Convergence Rates for the Degree Distribution in a Dynamic Network Model

Fabian Kück and Dominic Schuhmacher

Institute for Mathematical Stochastics
University of Göttingen
Goldschmidtstraße 7
37077 Göttingen
Germany

e-mail: fabian.kueck@mathematik.uni-goettingen.de
dominic.schuhmacher@mathematik.uni-goettingen.de

Abstract: In the stochastic network model of Britton and Lindholm [BL10], the number of individuals evolves according to a supercritical linear birth and death process, and a random social index is assigned to each individual at birth, which controls the rate at which connections to other individuals are created. We derive a rate for the convergence of the degree distribution in this model towards the mixed Poisson distribution determined by Britton and Lindholm based on heuristic arguments. In order to do so, we deduce the degree distribution at finite time and derive an approximation result for mixed Poisson distributions to compute an upper bound for the total variation distance to the asymptotic degree distribution.

MSC 2010 subject classifications: Primary 60J27; secondary 05C80, 62E20.
Keywords and phrases: mixed Poisson distribution, dynamic random graph, small worlds.

1. Introduction

"Network Science" is a relatively young, rapidly growing research area dealing with complex systems with an underlying graph structure (see e.g. the recent book by van der Hofstad [vdH15]). One of the first random graph models is the well-known Erdős-Rényi model, which is a static model, i.e. it describes a random network at a fixed time. Although this model shows interesting behaviour, further random graph models were needed for the description of real networks since some empirical networks have important properties that are not represented by Erdős-Rényi Graphs. One of these is the power law property of the degree distribution. Preferential attachment models exhibit such a power law behaviour asymptotically and were popularized by Barabási and Albert [BA99]. Peköz, Röllin and Ross [PRR13] established convergence rates for the degree distribution in a discrete-time preferential attachment model in terms of the total variation distance. Besides preferential models, the so-called fitness models have gained huge popularity in the recent years. In those models the attachment does not only depend on the degree but also on a random intrinsic fitness that is determined at the birth of each node. The best-known fitness model was introduced in [BB01]. In this model the attachment mechanism is a combination with preferential attachment. Pure fitness models were for example considered in [CCDLRM02] and [SHR13]. The time-continuous random graph model that was introduced by Britton and Lindholm [BL10] and that we investigate here can be seen as pure fitness model that is particular realistic due to time continuity and possibly dying nodes as well as edges. Depending on the application, either preferential models or the model by Britton and Lindholm can be more realistic. Note that it can be shown that the degrees of nodes in the model by Britton and Lindholm can in particular exhibit power laws such that this model displays an interesting alternative mechanism for producing graphs with this property.

Let us give a *loop-free* version of the definition of the original *dynamic network model by Britton and Lindholm*; see [BL10] and [BLT11]:

We examine a finite undirected graph without loops that develops over time. The node process $(Y_t)_{t\geq 0}$ is a linear birth and death process with initial value one. Thus each node gives birth at constant rate λ and dies at constant rate μ independently from other nodes. We assume that $(Y_t)_{t\geq 0}$ has right-continuous trajectories. The process $(Y_t)_{t\geq 0}$ and all other random variables that are defined in what follows to describe the dynamic random graph are defined on a common underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

We assume $\lambda > \mu$, so that the node process $(Y_t)_{t\geq 0}$ is a supercritical continuous-time Markov branching process. From standard branching process theory, we obtain that for a random variable W_t with $\mathcal{L}(W_t) = \mathcal{L}(Y_t e^{-t(\lambda-\mu)}|Y_t > 0)$, we have

$$W_t \to W \text{ a.s.},$$

where W is $\mathrm{Exp}(\frac{\lambda}{\lambda-\mu})$ distributed (see e.g. [Har50], page 319).

We equip every node i with a positive random social index S_i, where the S_i are i.i.d. with finite expectation and independent of all other random variables.

This allows us to define the development of the edge set. At birth every node is isolated. During its lifetime and *as long as there is at least one other node*, node i generates and destroys edges according to a birth and death process with constant birth rate αS_i and per-edge death rate β. Here α and β are positive constants. The "second" node of each newly born edge is chosen uniformly at random from the set of all *other* living nodes, and all of the edge processes (including the choices of the second nodes) are independent of each other and all other events.

In addition to the direct destruction of edges in the above process, all edges connected to a certain node are removed when the node dies.

Remark 1.1. *The only difference to the definition by Britton and Lindholm is that we do not allow loops because these are not present in most applications. Note that the proofs become slightly simpler if we use the original model by Britton and Lindholm, essentially because times where $Y_t = 1$ need not obtain special treatment. The upper bounds remain largely the same; see also Remark 4.2 for the pure birth case.*

Note that we still allow multiple edges. However, it can be shown that those are negligible in the sense that the probability that a randomly picked node has at least one multiple edge converges to zero at an exponential rate (see Appendix A2). This allows us to formulate the main result also for the case where we ignore multiple edges (see Corollary 1.4 below).

We refer to the distribution of the number of edges incident to a node picked uniformly at random from all living nodes at time t given the number of nodes is positive as *degree distribution*, and denote it by ν_t. In [BL10], Britton and Lindholm give a rather heuristic argument for the convergence of the degree distribution in the original model towards a mixed Poisson distribution ν. It is the main purpose of this paper to give a rate in total variation distance rather than a mere convergence result, providing full proof for this rate and thereby also for the convergence. The distribution ν is given by

$$\nu = \mathrm{MixPo}\bigg(\frac{\alpha}{\beta+\mu}\big(S + \mathbb{E}(S)\big)\big(1 - e^{-(\beta+\mu)A}\big)\bigg),$$

where $A \sim \mathrm{Exp}(\lambda)$, S has the social index distribution, and A and S are independent. Here MixPo denotes the mixed Poisson distribution. The following result is an immediate consequence of the main results proved in this article, Theorems 4.1 and 4.3.

Theorem 1.2. *Let $\mathbb{E}(S^2) < \infty$. Then we obtain for the degree distribution ν_t in the Britton–Lindholm Model without loops*

(a) if $\mu = 0$, then $d_{TV}(\nu_t, \nu) = O\big(\sqrt{t}e^{-\frac{1}{2}\lambda t}\big)$ as $t \to \infty$;
(b) if $\mu > 0$, then $d_{TV}(\nu_t, \nu) = O\big(t^2 e^{-\frac{1}{6}(\lambda-\mu)t}\big)$ as $t \to \infty$.

Remark 1.3. *A positive death rate μ leads to a higher variability in the degree distributions for finite t since in particular the death of a highly connected node (hub) can have a large impact. Thus we would* *not expect the same rate as in the pure birth case. However, the actual factor in the exponential rate may be larger than the one stated in the theorem.*

Theorem 1.2 has consequences for the case where we ignore multiple edges. Let $\tilde{\nu}_t$ be the distribution of the number of neighbours of a node picked uniformly at random from all living nodes at time t given the number of nodes at time t is positive. We refer to $\tilde{\nu}_t$ shortly as distribution of the number of neighbours. The convergence of this distribution to the asymptotic degree distribution ν is an immediate consequence of the following corollary, which is proved in Appendix A2.

