha_\eta^A - 2^\gamma + 1 & 0 & 0 & \dots \\ \alpha_\eta^A & 0 & 0 & \dots \\ \vdots & \vdots & & \\ \alpha_\eta^A - (k+1)^\gamma + k^\gamma & 0 & 0 & \dots \\ \alpha_\eta^A & 0 & 0 & \dots \\ \vdots & \vdots & & \end{pmatrix} + \begin{pmatrix} -1 & 0 & 1 & 0 & \dots & & & & \\ 0 & -1 & 0 & 1 & 0 & \dots & & & \\ \vdots & & \ddots & & \ddots & & & & \\ & & & -1 & 0 & 1 & 0 & \dots & \\ & & & & -1 & 0 & 1 & 0 & \dots \\ & & & & & \ddots & & \ddots & \end{pmatrix} D^A - \alpha_\eta^A I ,$$

and L_η^B as

$$L_\eta^B = \begin{pmatrix} 0 & \alpha_\eta^B & 0 & \dots \\ 0 & \alpha_\eta^B - 2^\gamma + 1 & 0 & \dots \\ \vdots & \vdots & & \\ 0 & \alpha_\eta^B & 0 & \dots \\ 0 & \alpha_\eta^B - (k+1)^\gamma + k^\gamma & 0 & \dots \\ \vdots & \vdots & & \end{pmatrix} + \begin{pmatrix} -1 & 0 & 1 & 0 & \dots & & & & \\ 0 & -1 & 0 & 1 & 0 & \dots & & & \\ \vdots & & \ddots & & \ddots & & & & \\ & & & -1 & 0 & 1 & 0 & \dots & \\ & & & & -1 & 0 & 1 & 0 & \dots \\ & & & & & \ddots & & \ddots & \end{pmatrix} D^B - \alpha_\eta^B I .$$

For $\Lambda \in \{A, B\}$, we have $\eta(Q^\Lambda - \alpha_\eta^\Lambda I) = \eta L_\eta^\Lambda$, and the matrices L_η^Λ have negative entries on the diagonal and zero sum on each row. However, positivity of the coefficients off the diagonal is not guaranteed: we have to ensure that for all $k \geq 2$, $\alpha_\eta^\Lambda - (k+1)^\gamma + k^\gamma \geq 0$, which amounts to $\alpha_\eta^\Lambda - 3^\gamma + 2^\gamma \geq 0$. This is why we now make the following assumption.

Assumption 1. *The parameter $\gamma \in [0, 1]$ is such that*

$$3^\gamma - 2^\gamma \leq p_{min} ,$$

where $p_{min} = \min\{p_A, p_B\}$.

Remark 9. *It might be the case that this assumption is not necessary in the sense that the dynamics could eventually lead to distributions satisfying Assumption 1. We were unable to show such a result. While the idea of rewriting the function $F(\eta)$ in the form ηL_η is appealing, since L_η can be interpreted as an infinitesimal generator depending on η, we*

have nevertheless been unable to establish existence for all values of the parameters p and γ in $[0,1]$.

We have now rewrote our dynamics as

$$\partial_t \eta_t = s(\eta_t)\eta_t L^A_{\eta_t} + (1-s(\eta_t))\eta_t L^B_{\eta_t}\,, \tag{23}$$

which, under Assumption 1, can be seen as a non-linear Markov process.

5.1. **Existence and uniqueness of the solution to the non-linear dynamical system.** The next lemma ensures that (23) has a unique solution for any initial condition.

Proposition 15. *Under Assumption 1, for any measure $\eta_0 \in \mathcal{P}(E)$, there exists a unique solution to (23).*

Proof. Note that for a fixed sequence $\mu = (\mu_t)_{t\geq 0} \in C^0(\mathbb{R}_+, \mathcal{P}(E))$, the dynamics

$$\partial_t \eta_t = \eta_t \left\{ s(\mu_t)(Q^A - \alpha^A_{\mu_t} I) + (1-s(\mu_t))(Q^B - \alpha^B_{\mu_t} I) \right\}$$

do not preserve probability measures (it preserves positivity of the entries, but not the total mass). However, the dynamics

$$\partial_t \eta_t = \eta_t \left\{ s(\mu_t) L^A_{\mu_t} + (1-s(\mu_t)) L^B_{\mu_t} \right\} \tag{24}$$

do preserve probability measures. We denote by $(R^\mu_{s,t})_{t\geq s\geq 0}$ the corresponding inhomogeneous Markov semi-group.

Now let us fix $\nu_0 \in \mathcal{P}(E)$ and define the application

$$\begin{array}{rcl} \Psi : C^0(\mathbb{R}_+, \mathcal{P}(E)) & \longrightarrow & C^0(\mathbb{R}_+, \mathcal{P}(E)) \\ \mu = (\mu_t)_{t\geq 0} & \longmapsto & (\nu_0 R^\mu_{0,t})_{t\geq 0}\,. \end{array}$$

Let us show that Ψ takes values in the compact convex set $K = K(C)$ defined by

$$K = \left\{ (\nu_t)_{t\geq 0} \in C^0(\mathbb{R}_+, \mathcal{P}(E)), \forall t \geq 0, \sum_{x\in E} \nu_t(x)|x| \leq C \right\}\,,$$

where if $x = (k, \Lambda) \in E$, then $|x| = k$ (the *height* of x), and where $C > 0$ will be specified later. In other words, we want to show that, whatever the driving sequence (μ_t), the expected height along the trajectory is uniformly bounded by the same constant. Then Schauder's fixed point theorem will ensure the existence of a fixed point for Ψ, which is a solution to (23). To this end, let $(X^\mu_t)_{t\geq 0}$ be a process whose law is a solution to (24). We then define the height process as the first coordinate of the process $(X^\mu_t)_{t\geq 0}$. In the notation introduced above, the height process is therefore given by $(|X^\mu_t|)_{t\geq 0}$. We observe that the height process is always stochastically dominated by the homogeneous Markov process $(H_t)_{t\geq 0}$ on $\mathbb{N}^*$ which starts from the height measure induced by ν_0, and which, when at $k = 1$, jumps to 2 at rate 2^γ, and when at $k \geq 2$, jumps to $k+1$ at rate $(k+1)^\gamma$ and to 1 at rate

$$\theta_1 = \inf_{\eta\in\mathcal{P}(E)} \min_{\Lambda\in\{A,B\}} \left\{ s^\Lambda(\eta) \left(\alpha^\Lambda_\eta - 3^\gamma + 2^\gamma\right) + s^{\overline{\Lambda}}(\eta)\alpha^{\overline{\Lambda}}_\eta \right\}\,. \tag{25}$$

For this dominating process, we have

$$\partial_t \mathbb{E}[H_t] = \mathbb{E}[(H_t+1)^\gamma] - \theta_1(\mathbb{E}[H_t] - 1) \leq (\mathbb{E}[H_t]+1)^\gamma - \theta_1 \mathbb{E}[H_t] + \theta_1\,,$$

where we used the Jensen inequality. Letting $h^\star$ be the solution of $(h+1)^\gamma - \theta_1 h + \theta_1$, we have, for all $t \geq 0$,

$$\mathbb{E}[H_t] \leq \max\{h^\star, h_0\}\,,$$

where $h_0 = \sum_{x=(k,\Lambda)\in E} \nu_0(x)k$, the initial expected height. □