Corollary 1.4. *Let $\mathbb{E}(S^2) < \infty$. For the distribution of the number of neighbours in the Britton–Lindholm Model without loops, we have that $d_{TV}(\tilde{\nu}_t, \nu) = O(t^2 e^{-\frac{1}{6}(\lambda-\mu)t})$ as $t \to \infty$.*

Remark 1.5. *Britton and Lindholm also introduced a modified model where the "second" node of each newly born edge is not picked uniformly. Instead the probability for each node being the "second" node is proportional to its social index. We expect that one can prove similar results for this modified model analogously to the results for the model that we treat here. Note that the modification can lead to a significantly higher ratio of multiple edges. Therefore the modified version is less interesting for many applications.*

The rest of the paper is organized as follows. In Section 2 we collect some well known and less known facts about linear birth and death processes. Section 3 gives the degree distribution at finite time by using results about the birth and death processes of the edges. In Section 4, we derive upper bounds for the total variation distance between finite-time and asymptotic degree distributions, treating the pure birth case and the general case separately, which leads to Theorem 1.2 above. In order to achieve this, we derive a universal bound for the total variation distance between two general mixed Poisson distributions; see Theorem 4.5. Appendix A1 contains lemmas about linear birth and death processes that might be of general interest and further technical lemmas needed for the proof of our main theorem 4.3. In Appendix A2 we treat the negligibility of multiple edges.

For a few of the more elementary proofs throughout the paper, we refer to the technical report [KS15].

2. Linear birth and death processes and the age of a randomly picked individual

The node process $(Y_t)_{t\geq 0}$ is a linear birth and death process with birth rate λ, death rate $\mu < \lambda$ and initial value one, i.e. $Y_0 = 1$. According to (8.15) and (8.46) in [Bai64], the one-dimensional distributions of such a process are given by the following probability mass functions:

$$\begin{aligned} p_0(t) &= \mu\tilde{p}(t) \\ p_n(t) &= (1-\mu\tilde{p}(t))(1-\lambda\tilde{p}(t))(\lambda\tilde{p}(t))^{n-1}, \ n \geq 1, \end{aligned}$$

where

$$\tilde{p}(t) := \frac{e^{(\lambda-\mu)t}-1}{\lambda e^{(\lambda-\mu)t}-\mu} = \frac{1}{\lambda}\frac{1-e^{-(\lambda-\mu)t}}{1-\frac{\mu}{\lambda}e^{-(\lambda-\mu)t}}.$$

Remark 2.1. *Note that $p_0(t)$ is the probability that a linear birth and death process with initial value one goes extinct up to time t. Due to the branching property of a linear birth and death process, we have that $p_0(t)^m$ is the probability that a linear birth and death process with a general initial value m goes extinct up to time t. By taking the limit $t \to \infty$, we obtain that the probability of eventual extinction is $(\mu/\lambda)^m$ if $\lambda > \mu$ (see e.g. (8.59) in [Bai64]).*

By elementary computations using these probability mass functions, we obtain the following proposition (cf. (8.16), (8.17), (8.48) and (8.49) in [Bai64]).

Proposition 2.2. *We have*

$$\mathbb{E}(Y_t) = e^{(\lambda-\mu)t} \text{ and } Var(Y_t) = \frac{\lambda+\mu}{\lambda-\mu}(e^{2(\lambda-\mu)t} - e^{(\lambda-\mu)t}).$$

From [KS17] we know furthermore an expression and an upper bound for the conditional expectation of $1/Y_t$ given $Y_t > 0$ that is essentially of the anticipated order $e^{-(\lambda-\mu)t}$ as $t \to \infty$.

Proposition 2.3 (KS17, Lemma 3.1)**.** *For any $t > 0$, we have*

$$E\left(\frac{1}{Y_t}\,\middle|\, Y_t > 0\right) = \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)t} - \lambda}\log\left(\frac{\lambda e^{(\lambda-\mu)t} - \mu}{\lambda-\mu}\right) \leq \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)t} - \lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)t\right).$$

We may apply this bound in order to obtain an upper bound for $\mathbb{E}(Y_t^{-1/2}\,|\,Y_t > 0)$.

Proposition 2.4. *For $\lambda > \mu$ and $t \geq \frac{1}{\lambda-\mu}\log(2)$, we have*

$$\mathbb{E}\left(\frac{1}{\sqrt{Y_t}}\,\middle|\, Y_t > 0\right) \leq e^{-\frac{1}{2}(\lambda-\mu)t}\sqrt{\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)t\right)}.$$

Proof: For $\lambda > \mu$ and $t \geq \frac{1}{\lambda-\mu}\log(2)$, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sqrt{Y_t}}\,\middle|\, Y_t > 0\right) &\leq \sqrt{\mathbb{E}\left(\frac{1}{Y_t}\,\middle|\, Y_t > 0\right)} \\
&\leq \sqrt{\frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)t} - \lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)t\right)} \\
&\leq \sqrt{\frac{2(\lambda-\mu)}{\lambda e^{(\lambda-\mu)t}}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)t\right)} \\
&= e^{-\frac{1}{2}(\lambda-\mu)t}\sqrt{\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)t\right)},
\end{aligned}$$

where the second line follows from Lemma 3.1 in [KS17]. □

For the proof of our main theorem a finer analysis is required. Denote by B_t and D_t the numbers of births and deaths up to time t, respectively, where we set $B_0 = Y_0 = 1$ and $D_0 = 0$. By using a partial differential equation for the joint cumulant generating function of B_t and Y_t stated in [CM77], we obtain the following formulae for the first and second joint moments (see [KS15] for details).

Proposition 2.5 (KS15, Proposition 3.6)**.** *We have*

$$\begin{aligned}
\mathbb{E}(B_t) &= \frac{\lambda}{\lambda-\mu}e^{(\lambda-\mu)t} - \frac{\mu}{\lambda-\mu} \\
Cov(B_t, Y_t) &= \frac{\lambda(\lambda+\mu)}{(\lambda-\mu)^2}e^{2(\lambda-\mu)t} - \frac{2\lambda\mu}{\lambda-\mu}te^{(\lambda-\mu)t} - \frac{\lambda^2}{(\lambda-\mu)^2}e^{(\lambda-\mu)t} \\
Var(B_t) &= \frac{\lambda^2(\lambda+\mu)}{(\lambda-\mu)^3}e^{2(\lambda-\mu)t} - \frac{4\lambda^2\mu}{(\lambda-\mu)^2}te^{(\lambda-\mu)t} + \left(\frac{2\lambda^2\mu}{(\lambda-\mu)^3} - \frac{\lambda(\lambda+\mu)}{(\lambda-\mu)^2}\right)e^{(\lambda-\mu)t}.
\end{aligned}$$

Remark 2.6. *Note that $\mathbb{E}(Y_t) = \mathbb{E}(B_t - D_t) = \mathbb{E}(B_t) - \mathbb{E}(D_t)$ and $\frac{\mathbb{E}(B_t)-1}{\mathbb{E}(D_t)} = \frac{\lambda}{\mu} = \frac{\lambda}{\lambda+\mu}(\frac{\mu}{\lambda+\mu})^{-1}$ is the ratio of the probabilities of a birth and a death at each event time. Furthermore, the sum $\mathbb{E}(B_t)+\mathbb{E}(D_t) = \frac{\lambda+\mu}{\lambda-\mu}e^{(\lambda-\mu)T} - \frac{2\mu}{\lambda-\mu}$ is the expected number of events up to time t.*

In the rest of this section we summarize the results about the age of an individual picked uniformly at random at a fixed time $T > 0$ (given $Y_T > 0$). We briefly call the distribution of this age *the age distribution of* $(Y_t)_{t\geq 0}$ *at time* T. In the pure birth case, the age distribution has a simple form.

Proposition 2.7 (Neuts and Resnick [NR71, Theorem 1]; KS15, Proposition 3.9). *Let $\mu = 0$. The ages of the individuals at time T that have been born after time zero are i.i.d. truncated exponentially distributed, more precisely they have distribution $\mathcal{L}(Z|Z \leq T)$, where $Z \sim \mathrm{Exp}(\lambda)$.*

In the general case, we first state the conditional age distribution given the population size, which we know from [KS17].

Theorem 2.8 (KS17, Theorem 2.1). *Let F_{y_T} denote the cumulative distribution function of the age of an individual picked uniformly at random at time T given $Y_T = y_T$ for some $y_T > 0$. Then F_{y_T} is given by*

$$F_{y_T}(t) = \frac{y_T - 1}{y_T}\left(1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T}e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}}\right) + \frac{1}{y_T}\left(\frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu}\mathbb{1}_{\{t<T\}} + \mathbb{1}_{\{t=T\}}\right)$$

for $t \in [0, T]$.

A simple computation yields then the unconditional age distribution (see [KS17] for details).

Corollary 2.9 (cf. KS17, Corollary 2.3). *The cumulative distribution function F of the age of an individual picked uniformly at random at time T is given by*

$$\begin{aligned} F(t) = &\left(1 - \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda}\log\left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu}\right)\right)\left(1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T}e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}}\right) \\ &+ \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda}\log\left(\frac{\lambda e^{(\lambda-\mu)T} - \mu}{\lambda - \mu}\right)\left(\frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu}\mathbb{1}_{\{t<T\}} + \mathbb{1}_{\{t=T\}}\right) \end{aligned}$$

for $t \in [0, T]$.

The next corollary states that the age distribution converges exponentially fast to the $\mathrm{Exp}(\lambda)$ distribution in a certain sense for $\lambda > \mu$. It is an immediate consequence of Corollary 2.9.

Corollary 2.10 (KS17, Corollary 2.4). *Let $\lambda > \mu$, and let A denote the age of an individual picked uniformly at random at time T. Then there exists a random variable Z with $\mathcal{L}(Z|Y_T > 0) = \mathrm{Exp}(\lambda)$ such that*

$$\mathbb{E}\Big(\big|e^{-cA} - e^{-cZ}\big| \;\Big|\; Y_T > 0\Big) \leq \frac{\lambda}{c+\lambda}\frac{1}{e^{(\lambda-\mu)T} - 1} + \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda}\left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right)$$

for any $c > 0$.

Another random quantity we need to control for our main bound is, at any fixed time T, the time since the last event has occurred. Let $0 = T_1 < T_2 < \dots$ be the event times of $(Y_t)_{t\geq 0}$. Since $B_T + D_T$ is the number of events up to time T, the random variable $T - T_{B_T+D_T}$ describes the quantity we are interested in. The following result states that the $\mathrm{Exp}((Y_T - 1)\lambda)$ distribution is a stochastic upper bound given Y_T on $\{Y_T > 1\}$ and follows from Theorem 2.8 above (see [KS15] for details).

Theorem 2.11 (KS15, Theorem 3.19). *Given $Y_T = y_T$, the distribution of $T - T_{B_T+D_T}$ is stochastically dominated by the distribution with cumulative distribution function*

$$G(t) = \mathbb{1}_{\{y_T>1\}}(1 - e^{-(y_T-1)\lambda t}) + \mathbb{1}_{\{y_T\leq 1\}}\mathbb{1}_{\{t\geq T\}}.$$

3. Degree distribution at finite time

If at least two nodes are alive, each node i spawns edges to other nodes according to a birth and death process with constant birth rate $\alpha' = \alpha S_i$ and linear death rate with factor β. The one-dimensional distributions of this process are well-known.

Proposition 3.1. *Let $(Z_t)_{t\geq 0}$ be a birth and death process with constant birth rate α' and linear death rate $\beta' = \beta n$ if the process is in state $n \in \mathbb{N} = \{1, 2, \ldots\}$. If the process is started deterministically at $k \in \mathbb{N}$, we have*

$$Z_t \sim \mathrm{Po}\big(\tfrac{\alpha'}{\beta}(1 - e^{-\beta t})\big) * \mathrm{Bin}\big(k, e^{-\beta t}\big)$$

for every $t \geq 0$, where $*$ denotes convolution.

In what follows, we derive the degree distribution at finite time T based on this result.

3.1. The pure birth case

We deal with the case $\mu = 0$ first. Let the nodes be ordered by their birth times. Let S_i be the social index of node i and $A_i(T)$ its age at time T. For convenience, we define $A_2(T) = 0$ if $Y_T = 1$. Furthermore, given $Y_T = y_T$, let the random variable J_T be uniformly distributed on $\{1, \ldots, y_T\}$ and independent of the ages, the social indices and the rest of the path $(Y_t)_{0\leq t<T}$. We interpret J_T as the index of a node that is randomly picked at time T among all living nodes.

For the time being, we condition on $(Y_t)_{0\leq t\leq T} = (y_t)_{0\leq t\leq T}$, the social indices $(S_k)_{k\in\mathbb{N}} = (s_k)_{k\in\mathbb{N}}$ and $J_T = j_T$. Let $T = a_1 > \ldots > a_{y_T}$ denote the corresponding ages of the individuals.

Firstly, we consider the number of edges created by node j_T. Assuming $y_T > 1$, the edges created by j_T form a birth and death process with constant birth rate αs_{j_T} and linear death rate with factor β, started in zero at time $T - a_{\max(j_T,2)}$ since no edges are created if there is only one living node. By Proposition 3.1, the number of edges alive at time T that j_T has created has distribution

$$\mathrm{Po}\bigg(\frac{\alpha s_{j_T}}{\beta}(1 - e^{-\beta a_{\max(j_T,2)}})\bigg). \tag{3.1}$$

We have to add to this the number of edges alive at time T that *other* nodes have created and connect to j_T. Consider a fixed node $i \in \{1, \ldots, y_T\} \setminus \{j_T\}$ and some time interval of the form $[T - a_l, T - a_{l+1})$ for $l \geq i \vee 2$. The number of edges that connect i to j_T and that survive until the end of this interval, i.e. until the birth time $T - a_{l+1}$ of node $l + 1$, can be described by a birth and death process of the above type again. The birth rate is constant and equal to $\alpha s_i \frac{1}{l-1}$ because node i creates edges at rate αs_i and $\frac{1}{l-1}$ is the probability that an edge that is created in the interval $[T - a_l, T - a_{l+1})$ is connected to j_T. The death rate is linear again with factor β. By Proposition 3.1, the number of edges created by i in $[T - a_l, T - a_{l+1})$ that connect to j_T and survive until $T - a_{l+1}$ is $\mathrm{Po}\big(\frac{\alpha s_i}{(l-1)\beta}(1 - e^{-\beta(a_l - a_{l+1})})\big)$-distributed.