Remark 10. *Note that the same proof could have been used for the one-community case. We have chosen instead to present two different proofs: one with a more probabilistic flavor for the one-community case, and the other with a more analytical flavor for the two-community case. The former may have the advantage of being less of a black box. However, this perception may simply reflect our own probabilistic inclinations.*

Also note that the two proofs suppose to have written $F(\eta)$ as ηL_η, which, in the two-community case, is done under Assumption 1. One could think of other proofs using e.g. *the Cauchy–Lipschitz Theorem. However, we could not find a norm for which the application F would be Lipschitz and were therefore unable to apply such theorems.*

5.2. **Construction of a process $(X_t)_{t\geq 0}$ whose law solves the dynamical system.** Having established the existence and uniqueness of a solution $(\eta_t)_{t\geq 0}$ to Equation (23), we next construct a process $(X_t)_{t\geq 0}$ such that the law of X_t is equal to η_t, for all $t \in \mathbb{R}_+$, following the approach used in the one-community case. It is worth noting that the time-dependent infinitesimal generator L_{η_t} can be interpreted as a convex combination of the two infinitesimal generators $L^A_{\eta_t}$ and $L^B_{\eta_t}$, with weights that depend nonlinearly on η_t. Equivalently, it can be viewed as the superposition of two dynamics evolving at different rates, namely $s(\eta_t)$ and $1 - s(\eta_t)$, respectively. Alternatively, by analogy with a discrete-time evolution, consider the dynamics over a short time interval $[t, t+\varepsilon]$. One can interpret this evolution as follows: at time t, a generator is randomly selected by a coin toss and then governs the dynamics throughout the interval $[t, t+\varepsilon]$. The generator $L^A_{\eta_t}$ is selected with probability $s(\eta_t)$, while $L^B_{\eta_t}$ is selected with probability $1 - s(\eta_t)$. Let Π^A_{up}, Π^A_{down}, Π^B_{up} and Π^B_{down} be four Poisson point processes on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $\mathrm{d}t \times \mathrm{d}y$. For $\Lambda \in \{A, B\}$, Π^Λ_{up} governs the upward jumps associated with the infinitesimal generator L^Λ, while $\Pi^\Lambda_{\text{down}}$ governs the returns to state $(1, \Lambda)$.

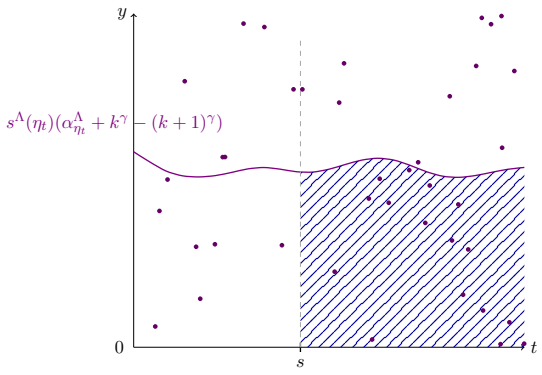
Now, let $\eta_0 \in \mathcal{P}(E)$ be the initial condition for Equation (23). We then define the random variable X_0 to be distributed according to η_0. We define the stopping time τ_1 below as the random variable corresponding to the first jump time of the process. A crucial observation is that the form of the infinitesimal generator L_η implies that any process whose law solves Equation (23) can only jump from state (k, Λ) to one of the states $(k+1, \Lambda)$, $(1, A)$, or $(1, B)$. If $X_0 = (k, \Lambda)$, with $k \geq 2$, then τ_1 is the point with the smallest abscissa $t > 0$ among the points of $\Pi^\Lambda_{\text{up}} \cup \Pi^\Lambda_{\text{down}} \cup \Pi^{\overline{\Lambda}}_{\text{down}}$ that lie:

- **in the case of Π^Λ_{up}:** between the curves $y = 0$ and $y = s^\Lambda(\eta_t)(k+1)^\gamma$;
- **in the case of $\Pi^\Lambda_{\text{down}}$:** between the curves $y = 0$ and $y = s^\Lambda(\eta_t)(\alpha^\Lambda_{\eta_t} + k^\gamma - (k+1)^\gamma)$;
- **in the case of $\Pi^{\overline{\Lambda}}_{\text{down}}$:** between the curves $y = 0$ and $y = s^{\overline{\Lambda}}(\eta_t)\alpha^{\overline{\Lambda}}_{\eta_t}$.

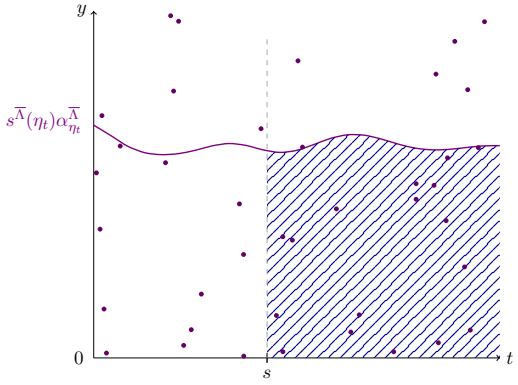
If $X_0 = (1, \Lambda)$, then one only looks at the point with the smallest abscissa $t > 0$ among the points of Π^Λ_{up} that lie between the curves $y = 0$ and $y = s^\Lambda(\eta_t)2^\gamma$ and those of $\Pi^{\overline{\Lambda}}_{\text{down}}$ that lie between the curves $y = 0$ and $y = s^{\overline{\Lambda}}(\eta_t)\alpha^{\overline{\Lambda}}_{\eta_t}$.

Note that, as in the one-community case, τ_1 is well defined, since these regions have finite area and, almost surely, contain only finitely many points, with no two points having

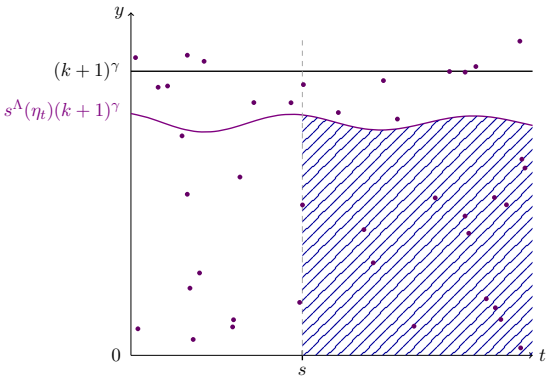
the same abscissa. If the point belongs to $\Pi^{\Lambda}_{\text{up}}$, then $X_{\tau_1} = (k+1, \Lambda)$, whereas if it belongs to Π^{A}_{down}, then $X_{\tau_1} = (1, A)$ and if it belongs to Π^{B}_{down}, then $X_{\tau_1} = (1, B)$. Moreover, for every $t \in [0, \tau_1)$, we have $X_t = X_0 = (k, \Lambda)$. We proceed in the same way for the second jump, and so on. Figure 4 gives a representation of the regions in which to look for the point with minimal abscissa when the process X is in state (k, Λ) (with $k \geq 2$) at time s ($\Pi^{\Lambda}_{\text{down}}$, $\Pi^{\overline{\Lambda}}_{\text{down}}$, and $\Pi^{\Lambda}_{\text{up}}$ are depicted in Figures 4a, 4b , and 4c, respectively). By induction, we define a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}^*}$ that is, almost surely, strictly increasing and goes to infinity. The fact that $\sup_{n\geq 0} \tau_n = \infty$ almost surely follows from the fact that $\gamma \in [0, 1]$. The jump times and the sequence of states visited thus define a process $(X_t)_{t\geq 0}$ whose law indeed solves Equation (23).



(A) Point cloud $\Pi^{\Lambda}_{\text{down}}$



(B) Point cloud $\Pi^{\overline{\Lambda}}_{\text{down}}$



(C) Point cloud $\Pi^{\Lambda}_{\text{up}}$

FIGURE 4. Representation of the three point clouds $\Pi^{\Lambda}_{\text{down}}$, $\Pi^{\overline{\Lambda}}_{\text{down}}$ and $\Pi^{\Lambda}_{\text{up}}$. The hatched regions indicate where to look for the point corresponding to the next jump when $X = (k, \Lambda)$ (when $k \geq 2$) at time s.

5.3. **The coupling.** In this section, we construct a coupling of two stochastic processes $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ whose respective laws at time t are η^1_t and η^2_t, two solutions to Equation (23). Let η^1_0 and η^2_0 be two probability measures on E, corresponding to the initial conditions for this evolution equation. Let (X_0, Y_0) be an optimal coupling of η^1_0 and η^2_0. As before, let Π^{A}_{up}, Π^{B}_{up}, Π^{A}_{down} and Π^{B}_{down} be four independent Poisson point

processes on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity $\mathrm{d}t \times \mathrm{d}y$. For both processes X and Y, we use the same four Poisson point processes: the first two govern the upward jumps along the respective branches $\mathbb{N}^* \times A$ and $\mathbb{N}^* \times B$ of the graph E, while the last two govern the returns to $(1, A)$ and $(1, B)$, respectively. The only difference is that, for each of the four point clouds, the curves drawn on the cone $\mathbb{R}_+ \times \mathbb{R}_+$ depend on the trajectory $(\eta_t^1)_{t\geq 0}$ in the case of X and on $(\eta_t^2)_{t\geq 0}$ in the case of Y.

Remark 11. *Unlike in the one-community case, if the two processes are in the same state (k, Λ), with $\Lambda \in \{A, B\}$, at a given time t, they can separate either because one jumps to one of the states $(1, A)$ or $(1, B)$ while the other remains unchanged, or because one jumps to state $(k + 1, \Lambda)$ while the other does not. In the proof of Proposition 13, we made essential use of the fact that the two processes in the constructed coupling could separate only when one of them jumped back to state 1. In the two-community case, however, the upward jump rates also depend nonlinearly on the law η_t at each time t, which prevents us from directly adapting the arguments developed in Section 4.4.*

5.4. **Convergence of the solution.**

Proposition 16. *Let $(\eta_t)_{t\geq 0}$ be the unique solution to (23) on $\mathbb{R}_+$, and $\eta^\star$ be the unique stationary distribution. There exists $0 < \gamma_0 \leq 1$ such that for $\gamma \in [0, \gamma_0]$, we have*

$$\|\eta_t - \eta^\star\|_{\mathrm{TV}} \underset{t\to\infty}{\longrightarrow} 0\,.$$

Before proving Proposition 16, let us state a useful lemma.

Lemma 17. *For $Y \sim \eta^\star$, provided $\gamma \leq \frac{1}{2}$, we have $\mathbb{E}[(|Y| + 1)^\gamma] \leq 2^\gamma + \frac{\gamma}{\theta_1}$, with θ_1 defined in (25).*

Proof. Let $(Y_t)_{t\geq 0}$ be a solution of (23) started at $Y_0 \sim \eta^\star$. As in the proof of Proposition 15, we start by observing that the height process $(|Y_t|)_{t\geq 0}$ is always stochastically dominated by the homogeneous Markov process $(H_t)_{t\geq 0}$ on $\mathbb{N}^*$ which starts from the height measure induced by $\eta^\star$, and which, when at $k = 1$, jumps to 2 at rate 2^γ, and when at $k \geq 2$, jumps to $k + 1$ at rate $(k + 1)^\gamma$ and to 1 at rate θ_1. For this dominating process, we have

$$\partial_t \mathbb{E}[(H_t + 1)^\gamma] = \mathbb{E}\left[((H_t + 2)^\gamma - (H_t + 1)^\gamma)\,(H_t + 1)^\gamma\right] - \theta_1\left(\mathbb{E}[(H_t + 1)^\gamma] - 2^\gamma\right).$$

Using $(H_t + 2)^\gamma - (H_t + 1)^\gamma \leq \gamma(H_t + 1)^{\gamma-1}$, we have

$$\begin{aligned}\partial_t \mathbb{E}[(H_t + 1)^\gamma] &\leq \gamma\mathbb{E}\left[(H_t + 1)^{2\gamma-1}\right] - \theta_1\left(\mathbb{E}[(H_t + 1)^\gamma] - 2^\gamma\right)\\ &\leq -\theta_1\mathbb{E}[(H_t + 1)^\gamma] + 2^\gamma\theta_1 + \gamma\,,\end{aligned}$$

where for the last inequality, we used that $\gamma \leq \frac{1}{2}$. Hence $\mathbb{E}[(H_t + 1)^\gamma]$ is upper-bounded by a function which converges to $2^\gamma + \frac{\gamma}{\theta_1}$. And since, for all $t \geq 0$, $\mathbb{E}[(|Y| + 1)^\gamma] \leq \mathbb{E}[(H_t + 1)^\gamma]$, we obtain the desired result. □

Proof of Proposition 16. Let $(X_t, Y_t)_{t\geq 0}$ be a coupling of $(\eta_t)_{t\geq 0}$ and $\eta^\star$, generated as in Section 5.3, with (X_0, Y_0) an arbitrary coupling of η_0 and $\eta^\star$ (*e.g.* the independent coupling). We will show that for all $t > 0$,

$$\partial_t \mathbb{P}(X_t \neq Y_t) \leq -a\mathbb{P}(X_t \neq Y_t)\,,$$

with $a > 0$ provided γ is small enough. This will imply that $\mathbb{P}(X_t \neq Y_t) \underset{t\to\infty}{\longrightarrow} 0$, and hence that $\|\eta_t - \eta^\star\|_{\mathrm{TV}} \underset{t\to\infty}{\longrightarrow} 0$.