We can extend the time interval by one birth time, i.e. we compute the distribution of the number of edges that i creates in $[T - a_l, T - a_{l+2})$, connect to j_T, and survive until $T - a_{l+2}$ by conditioning on the number Z of edges that i creates in $[T - a_l, T - a_{l+1})$, connect to j_T and survive until $T - a_{l+1}$. Given $Z = z$, by Proposition 3.1 and the Markov property, the number of edges that i creates in $[T - a_l, T - a_{l+2})$, connect to j_T, and survive until $T - a_{l+2}$ has distribution $\mathrm{Po}\big(\frac{\alpha s_i}{l\beta}(1 - e^{-\beta t(a_{l+1} - a_{l+2})})\big) * \mathrm{Bin}\big(z, e^{-\beta t}\big)$. We already know from above that $Z \sim \mathrm{Po}\big(\frac{\alpha s_i}{(l-1)\beta}(1 - e^{-\beta(a_l - a_{l+1})})\big)$. Thus if we do not condition on $Z = z$, we obtain the distribution

$$\mathrm{Po}\bigg(\frac{\alpha s_i}{l\beta}(1 - e^{-\beta(a_{l+1} - a_{l+2})}) + e^{-\beta(a_{l+1} - a_{l+2})}\frac{\alpha s_i}{(l-1)\beta}(1 - e^{-\beta(a_l - a_{l+1})})\bigg). \tag{3.2}$$

Starting at $l = \max(i, j_T)$ and iterating the procedure that leads to (3.2) until the whole interval $[T - a_{\max(i,j_T)}, T)$ is spanned, we see that the number of edges alive at time T that connect to j_T but have been created by other nodes has distribution

$$\begin{aligned}
&\mathrm{Po}\Bigg(\sum_{\substack{i=1\\ i\neq j_T}}^{y_T} \frac{\alpha s_i (1 - e^{-\beta a_{y_T}})}{(y_T - 1)\beta} + \sum_{\substack{i=1\\ i\neq j_T}}^{y_T} e^{-\beta a_{y_T}} \sum_{l=i\vee j_T}^{y_T - 1} \Bigg(\prod_{k=1}^{y_t - l - 1} e^{-\beta(a_{l+k} - a_{l+k+1})} \Bigg) \frac{\alpha s_i (1 - e^{-\beta(a_l - a_{l+1})})}{(l-1)\beta} \Bigg) \\
&= \mathrm{Po}\Bigg(\frac{\alpha}{\beta} \sum_{\substack{i=1\\ i\neq j_T}}^{y_T} \frac{s_i}{y_T - 1} (1 - e^{-\beta a_{y_T}}) + \frac{\alpha}{\beta} \sum_{\substack{i=1\\ i\neq j_T}}^{y_T} \sum_{l=i\vee j_T}^{y_T - 1} \frac{s_i}{l-1} (e^{-\beta a_{l+1}} - e^{-\beta a_l}) \Bigg). \qquad (3.3)
\end{aligned}$$

Since the numbers of outgoing and incoming edges are independent, the desired degree distribution is obtained by convoluting (3.1) and (3.3). Lifting the conditioning on $(Y_t)_{0\leq t\leq T}$, $(S_k)_{k\in\mathbb{N}}$ and J_T, we arrive at the following result.

Theorem 3.2. *For $\mu = 0$, the degree distribution in the Britton–Lindholm model without loops is the* $\mathrm{MixPo}(\Lambda_T)$ *distribution, where*

$$\begin{aligned}
\Lambda_T =& \frac{\alpha S_{J_T}}{\beta} (1 - e^{-\beta A_{\max(J_T, 2)}(T)}) + \frac{\alpha}{\beta} \sum_{\substack{i=1\\ i\neq J_T}}^{Y_T} \frac{S_i}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \\
&+ \frac{\alpha}{\beta} \sum_{\substack{i=1\\ i\neq J_T}}^{Y_T} \sum_{l=i\vee J_T}^{Y_T - 1} \frac{S_i}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}). \qquad (3.4)
\end{aligned}$$

3.2. The general case

For general μ, the degree distribution can be determined similarly to the pure birth case.

As in Section 2, let $0 = T_1 < \ldots < T_{B_T + D_T}$ be the event times up to T. We still enumerate nodes according to their birth times. Denote by $0 = T_1^+ < \ldots < T_{B_T}^+$ their birth times and by T_i^- the death time of the i-th node. Finally, given the subset of $\{1, \ldots, B_T\}$ that contains the indices of all living nodes at time T, let J_T be uniformly distributed on this set and independent of all other random variables as before. Note that S_{J_T} is (stochastically) independent of $((Y_t)_{0\leq t\leq T}, J_T)$.

We condition on $(Y_t)_{0\leq t\leq T} = (y_t)_{0\leq t\leq T}$ with $y_T > 0$, the social indices $(S_k)_{k\in\mathbb{N}} = (s_k)_{k\in\mathbb{N}}$ and $J_T = j_T$ again. Let b_T and d_T denote the corresponding number of births and deaths up to time T, respectively. Furthermore, let $0 = t_1 < t_2 < \ldots < t_{b_T + d_T}$ and $0 = t_1^+ < t_2^+ < \ldots < t_{b_T}^+$ be the corresponding event times and birth times up to time T, respectively.

Since deaths of other nodes reduce the number of edges created by node j_T, we cannot derive the distribution of the number of outgoing edges at time T in the same way as for the pure birth process. However, we can derive the total number of edges incident to j_T in a similar way as the number of incoming edges in the pure birth case. Consider a fixed node $i \neq j_T$ that is alive at time T (provided there are any) and some time interval of the form $[t_l, t_{l+1})$ for $t_l \geq t_i^+ \vee t_{j_T}^+$ and $t_{l+1} \leq T$. Note that the nodes i and j_T create edges with rate αs_i and αs_{j_T}, respectively, and that the probability that an edge that is created by i is connected to j_T is $\frac{1}{y_{t_l} - 1}$ and equal to the probability that an edge that is created by j_T is connected to i. Thus the number of edges created between i and j_T that survive until time t_{l+1} can be described by a birth and death process with constant birth rate $\alpha(s_i + s_{j_T})\frac{1}{y_{t_l} - 1}$ and linear death rate with factor β like before. By Proposition 3.1 we obtain that the number of edges that are created in $[t_l, t_{l+1})$ between i and j_T and survive until t_{l+1} has distribution

$$\mathrm{Po}\Bigg(\frac{\alpha(s_i + s_{j_T})}{(y_{t_l} - 1)\beta} (1 - e^{-\beta(t_{l+1} - t_l)}) \mathbb{1}_{\{y_{t_l} > 1\}} \Bigg).$$

Let the function $r = r_{(y_t)_{0\le t\le T}} : \{1,\dots,b_T\} \to \{1,\dots,b_T+d_T\}$ be defined such that $t_j^+ = T_{r(j)}$ for all $j \in \{1,\dots,b_T\}$, i.e. r maps birth number to event number.