In what follows, we use the notation

$$\beta_\eta^\Lambda = \sum_{k\geq 1} \eta(k,\Lambda)\left((k+1)^\gamma - k^\gamma\right) = \alpha_\eta^\Lambda - p_\Lambda\,.$$

One can already notice that, since the total-variation distance between two probability measures $\eta,\nu\in\mathcal{P}(E)$ is the supremum, over all functions f on E taking values between 0 and 1, of the absolute difference $|\mathbb{E}_\eta[f]-\mathbb{E}_\nu[f]|$, and since the function $k\mapsto(k+1)^\gamma-k^\gamma$ is between 0 and $2^\gamma-1$, we have

$$|\beta_\eta^\Lambda - \beta_\nu^\Lambda| \leq (2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}\,. \tag{26}$$

Now, in order to control $\mathbb{P}(X_{t+\varepsilon}\neq Y_{t+\varepsilon}\,|\,\mathcal{F}_t)$, we observe the following:

- when $X_t\neq Y_t$, then the two processes get together (either at $(1,A)$ or at $(1,B)$) at rate at least
$$\theta := \sum_{\Lambda\in\{A,B\}} \inf_{\eta\in\mathcal{P}(E)} s^\Lambda(\eta)\left(\alpha_\eta^\Lambda - 3^\gamma + 2^\gamma\right);$$
- when $X_t=Y_t=(1,\Lambda)$ with $\Lambda\in\{A,B\}$, then $X_{t+\varepsilon}\neq Y_{t+\varepsilon}$ when one of the two processes jumps upward to $(2,\Lambda)$ but not the other, which happens at rate $|s(\eta_t)-s(\eta^\star)|2^\gamma$;
- when $X_t=Y_t=(k,A)$ with $k\geq 2$, there are three main contributions to separation at time $t+\varepsilon$:
  - either one jumps upward to $(k+1,A)$ but not the other, which happens at rate $|s(\eta_t)-s(\eta^\star)|(k+1)^\gamma$;
  - or one jumps downward to $(1,A)$ but not the other, which happens at rate
$$\begin{aligned}&\left|s(\eta_t)(p_A+\beta_{\eta_t}^A-(k+1)^\gamma+k^\gamma) - s(\eta^\star)(p_A+\beta_{\eta^\star}^A-(k+1)^\gamma+k^\gamma)\right|\\ &\qquad\leq s(\eta_t)|\beta_{\eta_t}^A-\beta_{\eta^\star}^A| + |s(\eta_t)-s(\eta^\star)|(p_A+\beta_{\eta^\star}^A)\,;\end{aligned}$$
  - or one jumps downward to $(1,B)$ but not the other, which happens at rate
$$\begin{aligned}&\left|(1-s(\eta_t))(p_B+\beta_{\eta_t}^B) - (1-s(\eta^\star))(p_B+\beta_{\eta^\star}^B)\right|\\ &\qquad\leq (1-s(\eta_t))|\beta_{\eta_t}^B-\beta_{\eta^\star}^B| + |s(\eta_t)-s(\eta^\star)|(p_B+\beta_{\eta^\star}^B)\,.\end{aligned}$$

  Summing those three possibilities, using the upper-bound (26), and the fact that $\beta_{\eta^\star}^A+\beta_{\eta^\star}^B\leq 2^\gamma-1$, we obtain a separation rate smaller than
$$|s(\eta_t)-s(\eta^\star)|\left((k+1)^\gamma+2^\gamma\right) + (2^\gamma-1)\|\eta_t-\eta^\star\|_{\mathrm{TV}}\,. \tag{27}$$
- when $X_t=Y_t=(k,B)$ with $k\geq 2$, one just has to adjust the preceding case but the same upper bound (27) holds for the separation rate.

All in all, we get that when the two processes are different at time t, they get together at rate at least θ, whereas when they are together at $x\in E$, denoting by $|x|$ the height of x, they separate at rate at most

$$r_{|x|}(t) := (2^\gamma-1)\left(1+\frac{2^\gamma\left((|x|+1)^\gamma+2^\gamma\right)}{4p_Ap_B}\right)\|\eta_t-\eta^\star\|_{\mathrm{TV}}\,,$$

where we used the upper-bound (12) on $|s(\eta_t)-s(\eta^\star)|$. For $\varepsilon>0$, we then see that the probability that they are together at time $t+\varepsilon$ given that they are separated at time t is larger than the probability that an exponential clock with parameter θ rings between t and $t+\varepsilon$ (so that the two processes are together at level 1), but another independent clock

with parameter 2^γ (a crude upper-bound for the rate at which one of the two processes jumps from level 1 to level 2) does not ring in that same time interval:

$$\mathbb{P}(X_{t+\varepsilon} = Y_{t+\varepsilon} \,|\, X_t \neq Y_t) \geq (1 - \mathrm{e}^{-\theta\varepsilon})\,\mathrm{e}^{-2^\gamma \varepsilon} = \theta\varepsilon + o(\varepsilon)\,,$$

as $\varepsilon \to 0$. On the other hand, the probability that they are separated at time $t+\varepsilon$ given that they are together at x at time t is smaller than the probability that there exists a point in a zone with area $\int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s$:

$$\mathbb{P}(X_{t+\varepsilon} \neq Y_{t+\varepsilon} \,|\, X_t = Y_t = x) \leq 1 - \mathrm{e}^{-\int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s} \leq \int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s\,.$$

We thus have

$$\mathbb{P}(X_{t+\varepsilon} \neq Y_{t+\varepsilon}) \leq \mathbb{P}(X_t \neq Y_t)(1 - \theta\varepsilon + o(\varepsilon)) + \sum_{x\in E} \mathbb{P}(X_t = Y_t = x) \int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s\,.$$

Using that $\mathbb{P}(X_t = Y_t = x) \leq \mathbb{P}(Y_t = x) = \eta^\star(x)$, we get

$$\frac{\mathbb{P}(X_{t+\varepsilon} \neq Y_{t+\varepsilon}) - \mathbb{P}(X_t \neq Y_t)}{\varepsilon} \leq -(\theta + o(1))\mathbb{P}(X_t \neq Y_t) + \sum_{x\in E} \eta^\star(x) \frac{\int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s}{\varepsilon}\,.$$

For $x \in E$ fixed, we have $\varepsilon^{-1}\int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s \to r_{|x|}(t)$ as $\varepsilon \to 0$. Moreover,

$$\varepsilon^{-1}\int_t^{t+\varepsilon} r_{|x|}(s)\mathrm{d}s \leq C(|x|+1)^\gamma\,,$$

for some constant $C > 0$, and Lemma 17 ensures that, if $\gamma \leq \frac{1}{2}$, then $\sum_{x\in E} \eta^\star(x)(|x|+1)^\gamma < +\infty$. Hence, by dominating convergence,

$$\partial_t \mathbb{P}(X_t \neq Y_t) \leq -\theta\mathbb{P}(X_t \neq Y_t) + (2^\gamma - 1)\left(1 + \frac{2^\gamma\left(\mathbb{E}[(|Y|+1)^\gamma] + 2^\gamma\right)}{4p_A p_B}\right)\|\eta_t - \eta^\star\|_{\mathrm{TV}}\,,$$

where $Y \sim \eta^\star$. Finally, using Lemma 17 and the fact that $\|\eta_t - \eta^\star\|_{\mathrm{TV}} \leq \mathbb{P}(X_t \neq Y_t)$, we obtain

$$\partial_t \mathbb{P}(X_t \neq Y_t) \leq -\left(\theta - (2^\gamma - 1)\left(1 + \frac{2^\gamma\left(2^{\gamma+1} + \frac{\gamma}{\theta_1}\right)}{4p_A p_B}\right)\right)\mathbb{P}(X_t \neq Y_t)\,,$$

and we see that we can find γ small enough so that

$$\theta - (2^\gamma - 1)\left(1 + \frac{2^\gamma\left(2^{\gamma+1} + \frac{\gamma}{\theta_1}\right)}{4p_A p_B}\right) > 0\,. \tag{28}$$

For instance, for $p_A = p_B = \frac{1}{2}$ and $q_A = q_B = q$, taking $\gamma \leq \frac{1}{7}$ is sufficient. Indeed, in this case, we have $r_A = r_B$, and the dynamics is completely symmetric between A and B. In particular, $s^A(\eta) = s^B(\eta)$ all along the trajectory. Using that $\alpha_\eta^A + \alpha_\eta^B \geq 1$, we then have

$$\theta \geq \frac{1}{2} - 3^\gamma + 2^\gamma \quad \text{and} \quad \theta_1 \geq \frac{1}{2}\left(1 - 3^\gamma + 2^\gamma\right)\,.$$

One may then check that inequality (28) holds for $\gamma = 1/7$. □

5.5. **Proof of Theorem 1 and 2.** In Proposition 4, we have shown that $(X(t))_{t\geq 0}$ is an asymptotic pseudo-trajectory of the semiflow induced by the well-posed dynamical system (7). In particular, the set of limit points of $(X(t))_{t\geq 0}$ is included in the uniformly attracting set $\{\eta^\star\}$ of the semiflow generated by (7) (see Proposition 13 for one community, Proposition 16 for two communities). In particular, the only limit point of the compact sequence $(\eta_n)_{n\geq 0}$ is $\eta^\star$. For one community, we immediately have

$$\frac{m_n}{n} = \frac{m_0}{n} + \eta_n R \underset{n\to+\infty}{\longrightarrow} \eta^\star R \quad \text{a.s.}$$

For two communities, by Lemma 3, we have

$$\frac{m_n}{n} = \frac{m_0}{n} + \eta_n R + \frac{1}{n}\sum_{i=1}^{n} (B_i - p, p - B_i, 0, \dots) \underset{n\to+\infty}{\longrightarrow} \eta^\star R \quad \text{a.s.},$$

by the law of large numbers.

5.6. **Expression of the limiting distribution $\mu^\star$.** We adopt the notation of Lemmas 8 and 10 and recall the expression for the limit $\eta^\star$ obtained in Proposition 16. We have just shown in Section 5.5 that the limit $\mu^\star$ in Theorem 1 is given by $\mu^\star = \eta^\star R$, which yields

$$\mu^\star(1,\Lambda) = p_\Lambda - \eta^*(1,\Lambda) = p_\Lambda \frac{\lambda^\star}{\lambda^\star + s^\star_\Lambda},$$

and for all $k \geq 2$,

$$\begin{aligned}\mu^\star(k,\Lambda) &= \eta^\star(k-1,\Lambda) - \eta^\star(k,\Lambda) \\ &= p_\Lambda \frac{\lambda^\star}{\lambda^\star + s^\star_\Lambda k^\gamma} \prod_{j=1}^{k-1} \frac{s^\star_\Lambda j^\gamma}{\lambda^\star + s^\star_\Lambda j^\gamma} = p_\Lambda \frac{\alpha^\Lambda_{\eta^\star}/r_\Lambda}{\alpha^\Lambda_{\eta^\star}/r_\Lambda + k^\gamma} \prod_{j=1}^{k-1} \frac{j^\gamma}{\alpha^\Lambda_{\eta^\star}/r_\Lambda + j^\gamma},\end{aligned}$$

where the last equality follows from Remark 6. In the single-community case, the limiting measure can be expressed more concisely as follows:

$$\mu^\star(k) = \frac{\lambda^\star}{\lambda^\star + k^\gamma} \prod_{j=1}^{k-1} \frac{j^\gamma}{\lambda^\star + j^\gamma} = \frac{\alpha^\star}{\alpha^\star + k^\gamma} \prod_{j=1}^{k-1} \frac{j^\gamma}{\alpha^\star + j^\gamma},$$

for all $k \in \mathbb{N}^*$, since $\lambda^\star = \alpha_{\eta^\star} =: \alpha^\star$ in the one-community case.

Remark 12. *In both the single- and two-community settings, the limiting distribution has a power-law tail if and only if $\gamma = 1$. Otherwise, it is stretched-exponential. This behavior is therefore specific to the case $\gamma = 1$. In particular, $\mu^\star(k,\Lambda) \propto k^{-3}$.*

6. Model 2: Proof strategy and main results

In this section, we give all the elements needed to adapt the proof to Model 2 in the case $q_A = q_B = q$. In this case, we still have

$$\eta_{n+1} = \left(1 - \frac{1}{n+1}\right) + \frac{1}{n+1}\delta_{X_{n+1}},$$

but now we have

$$\mathbb{E}[\delta_{X_{n+1}} \,|\, \mathcal{F}_n] = p_A \frac{m_n P^A}{m_n P^A(E)} + p_B \frac{m_n P^B}{m_n P^B(E)},$$

where, for $\Lambda \in \{A, B\}$,

$$P^\Lambda = (1-q)D^\Lambda + qD^{\overline{\Lambda}}.$$

Setting $\widetilde{Q}^{\Lambda} = RP^{\Lambda}$, and recalling that $\frac{m_n}{n} \approx \eta_n R$, we are then incited to write

$$\eta_{n+1} = \eta_n + \frac{1}{n+1}\left(p_A \frac{\eta_n \widetilde{Q}^A}{\eta_n \widetilde{Q}^A(E)} + p_B \frac{\eta_n \widetilde{Q}^B}{\eta_n \widetilde{Q}^B(E)} - \eta_n + \widetilde{\varepsilon}_{n+1} \right), \tag{29}$$

where $\widetilde{\varepsilon}_{n+1} = \delta_{X_{n+1}} - p_A \frac{\eta_n \widetilde{Q}^A}{\eta_n \widetilde{Q}^A(E)} + p_B \frac{\eta_n \widetilde{Q}^B}{\eta_n \widetilde{Q}^B(E)}$. As for Model 1, it is convenient to slightly change the step sizes and rewrite (29) as

$$\eta_{n+1} = \eta_n + \widetilde{\gamma}_{n+1} \left\{ G(\eta_n) + \widetilde{\varepsilon}'_{n+1} \right\},$$

where

$$\begin{aligned} \widetilde{\gamma}_{n+1} &= \frac{1}{n+1}\left(\frac{p_A}{\eta_n \widetilde{Q}^A(E)} + \frac{p_B}{\eta_n \widetilde{Q}^B(E)} \right), \\ \widetilde{\varepsilon}'_{n+1} &= \frac{\widetilde{\varepsilon}_{n+1}}{\frac{p_A}{\eta_n \widetilde{Q}^A(E)} + \frac{p_B}{\eta_n \widetilde{Q}^B(E)}}, \end{aligned}$$

and where

$$G(\eta) = \widetilde{s}(\eta)\left(\eta \widetilde{Q}^A - \widetilde{\alpha}^A_\eta \eta \right) + (1 - \widetilde{s}(\eta))\left(\eta \widetilde{Q}^B - \widetilde{\alpha}^B_\eta \eta \right), \tag{30}$$

with $\widetilde{\alpha}^{\Lambda}_\eta = \eta \widetilde{Q}^{\Lambda}(E)$ and $\widetilde{s}(\eta) = \frac{p_A \widetilde{\alpha}^B_\eta}{p_A \widetilde{\alpha}^B_\eta + p_B \widetilde{\alpha}^A_\eta}$. Note that, since $\widetilde{Q}^{\Lambda} = (1-q)Q^{\Lambda} + qQ^{\overline{\Lambda}}$, the function G defined in (30) has the same form as F defined in (5). More precisely, we have

$$\begin{aligned} G(\eta) = &\left((1-q)\widetilde{s}(\eta) + q(1 - \widetilde{s}(\eta)) \right)\left(\eta Q^A - \alpha^A_\eta \eta \right) \\ &+ \left((q\widetilde{s}(\eta) + (1-q)(1 - \widetilde{s}(\eta)) \right)\left(\eta Q^B - \alpha^B_\eta \eta \right), \end{aligned}$$

so that one can go from F to G by merely replacing $s(\eta)$ by $(1-q)\widetilde{s}(\eta) + q(1-\widetilde{s}(\eta))$.