Applying the same iterative procedure as in Subsection 3.1, starting at $l = r(i) \vee r(j_T)$ and continuing until the whole interval $[t_i^+ \vee t_{j_T}^+, T)$ is spanned, we can see that the number of edges between i and j_T alive at time T has distribution

$$\begin{aligned}\mathrm{Po}\bigg(&\frac{\alpha(s_i+s_{j_T})}{(y_T-1)\beta}(1-e^{-\beta(T-t_{b_T+d_T})})\mathbb{1}_{\{y_T>1\}}\\ &+e^{-\beta(T-t_{b_T+d_T})}\sum_{l=r(i)\vee r(j_T)}^{b_T+d_T-1}\bigg(\prod_{k=1}^{b_T+d_T-l-1}e^{-\beta(t_{l+k+1}-t_{l+k})}\frac{\alpha(s_i+s_{j_T})}{(y_{t_l}-1)\beta}(1-e^{-\beta(t_{l+1}-t_l)})\bigg)\mathbb{1}_{\{y_{t_l}>1\}}\bigg).\end{aligned} \tag{3.5}$$

Since the processes of edges between i and j_T are mutually independent for different i, we obtain the desired degree distribution by convoluting the distribution (3.5) for i running in $\{i' \in \{1,\dots,b_T\}\colon i' \neq j_T, t_{i'}^- > T\}$ and lifting the conditioning on $(Y_t)_{0\le t\le T}$, $(S_k)_{k\in\mathbb{N}}$ and J_T.

Theorem 3.3. *The degree distribution in the Britton–Lindholm model without loops is the* $\mathrm{MixPo}(\Lambda_T^*)$ *distribution, where* Λ_T^* *is a random variable with* $\mathcal{L}(\Lambda_T^*) = \mathcal{L}(\Lambda_T|Y_T>0)$ *and*

$$\begin{aligned}\Lambda_T :=&\frac{\alpha}{\beta}\sum_{\substack{i=1\\ i\neq J_T}}^{B_T}\frac{S_i+S_{J_T}}{(Y_T-1)}\mathbb{1}_{\{T_i^->T\}}(1-e^{-\beta(T-T_{B_T+D_T})})\mathbb{1}_{\{Y_T>1\}}\\ &+\frac{\alpha}{\beta}\sum_{\substack{i=1\\ i\neq J_T}}^{B_T}\mathbb{1}_{\{T_i^->T\}}\sum_{l=r(i)\vee r(J_T)}^{B_T+D_T-1}\frac{S_i+S_{J_T}}{(Y_{T_l}-1)}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\mathbb{1}_{\{Y_{T_l}>1\}}.\end{aligned} \tag{3.6}$$

4. Bounds on the total variation distance between the finite-time and the asymptotic degree distributions

Since in the pure birth case we obtain a much better bound with considerably less work, we treat the cases $\mu = 0$ and $\mu > 0$ separately in Subsections 4.1 and 4.2, giving the ideas of the proofs. The proofs themselves are deferred to Subsection 4.3.

4.1. The pure birth case

Let $\mu = 0$. Fix $T > 0$, and let the random variable Λ_T be defined as in Theorem 3.2. Furthermore, let S_{J_∞} and A_{J_∞} be independent random variables such that $S_{J_\infty}(\omega) = S_{J_T}(\omega)$ and $A_{J_\infty}(\omega) = F_\infty^{-1}(F_T(A_{J_T}(T,\omega)))$ for all $\omega \in \Omega$, where F_T and F_∞ are the cumulative distribution functions of $A_{J_T}(T)$ and the $\mathrm{Exp}(\lambda)$ distribution, respectively. Note that we obtain from Proposition 2.7 that $\mathcal{L}(A_{J_T}) \to \mathrm{Exp}(\lambda)$ weakly. The random variable A_{J_∞} is $\mathrm{Exp}(\lambda)$ distributed since F_T is continuous and hence $F_T(A_{J_T}(T))$ is uniformly distributed on $[0,1]$. Moreover, since the exponential distribution stochastically dominates the truncated exponential distribution, we have $F_\infty(A_{J_T}(T,\omega)) \le F_T(A_{J_T}(T,\omega))$ for all $\omega \in \Omega$. Since F_∞^{-1} is increasing, it follows that $A_{J_T}(T,\omega) \le A_{J_\infty}(\omega)$ for all $\omega \in \Omega$.

Let S be a generic random variable that is distributed according to the distribution of the social indices and let

$$\mathrm{M} = \frac{\alpha S_{J_\infty}}{\beta}(1-e^{-\beta A_{J_\infty}}) + \frac{\alpha\mathbb{E}(S)}{\beta}(1-e^{-\beta A_{J_\infty}}). \tag{4.1}$$

We know from Theorem 3.2 that $\mathrm{MixPo}(\Lambda_T)$ is the degree distribution at time T in the pure birth case. Theorem 4.1 below implies that $\mathrm{MixPo}(\Lambda_T)$ converges at rate of just a bit slower than $e^{-\frac{1}{2}\lambda T}$ to

MixPo(M) as $T \to \infty$, which is the asymptotic degree distribution already stated in Section 3.2 of [BL10]. Note that the theorem below is much more powerful as it gives an exact distance bound for finite T.

Theorem 4.1. *Let σ_S be the standard deviation of S. Then for $T \geq \frac{\log(2)}{\lambda}$, we have*

$$d_{TV}(\mathrm{MixPo}(\Lambda_T), \mathrm{MixPo}(\mathrm{M})) \leq \frac{\sqrt{32}\alpha}{\sqrt{\lambda}} \sigma_S \sqrt{T} e^{-\frac{1}{2}\lambda T} + 4\alpha \mathbb{E}(S)\left(T + \frac{\lambda}{\beta(\beta+\lambda)}\right) e^{-\lambda T}.$$

The main idea of the proof is as follows. Theorem 4.5 below is a simple but crucial result that allows to bound $d_{TV}(\mathrm{MixPo}(\Lambda_T), \mathrm{MixPo}(\mathrm{M}))$ by $\mathbb{E}(|\Lambda_T - \mathrm{M}|)$. In order to bound $\mathbb{E}(|\Lambda_T - \mathrm{M}|)$ further, we use that the expected value of the second summand of the right-hand side of (3.4) becomes small as $T \to \infty$ since the age $A_{Y_T}(T)$ of the youngest individual at time T converges quickly to 0, and compare the other summands of the right-hand sides of (3.4) and (4.1).

For the comparison of the last summands in (3.4) and (4.1), respectively, we note that the average of the social indices becomes close to $\mathbb{E}(S)$ by the Law of Large Numbers. Then we use again that the age $A_{Y_T}(T)$ of the youngest individual at time T converges quickly to 0 and that $A_{J_T}(T)$ converges quickly to A_{J_∞} by Proposition 2.7.

Finally, the expected absolute value of the difference between the first summand of (3.4) and the first summand of (4.1) again becomes small since $A_{J_T}(T)$ converges quickly to A_{J_∞}.