In particular, considering the increasing sequence $(\widetilde{\tau}_n)_{n\geq 0}$ defined by $\widetilde{\tau}_0 = 0$ and for all $n \geq 1$,

$$\widetilde{\tau}_n = \widetilde{\gamma}_1 + \widetilde{\gamma}_2 + \ldots + \widetilde{\gamma}_n,$$

and introducing the linearized version $(\widetilde{X}(t))_{t\geq 0}$ of $(\eta_n)_{n\in\mathbb{N}}$ defined, for all $n \geq 1$ and all $t \in [\widetilde{\tau}_n, \widetilde{\tau}_{n+1}]$, by

$$\widetilde{X}(t) = \eta_n + \frac{t - \widetilde{\tau}_n}{\widetilde{\tau}_{n+1} - \widetilde{\tau}_n}(\eta_{n+1} - \eta_n),$$

one can easily adapt the arguments of Section 2 to show that the interpolated process $(\widetilde{X}(t))_{t\geq 0}$ is an asymptotic pseudo-trajectory of the semiflow induced by G.

Similarly, all the arguments of Section 5 apply to Model 2, including (12) which still holds. Indeed,

$$\begin{aligned} |(1-q)\widetilde{s}(\eta) + q(1-\widetilde{s}(\eta)) - (1-q)\widetilde{s}(\nu) - q(1-\widetilde{s}(\nu))| &\leq |\widetilde{s}(\eta) - \widetilde{s}(\nu)| \\ &\leq \frac{\left| p_A \widetilde{\alpha}^B_\eta \left(p_A \widetilde{\alpha}^B_\nu + p_B \widetilde{\alpha}^A_\nu \right) - p_A \widetilde{\alpha}^B_\nu \left(p_A \widetilde{\alpha}^B_\eta + p_B \widetilde{\alpha}^A_\eta \right) \right|}{\left(p_A \widetilde{\alpha}^B_\eta + p_B \widetilde{\alpha}^A_\eta \right)\left(p_A \widetilde{\alpha}^B_\nu + p_B \widetilde{\alpha}^A_\nu \right)} \\ &\leq \frac{p_A p_B \left| \widetilde{\alpha}^A_\nu \widetilde{\alpha}^B_\eta - \widetilde{\alpha}^A_\eta \widetilde{\alpha}^B_\nu \right|}{\left(p_A \left((1-q)p_B + q p_A \right) + p_B \left((1-q)p_A + q p_B \right) \right)^2}. \end{aligned}$$

For the numerator, we have

$$\begin{aligned}\left|\widetilde{\alpha}_\nu^A\widetilde{\alpha}_\eta^B-\widetilde{\alpha}_\eta^A\widetilde{\alpha}_\nu^B\right| &\le \widetilde{\alpha}_\nu^A\left|\widetilde{\alpha}_\eta^B-\widetilde{\alpha}_\nu^B\right|+\widetilde{\alpha}_\nu^B\left|\widetilde{\alpha}_\nu^A-\widetilde{\alpha}_\eta^A\right|\\ &\le \left(\widetilde{\alpha}_\nu^A+\widetilde{\alpha}_\nu^B\right)(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}\\ &\le (2^\gamma-1)\|\eta-\nu\|_1\,,\end{aligned}$$

where we used that $\left|\widetilde{\alpha}_\eta^\Lambda-\widetilde{\alpha}_\nu^\Lambda\right|\le(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}$ and that

$$\widetilde{\alpha}_\nu^A+\widetilde{\alpha}_\nu^B=1+\sum_{x\in E}\nu(x)\left((|x|+1)^\gamma-|x|^\gamma\right)\le 2^\gamma\le 2\,.$$

As for the denominator, it is equal to $\left(q(p_A^2+p_B^2)+2(1-q)p_Ap_B\right)^2$, which is larger $(2p_Ap_B)^2$, since $p_A^2+p_B^2\ge 2p_Ap_B$. All in all, we have

$$|(1-q)\widetilde{s}(\eta)+q(1-\widetilde{s}(\eta))-(1-q)\widetilde{s}(\nu)-q(1-\widetilde{s}(\nu))|\le\frac{2^\gamma-1}{4p_Ap_B}\|\eta-\nu\|_1\,.$$

7. Remaining proofs

7.1. **Proof of (12) and (13).** We have

$$\begin{aligned}|s(\eta)-s(\nu)| &= \frac{\left|r_A\alpha_\eta^B\left(r_A\alpha_\nu^B+r_B\alpha_\nu^A\right)-r_A\alpha_\nu^B\left(r_A\alpha_\eta^B+r_B\alpha_\eta^A\right)\right|}{\left(r_A\alpha_\eta^B+r_B\alpha_\eta^A\right)\left(r_A\alpha_\nu^B+r_B\alpha_\nu^A\right)}\\ &= \frac{r_Ar_B\left|\alpha_\eta^B\alpha_\nu^A-\alpha_\nu^B\alpha_\eta^A\right|}{\left(r_A\alpha_\eta^B+r_B\alpha_\eta^A\right)\left(r_A\alpha_\nu^B+r_B\alpha_\nu^A\right)}\,.\end{aligned}$$

Since α_η^Λ and α_ν^Λ are both larger than p_Λ, the denominator is larger than $(r_Ap_B+r_Bp_A)^2$, and we have

$$\frac{r_Ar_B}{(r_Ap_B+r_Bp_A)^2}=\frac{1}{p_Ap_B}\frac{r_Ap_B\cdot r_Bp_A}{(r_Ap_B+r_Bp_A)^2}\le\frac{1}{4p_Ap_B}\,.$$

For the numerator, we see that

$$\left|\alpha_\eta^B\alpha_\nu^A-\alpha_\nu^B\alpha_\eta^A\right|\le\alpha_\eta^B\left|\alpha_\nu^A-\alpha_\eta^A\right|+\alpha_\eta^A\left|\alpha_\eta^B-\alpha_\nu^B\right|\,.$$

Since both $\left|\alpha_\nu^A-\alpha_\eta^A\right|$ and $\left|\alpha_\eta^B-\alpha_\nu^B\right|$ are smaller than $(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}$, and since $\alpha_\eta^A+\alpha_\eta^B\le 1+2^\gamma-1=2^\gamma$, we have

$$|s(\eta)-s(\nu)|\le\frac{2^\gamma(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}}{4p_Ap_B}=\frac{2^\gamma(2^\gamma-1)\|\eta-\nu\|_1}{8p_Ap_B}\,,$$

which establishes (12). Moving on to (13), we write $\lambda(\eta)=\alpha_\eta^As(\eta)+\alpha_\eta^B(1-s(\eta))$ and apply the triangle inequality to get

$$|\lambda(\eta)-\lambda(\nu)|\le(\alpha_\eta^A+\alpha_\eta^B)|s(\eta)-s(\nu)|+s(\nu)|\alpha_\eta^A-\alpha_\nu^A|+(1-s(\nu))|\alpha_\eta^B-\alpha_\nu^B|\,.$$

Combining the previous bound on $|s(\eta)-s(\nu)|$, the fact that $\alpha_\eta^A+\alpha_\eta^B\le 2^\gamma$, and that both $\left|\alpha_\nu^A-\alpha_\eta^A\right|$ and $\left|\alpha_\eta^B-\alpha_\nu^B\right|$ are smaller than $(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}$, we obtain

$$|\lambda(\eta)-\lambda(\nu)|\le(2^\gamma-1)\|\eta-\nu\|_{\mathrm{TV}}\left(1+\frac{2^\gamma}{4p_Ap_B}\right)\le\frac{1}{2}\|\eta-\nu\|_1\left(1+\frac{1}{2p_Ap_B}\right)\,,$$

which establishes (13). □

7.2. **Proof of (14).** We start by decomposing

$$\sup_{m\geq n}\left\|\sum_{i=n}^{m}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1 \leq \underbrace{\sup_{m\geq n}\left\|\sum_{i=n}^{m}\frac{1}{i+1}(\delta_{X_{i+1}}-\mathbb{E}\left[\delta_{X_{i+1}}\,|\,\mathcal{F}_i\right])\right\|_1}_{(\mathtt{I})} + \underbrace{\max_{\Lambda\in\{A,B\}}\sum_{i=n}^{+\infty}\frac{1}{i+1}\left\|\frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)}-\frac{\eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}\right\|_1}_{(\mathtt{II})},$$

with $\widetilde{m}_i=\frac{m_i}{i}$. For term (I), using the Burkholder–Davis–Gundy inequality, we see that there exists $C_2>0$ such that

$$\mathbb{E}[(\mathtt{I})^2]\leq C_2\mathbb{E}\left[\sum_{i=n}^{+\infty}\frac{1}{(i+1)^2}\|\delta_{X_{i+1}}-\mathbb{E}\left[\delta_{X_{i+1}}\,|\,\mathcal{F}_i\right]\|_1^2\right].$$

Now we have

$$\mathbb{E}\left[\left\|\delta_{X_{i+1}}-\mathbb{E}\left[\delta_{X_{i+1}}\,|\,\mathcal{F}_i\right]\right\|_1^2\right]=\mathbb{E}\left[\left(\sum_{x\in E}|\delta_{X_{i+1}}(x)-\mathbb{E}\left[\delta_{X_{i+1}}(x)\,|\,\mathcal{F}_i\right]|\right)^2\right]\leq 4\,,$$

where we used the triangle inequality and the fact that $\sum_{x\in E}\delta_{X_{i+1}}(x)=1$. Hence we get

$$\mathbb{E}[(\mathtt{I})^2]\leq 4C_2\sum_{i\geq n}\frac{1}{(i+1)^2}\leq\frac{4C_2}{n}\,. \tag{31}$$

As for term (II), we have

$$\sum_{i=n}^{+\infty}\frac{1}{i+1}\left\|\frac{\widetilde{m}_i D^\Lambda}{\widetilde{m}_i D^\Lambda(E)}-\frac{\eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}\right\|_1 \leq \underbrace{\sum_{i=n}^{+\infty}\frac{1}{i+1}\left\|\frac{\widetilde{m}_i D^\Lambda-\eta_i Q^\Lambda}{\eta_i Q^\Lambda(E)}\right\|_1}_{(\mathtt{II.a})} + \underbrace{\sum_{i=n}^{+\infty}\frac{1}{i+1}\|\widetilde{m}_i D^\Lambda\|_1\left|\frac{1}{\widetilde{m}_i D^\Lambda(E)}-\frac{1}{\eta_i Q^\Lambda(E)}\right|}_{(\mathtt{II.b})}.$$

For term (II.a), using the fact that $\eta_i Q^\Lambda(E)$ is lower bounded by p_Λ and recalling that $\widetilde{m}_i D^\Lambda-\eta_i Q^\Lambda=(\widetilde{m}_i-\eta_i R)D^\Lambda$, we see that Lemma 3 together with a variance bound for the binomial distribution imply that

$$\mathbb{E}\left[(\mathtt{II.a})\right]\lesssim\sum_{i=n}^{+\infty}\frac{1}{(i+1)^2}\mathbb{E}\left[\left|\sum_{j=1}^{i}(B_j-p)\right|\right]\lesssim\sum_{i=n}^{+\infty}\frac{1}{(i+1)^{3/2}}\lesssim\frac{1}{\sqrt{n}}\,. \tag{32}$$

As for term (II.b), noticing that $\|\widetilde{m}_i D^\Lambda\|_1=\widetilde{m}_i D^\Lambda(E)$, that $\eta_i Q^\Lambda(E)$ is larger than p_Λ, and applying Lemma 3 again, we have

$$\mathbb{E}\left[(\mathtt{II.b})\right]\lesssim\sum_{i=n}^{+\infty}\frac{1}{i+1}\mathbb{E}\left[|\eta_i Q^\Lambda(E)-\widetilde{m}_i D^\Lambda(E)|\right]\lesssim\sum_{i=n}^{+\infty}\frac{1}{(i+1)^{3/2}}\lesssim\frac{1}{\sqrt{n}}\,. \tag{33}$$

Combining (31) (together with the Jensen's inequality), (32) and (33), we obtain

$$\mathbb{E}\left[\sup_{m\geq n}\left\|\sum_{i=n}^{m}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1\right]\lesssim\frac{1}{\sqrt{n}}\,.$$

Markov's inequality then gives, for all $\varepsilon>0$,

$$\sum_{k\geq 1}\mathbb{P}\left(\sup_{m\geq 2^k}\left\|\sum_{i=2^k}^{m}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1>\varepsilon\right)\lesssim\sum_{k\geq 1}2^{-k/2}<+\infty\,,$$

so that, by the Borel–Cantelli lemma, we have $\sup_{m\geq 2^k}\left\|\sum_{i=2^k}^{m}\frac{1}{i+1}\varepsilon_{i+1}\right\|_1\to 0$ almost surely as $k\to+\infty$, and we can conclude by monotonicity. □

7.3. **Proof of Lemma 10.** We aim to show that the equation $s(\pi_u)=u$ admits a unique solution $u\in[0,1]$. The following equivalences hold:

$$\begin{aligned}
s(\pi_u)=u &\iff r\pi_uQ^B(E)=u(1-r)\pi_uQ^A(E)+ur\pi_uQ^B(E)\\
&\iff u\pi_uQ^A(E)=r\underbrace{[u\pi_uQ^A(E)+(1-u)\pi_uQ^B(E)]}_{=\pi_uQ_u(E)=\lambda(u)}\\
&\iff \psi(u):=\frac{u\pi_uQ^A(E)}{u\pi_uQ^A(E)+(1-u)\pi_uQ^B(E)}=r\,.
\end{aligned}$$