Remark 4.2. *For the original model of Britton and Lindholm with loops, we can adapt the proof of Theorem 4.1 in such a way that the upper bound remains exactly the same.*

4.2. The general case

Let $S_{J_\infty}(\omega) = S_{J_T}(\omega)$ as in the pure birth case. Furthermore, let $A_{J_\infty} = Z$, where Z is the random variable from Corollary 2.10. From the proof of this corollary (see [KS17], proof of Corollary 2.4), we obtain

$$A_{J_\infty}(\omega) = \mathbb{1}_{\{J_T(\omega)<Y_T(\omega)\}} F_\infty^{-1}(F_*(A_{J_T}(T,\omega)) + \mathbb{1}_{\{J_T(\omega)=Y_T(\omega)\}} \tilde{Z}(\omega),$$

where F_∞ is the cumulative distribution function of the $\mathrm{Exp}(\lambda)$ distribution as before, $\tilde{Z} \sim F_\infty$ independent of everything and $F_*(t) = 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1-e^{-(\lambda-\mu)T}}$. Note that, given $Y_T > 0$, we have $A_{J_\infty} \sim \mathrm{Exp}(\lambda)$.

Recall the definition of the parameter random variable Λ_T from (3.6) and let Λ_T^* be a random variable with $\mathcal{L}(\Lambda_T^*) = \mathcal{L}(\Lambda_T | Y_T > 0)$ as before. Moreover, set

$$\mathrm{M} = \frac{\alpha S_{J_\infty}}{\beta+\mu}(1 - e^{-(\beta+\mu)A_{J_\infty}}) + \frac{\alpha \mathbb{E}(S)}{\beta+\mu}(1 - e^{-(\beta+\mu)A_{J_\infty}}) \tag{4.2}$$

and let M^* be a random variable with $\mathcal{L}(\mathrm{M}^*) = \mathcal{L}(\mathrm{M}|Y_T > 0)$.

From Theorem 3.3, we know that $\mathrm{MixPo}(\Lambda_T^*)$ is the degree distribution at time T in the general case. Theorem 4.3 below implies that $\mathrm{MixPo}(\Lambda_T^*)$ converges at a rate of just a bit over $e^{-\frac{1}{6}(\lambda-\mu)T}$ to the $\mathrm{MixPo}(\mathrm{M}^*)$ distribution, which is the asymptotic degree distribution stated in Section 3.2 of [BL10] as $T \to \infty$. Note again that this theorem is much more powerful since it gives an exact bound for the total variation distance for finite T.

Theorem 4.3. *Let $\sigma_S < \infty$ be the standard deviation of S. Then for $T \geq \frac{2\log(4(\lambda(\lambda-\mu)^{-1}))}{\lambda-\mu}$, we have*

$$\begin{aligned}
&d_{TV}(\mathrm{MixPo}(\Lambda_T^*), \mathrm{MixPo}(\mathrm{M}^*)) \\
&\leq \alpha\left(\left(\frac{5\sqrt{6}}{2}\frac{\lambda}{\lambda-\mu} + \frac{2}{5} + \left(\beta + \frac{\mu}{2}\right)\left(\frac{229}{5(\lambda-\mu)} + \frac{2}{\lambda+\mu}\right)\right)\mathbb{E}(S) + \frac{27}{10}\sqrt{2}\sigma_S\right)(\lambda-\mu)T^2 e^{-\frac{1}{6}(\lambda-\mu)T} \\
&+ \alpha\beta\left(\left(\frac{1}{2} + \frac{\mu}{4\beta}\right)\frac{\mu}{\lambda} + \left(\frac{8}{\mu} + \frac{4}{\beta}\right)\frac{\lambda^3(\lambda+\mu)}{(\lambda-\mu)^3} + \left(6T + \frac{\sqrt{6}}{2} + \frac{3}{\beta}(\mu T + 3) + \frac{5\lambda}{\beta^2}\right)\lambda\right)\mathbb{E}(S)T^2 e^{-(\lambda-\mu)T}
\end{aligned}$$

$$+ 3\sqrt{2}\alpha\lambda\sigma_S T^2 e^{-(\lambda-\mu)T}.$$

Note that the right-hand side is of the order

$$O(T^2)e^{-\frac{1}{6}(\lambda-\mu)T}$$

as $T \to \infty$.

The main idea of the proof of this theorem is the same as in the pure birth case: We make use of Theorem 4.5 to obtain $\mathbb{E}(|\Lambda_T - \mathrm{M}|\ |Y_T > 0)$ as an upper bound for $d_{TV}(\mathrm{MixPo}(\Lambda_T^*), \mathrm{MixPo}(\mathrm{M}^*))$ and establish a further bound for this expected value. In order to do so, we use that the expected value of the first summand of the right-hand side of (3.6) converges quickly to zero since the time $T - T_{B_T+D_T}$ since the last event before T converges quickly to 0, and compare the remaining summand of the right-hand side of (3.6) with the right-hand side of (4.2). For this comparison, we make vital use of the fact that the average of the social indices of the nodes living at time T is close to $\mathbb{E}(S)$ by the Law of Large Numbers again and that, given T_l for some large $l \in \mathbb{N}$, the percentage of the nodes living at time T_l that survive up to time T is approximately $e^{-\mu(T-T_l)}$ by the Law of Large Numbers (see Lemma A.11 in the appendix).

A further important ingredient is that the reciprocal of the node process $(Y_t)_{t\geq 0}$ conditioned on survival is a supermartingale (see Lemma A.4 in the appendix), which makes it easy to deal with the expected value of its maximum. Finally, we also use that the age $A_{J_T}(T)$ of the randomly picked individual converges quickly to A_{J_∞} by Corollary 2.10.

The fact that nodes may die complicates the procedure considerably since the population size after a fixed number of events is random in this case and additional dependencies have to be treated (e.g. the inter-event times depend on the random population size at the previous event time).

In order to cope with additional dependencies on the random index J_T, we note that the probability that the node that is randomly picked is not older than $\frac{T}{2}$ at time T decreases exponentially in T (see Lemma A.6(i) in the appendix). Thus we essentially only have to consider the time interval $[\frac{T}{2}, \infty)$ instead of the interval $[T_{r(J_T)}, \infty)$, where $T_{r(J_T)}$ is the birth time of the randomly picked individual. The choice of $\frac{T}{2}$ as the left endpoint of the interval makes sure that we always have a large number of individuals in the time interval with high probability. We define the random number $\mathcal{K}(T)$ as index of the last event time before $\frac{T}{2}$ now such that $r(J_T) > \mathcal{K}(T)$ is equivalent to $T_{r(J_T)} \geq \frac{T}{2}$.

Notation 4.4. *Let* $\mathcal{K}(T) := \max\{k : T_k < \frac{T}{2}\}$*, so that we have* $Y_{T_{\mathcal{K}(T)}} = Y_{\frac{T}{2}}$ *almost surely.*

4.3. Proofs

For the proofs of both Theorem 4.1 and Theorem 4.3 we are interested in the total variation distance between two mixed Poisson distributions, say $\mathrm{MixPo}(\tilde{\Lambda})$ and the $\mathrm{MixPo}(\tilde{\mathrm{M}})$. Theorem 2.1 in [Yan91] gives us an upper bound for the total variation distance between two Poisson distributions. By conditioning we can generalize this result to mixed Poisson distributions.