Observe that the function ψ is such that $\psi(0)=0$ and $\psi(1)=1$. We will show that ψ is strictly increasing and continuous by proving that its derivative is strictly positive. This will allow us to establish the existence and uniqueness of a solution u to the equation $\psi(u)=r$ as $r\in[0,1]$.
Let φ be the function defined on $(0,\infty)$ by

$$\varphi(x)=\sum_{k=1}^{\infty}\prod_{j=1}^{k}\frac{1}{1+xj^{-\gamma}}\,.$$

Note that φ is strictly decreasing and differentiable on $(0,\infty)$, with derivative given by

$$\varphi'(x)=-\sum_{k=1}^{\infty}\left(\prod_{j=1}^{k}\frac{1}{1+xj^{-\gamma}}\right)\sum_{i=1}^{k}\frac{1}{i^{\gamma}+x}\,.$$

With this notation, a straightforward calculation shows that

$$\pi_uQ^A(E)=\frac{\lambda(u)}{u}p\,\varphi\left(\frac{\lambda(u)}{u}\right)\quad\text{and}\quad\pi_uQ^B(E)=\frac{\lambda(u)}{1-u}(1-p)\,\varphi\left(\frac{\lambda(u)}{1-u}\right)\,,$$

which yields the following simplified expression for ψ:

$$\psi(u)=\frac{p\,\varphi\left(\frac{\lambda(u)}{u}\right)}{p\,\varphi\left(\frac{\lambda(u)}{u}\right)+(1-p)\,\varphi\left(\frac{\lambda(u)}{1-u}\right)}=p\,\varphi\left(\frac{\lambda(u)}{u}\right)\,,$$

since

$$p\,\varphi\left(\frac{\lambda(u)}{u}\right)+(1-p)\,\varphi\left(\frac{\lambda(u)}{1-u}\right)=1\,.\tag{34}$$

Now, let h be the function defined by $h(u) = \frac{\lambda(u)}{u}$. We show that h is strictly decreasing. For $u \in (0,1)$, Equation (34) can be rewritten as

$$p\,\varphi\left(h(u)\right) + (1-p)\,\varphi\left(h(u)\frac{u}{1-u}\right) = 1\,.$$

Defining f by

$$\begin{array}{rcl} f : (0,1)\times(0,\infty) & \longrightarrow & \mathbb{R} \\ (u,y) & \longmapsto & p\,\varphi(y) + (1-p)\,\varphi(y\times\frac{u}{1-u}) - 1\,, \end{array}$$

we have the following equivalence:

$$f(u,y) = 0 \iff y = \frac{\lambda(u)}{u}\,.$$

By the implicit function theorem, the derivative of h is therefore given by

$$h'(u) = -\frac{\frac{\partial f}{\partial u}(u,h(u))}{\frac{\partial f}{\partial y}(u,h(u))} = -\frac{\frac{(1-p)}{(1-u)^2}\varphi'\left(\frac{\lambda(u)}{1-u}\right)}{p\varphi'\left(\frac{\lambda(u)}{u}\right) + \frac{(1-p)u}{1-u}\varphi'\left(\frac{\lambda(u)}{1-u}\right)}\,.$$

Since h' is strictly negative on $(0,1)$, h is strictly decreasing. Moreover, as φ is also strictly decreasing, it follows by composition that ψ is strictly increasing. Standard arguments show that ψ is continuous on $[0,1]$. We have thus established the existence of a unique s satisfying $\psi(u) = r$, and hence $s(\pi_u) = u$. This completes the proof. □

Acknowledgments. The authors warmly thank Pierre Monmarché for insightful discussions at various stages of this work and for his kindness. They also thank Nicolas Fournier for stimulating discussions on jump processes and for his interest in this work.

References

[1] N. Antunes, S. Banerjee, S. Bhamidi, and V. Pipiras. Attribute network models, stochastic approximation and network sampling. *The Annals of Applied Probability*, 36(1):652 – 702, 2026. 5

[2] K. B. Athreya. Preferential attachment random graphs with general weight function. *Internet Mathematics*, 4(4):401–418, 2007. 4

[3] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999. 3

[4] M. Benaïm. Dynamics of stochastic approximation algorithms. In P. Bernard, editor, *École d'Été de Probabilités de Saint-Flour*, volume 33 of *Lecture Notes in Mathematics*, pages 1–68. Springer, Berlin, Heidelberg, 1999. 7, 10, 13

[5] S. Bhamidi, R. Fan, N. Fraiman, and A. Nobel. Community modulated recursive trees and population dependent branching processes. *Random Structures & Algorithms*, 60 (2):201–232, 2022. 3, 4

[6] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001. 3

[7] L.-P. Chaintron and D. Antoine. Propagation of chaos: A review of models, methods and applications. I. Models and methods. *Kinetic and Related Models*, 15(6):895–1015, 2022. 20, 24

[8] L.-P. Chaintron and D. Antoine. Propagation of chaos: A review of models, methods and applications. II. Applications. *Kinetic and Related Models*, 15(6):1017–1173, 2022. 24

[9] N. Champagnat and D. Villemonais. Exponential convergence to quasi-stationary distribution and Q-process. *Probability Theory and Related Fields*, 164(1):243–283, Feb 2016. 19

[10] N. Champagnat and D. Villemonais. General criteria for the study of quasi-stationarity. *Electronic Journal of Probability*, 28:1–84, 2023. 19

[11] P. Jagers and O. Nerman. *The asymptotic composition of supercritical, multi-type branching populations*, pages 40–54. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996. 6

[12] J. Jordan. Geometric preferential attachmrev. lett.ent in non-uniform metric spaces. *Electronic Journal of Probability*, 18:1 – 15, 2013. 3, 4, 5, 6

[13] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Physical Review Letters*, 85:4629–4632, Nov 2000. 4

[14] C. Mailler and J.-F. Marckert. Measure-valued pólya urn processes. *Electronic Journal of Probability*, 22:1–33, 2017. 6

[15] C. Mailler and D. Villemonais. Stochastic approximation on noncompact measure spaces and application to measure-valued pólya processes. *The Annals of Applied Probability*, 30(5):2393–2438, 2020. 6, 7, 19

[16] P. Monmarché. Elementary coupling approach for non-linear perturbation of markov processes with mean-field jump mechanisms and related problems. *ESAIM: Probability and Statistics*, 27:278–323, 2023. 8

[17] T. Móri. On random trees. *Studia Scientiarum Mathematicarum Hungarica*, 39(1–2): 143–155, 2002. 3

[18] R. Oliveira and J. Spencer. Connectivity Transitions in Networks with Super-Linear Preferential Attachment. *Internet Mathematics*, 2(2), 2005. 4

[19] A. Rudas. Random tree growth with general weight function, 2004. URL https://arxiv.org/abs/math/0410532. 4

[20] A. Rudas, B. Tóth, and B. Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007. 4, 5, 6, 17

[21] S. Sethuraman and S. C. Venkataramani. On the growth of a superlinear preferential attachment scheme". In P. Friz, W. König, C. Mukherjee, and S. Olla, editors, *Probability and Analysis in Interacting Physical Systems*, pages 243–265. Springer International Publishing, 2019. 4

Sorbonne Université, Université Paris Cité, CNRS, Laboratoire de Probabilités, Statistique et Modélisation, LPSM, F-75005 Paris, France

Email address: abenhamou@lpsm.paris

Email address: broutinn@lpsm.paris

Email address: ccazaux@lpsm.paris