Theorem 4.5. *Let* $\tilde{\Lambda}$ *and* $\tilde{\mathrm{M}}$ *be positive real valued random variables. Then we have for the total variation distance between the mixed Poisson distributions* $\mathrm{MixPo}(\tilde{\Lambda})$ *and* $\mathrm{MixPo}(\tilde{\mathrm{M}})$*:*

$$d_{TV}\big(\mathrm{MixPo}(\tilde{\Lambda}), \mathrm{MixPo}(\tilde{\mathrm{M}})\big) \leq \mathbb{E}\big(\min\big(\big|\sqrt{\tilde{\Lambda}} - \sqrt{\tilde{\mathrm{M}}}\big|, \big|\tilde{\Lambda} - \tilde{\mathrm{M}}\big|\big)\big).$$

Proof. Let X and Y be $\mathrm{MixPo}(\tilde{\Lambda})$ and $\mathrm{MixPo}(\tilde{\mathrm{M}})$ distributed, respectively. Then it follows

$$d_{TV}\big(\mathrm{MixPo}(\tilde{\Lambda}), \mathrm{MixPo}(\tilde{\mathrm{M}})\big) = \sup_{A\subset\mathbb{N}_0} \big|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)\big|$$

$$
\begin{aligned}
&= \sup_{A\subset\mathbb{N}_0} \big|\mathbb{E}(\mathbb{P}(X\in A|\tilde\Lambda)) - \mathbb{E}(\mathbb{P}(Y\in A|\tilde{\mathrm{M}}))\big| \\
&\le \sup_{A\subset\mathbb{N}_0} \mathbb{E}\big(|\mathbb{P}(X\in A|\tilde\Lambda) - \mathbb{P}(Y\in A|\tilde{\mathrm{M}})|\big) \\
&\le \mathbb{E}\bigg(\sup_{A\subset\mathbb{N}_0} |(\mathbb{P}(X\in A|\tilde\Lambda)) - (\mathbb{P}(Y\in A|\tilde{\mathrm{M}}))|\bigg) \\
&= \mathbb{E}\big(d_{TV}(\mathcal{L}(X|\tilde\Lambda), \mathcal{L}(Y|\tilde{\mathrm{M}})\big) \\
&\le \mathbb{E}\big(\min\big(\big|\sqrt{\tilde\Lambda} - \sqrt{\tilde{\mathrm{M}}}\big|, |\tilde\Lambda - \tilde{\mathrm{M}}|\big)\big),
\end{aligned}
$$

where the last line follows from Theorem 2.1 in [Yan91]. □

Remark 4.6. *Note that the mixed Poisson distribution depends on the parameter random variable only via its distribution; therefore Theorem 4.5 yields*

$$
d_{TV}(\mathrm{MixPo}(\tilde\Lambda), \mathrm{MixPo}(\tilde{\mathrm{M}})) \le \inf_{\substack{\hat\Lambda:\hat\Lambda \overset{\mathcal{D}}{=} \tilde\Lambda \\ \hat{\mathrm{M}}:\hat{\mathrm{M}} \overset{\mathcal{D}}{=} \tilde{\mathrm{M}}}} \mathbb{E}(\min(|\hat\Lambda^{\frac12} - \hat{\mathrm{M}}^{\frac12}|, |\hat\Lambda - \hat{\mathrm{M}}|).
$$

Proof of Theorem 4.1

From Theorem 4.5 follows

$$
d_{TV}(\mathrm{MixPo}(\Lambda_T), \mathrm{MixPo}(\mathrm{M})) \le \mathbb{E}(|\mathrm{M} - \Lambda_T|),
$$

and $\mathbb{E}(|\mathrm{M} - \Lambda_T|)$ is smaller than or equal to

$$
\begin{aligned}
&\mathbb{E}\bigg|\frac{\alpha}{\beta}\bigg(S_{J_\infty}(1-e^{-\beta A_{J_\infty}}) + \mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - S_{J_T}(1-e^{-\beta A_{\max(J_T,2)}(T)}) \\
&\qquad - \sum_{\substack{i=1\\ i\ne J_T}}^{Y_T} \frac{S_i}{Y_T-1}(1-e^{-\beta A_{Y_T}(T)}) - \sum_{\substack{i=1\\ i\ne J_T}}^{Y_T}\sum_{l=i\vee J_T}^{Y_T-1} \frac{S_i}{l-1}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\bigg)\bigg| \\
&\le \frac{\alpha}{\beta}\mathbb{E}\bigg|\sum_{\substack{i=1\\ i\ne J_T}}^{Y_T} \frac{S_i}{Y_T-1}(1-e^{-\beta A_{Y_T}(T)})\bigg| \\
&\quad + \frac{\alpha}{\beta}\mathbb{E}\bigg|\sum_{\substack{i=1\\ i\ne J_T}}^{Y_T}\sum_{l=i}^{Y_T-1} \frac{(S_i - \mathbb{E}(S))}{l-1}\mathbb{1}_{\{J_T\le l\}}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\bigg| \\
&\quad + \frac{\alpha}{\beta}\mathbb{E}\bigg|\mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1\\ i\ne J_T}}^{Y_T}\sum_{l=i\vee J_T}^{Y_T-1} \frac{\mathbb{E}(S)}{l-1}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}))\bigg| \\
&\quad + \frac{\alpha}{\beta}\mathbb{E}\bigg|S_{J_\infty}(1-e^{-\beta A_{J_\infty}}) - S_{J_T}(1-e^{-\beta A_{\max(J_T,2)}(T)})\bigg|,
\end{aligned}
\tag{4.3}
$$

where we use the convention $\frac{0}{0} := 0$.

For the first line of the right-hand side, we have

$$
\mathbb{E}\bigg|\sum_{\substack{i=1\\ i\ne J_T}}^{Y_T} \frac{S_i}{Y_T-1}(1-e^{-\beta A_{Y_T}(T)})\bigg| = \mathbb{E}\bigg(\mathbb{E}\bigg(\sum_{\substack{i=1\\ i\ne J_T}}^{Y_T} \frac{S_i}{Y_T-1}(1-e^{-\beta A_{Y_T}(T)})\bigg|Y_T, J_T\bigg)\bigg)
$$

$$= \mathbb{E}\biggl(\sum_{\substack{i=1\\ i\neq J_T}}^{Y_T} \frac{\mathbb{E}(S_i)}{Y_T-1}(1-\mathbb{E}(e^{-\beta A_{Y_T}(T)}|Y_T))\biggr), \tag{4.4}$$

where the last equality holds since, given Y_T, the age $A_{Y_T}(T)$ and J_T are independent.

Given $Y_T = y_T > 1$, the age $A_{y_T}(T)$ is the minimum of y_T-1 i.i.d. truncated exponentially distributed random variables by Proposition 2.7. Since the minimum of y_T-1 independent $\mathrm{Exp}(\lambda)$ distributed random variables is $\mathrm{Exp}((y_T-1)\lambda)$ distributed, the distribution of $A_{y_T}(T)$ is stochastically dominated by the $\mathrm{Exp}((y_T-1)\lambda)$ distribution. In general, we have for random variables $\hat{X}$ and $\hat{Y}$ with $\hat{X} \leq_{st} \hat{Y}$ that $f(\hat{X}) \leq_{st} f(\hat{Y})$ and consequently $\mathbb{E}(f(\hat{X})) \leq \mathbb{E}(f(\hat{Y}))$ for every increasing function f. Thus the right-hand side of (4.4) is smaller than or equal to

$$\mathbb{E}\biggl(\mathbb{E}(S)\frac{1}{\lambda}\frac{\beta}{Y_T-1+\frac{\beta}{\lambda}}\mathbb{1}_{\{Y_T>1\}}\biggr) \leq \frac{\beta}{\lambda}\mathbb{E}(S)\mathbb{E}\biggl(\frac{1}{Y_T-1}\mathbb{1}_{\{Y_T>1\}}\biggr).$$

Note that

$$\begin{aligned}\mathbb{E}\biggl(\frac{1}{Y_T-1}\mathbb{1}_{\{Y_T>1\}}\biggr) &= \sum_{n=2}^{\infty}\frac{p_n(T)}{n-1} = \frac{\lambda}{\lambda e^{\lambda T}}\sum_{n=2}^{\infty}\frac{1}{n-1}\biggl(\frac{\lambda e^{\lambda T}-\lambda}{\lambda e^{\lambda T}}\biggr)^{n-1} = \frac{\lambda}{\lambda e^{\lambda T}}\sum_{n=1}^{\infty}\frac{1}{n}\biggl(\frac{\lambda e^{\lambda T}-\lambda}{\lambda e^{\lambda T}}\biggr)^{n}\\ &\leq \frac{\lambda}{\lambda e^{\lambda T}-\lambda}\sum_{n=1}^{\infty}\frac{1}{n}\biggl(\frac{\lambda e^{\lambda T}-\lambda}{\lambda e^{\lambda T}}\biggr)^{n} = \sum_{n=1}^{\infty}\frac{p_n(T)}{n} = \mathbb{E}\biggl(\frac{1}{Y_T}\biggr),\end{aligned}$$

where $p_n(T)$ is the probability mass function from Section 2. Thus the right-hand side of (4.4) is smaller than or equal to

$$\frac{\beta}{\lambda}\mathbb{E}(S)\mathbb{E}\biggl(\frac{1}{Y_T}\biggr).$$

For the second line of the right-hand side of (4.3), we obtain

$$\begin{aligned}&\mathbb{E}\biggl|\sum_{\substack{i=1\\ i\neq J_T}}^{Y_T}\sum_{l=i}^{Y_T-1}\frac{(S_i-\mathbb{E}(S))}{l-1}\mathbb{1}_{\{J_T\leq l\}}(e^{-\beta A_{l+1}(T)}-e^{-\beta A_l(T)})\biggr|\\ &= \mathbb{E}\biggl|\sum_{l=2}^{Y_T-1}\frac{1}{l-1}\sum_{\substack{i=1\\ i\neq J_T}}^{l}(S_i-\mathbb{E}(S))\mathbb{1}_{\{J_T\leq l\}}(e^{-\beta A_{l+1}(T)}-e^{-\beta A_l(T)})\biggr|\\ &= \mathbb{E}\biggl(\mathbb{E}\biggl(\biggl|\sum_{l=J_T\vee 2}^{Y_T-1}\frac{1}{l-1}\sum_{\substack{i=1\\ i\neq J_T}}^{l}(S_i-\mathbb{E}(S))(e^{-\beta A_{l+1}(T)}-e^{-\beta A_l(T)})\biggr|\ \biggm|\ Y_T, J_T\biggr)\biggr)\\ &\leq \mathbb{E}\biggl(\sum_{l=J_T\vee 2}^{Y_T-1}\mathbb{E}\biggl|\frac{1}{l-1}\sum_{i=2}^{l}(S_i-\mathbb{E}(S))\biggr|\mathbb{E}\biggl(e^{-\beta A_{l+1}(T)}(1-e^{-\beta(A_l(T)-A_{l+1}(T))})|Y_T\biggr)\biggr)\\ &\leq \mathbb{E}\biggl(\sum_{l=J_T\vee 2}^{Y_T-1}\frac{1}{l-1}\sqrt{\sum_{i=2}^{l}\mathbb{E}\biggl((S_i-\mathbb{E}(S))^2\biggr)}\mathbb{E}(1-e^{-\beta(A_l(T)-A_{l+1}(T))}|Y_T)\biggr)\\ &\leq \mathbb{E}\biggl(\sum_{l=J_T\vee 2}^{Y_T-1}\frac{1}{\sqrt{l-1}}\sigma_S\mathbb{E}(1-e^{-\beta(A_l(T)-A_{l+1}(T))}|Y_T)\biggr),\end{aligned} \tag{4.5}$$

where the fourth line holds since the second sum in the third line is (stochastically) independent of J_T, the social indices are independent of all other random variables and, given Y_T, the index J_T is independent of the social indices and ages, and the second last line is obtained by applying $\mathbb{E}|Z-\mathbb{E}(Z)| \leq \sqrt{\mathrm{Var}(Z)}$ to $Z = \frac{1}{l-1}\sum_{i=2}^{l} S_i$.

Given Y_T and $A_{l+1}(T)$, the difference $A_l(T) - A_{l+1}(T)$ is the minimum of $l-1$ i.i.d. truncated exponentially distributed random variables by Proposition 2.7 (for $l \geq 2$). Thus it is stochastically dominated by an $\mathrm{Exp}((l-1)\lambda)$ distributed random variable Z_l, where Z_l can be assumed to independent of $A_{l+1}(T)$. This implies for $a > 0$:

$$\begin{aligned}\mathbb{P}(A_l(T) - A_{l+1}(T) \leq a | Y_T) &= \int_0^T \mathbb{P}(A_l(T) - A_{l+1}(T) \leq a | Y_T, T - A_{l+1}(T) = x)\mathbb{P}(T - A_{l+1}(T) \in dx | Y_T) \\ &\geq \mathbb{P}(Z_l \leq a | Y_T).\end{aligned}$$

Thus, also given Y_T alone, the difference $A_l(T) - A_{l+1}(T)$ is stochastically dominated by Z_l. By the same argument as for (4.4), it follows that (4.5) is smaller than or equal to

$$\begin{aligned}\mathbb{E}\bigg(\sum_{l=2}^{Y_T-1} \sigma_S \frac{1}{\sqrt{l-1}} \frac{\beta}{(l-1)\lambda+\beta} \mathbb{1}_{\{J_T \leq l\}}\bigg) &= \mathbb{E}\bigg(\sum_{l=2}^{Y_T-1} \sigma_S \frac{1}{\sqrt{l-1}} \frac{\beta}{(l-1)\lambda+\beta} \mathbb{E}(\mathbb{1}_{\{J_T \leq l\}}|Y_T)\bigg) \\ &\leq \mathbb{E}\bigg(\sum_{l=2}^{Y_T-1} \sigma_S \frac{1}{\sqrt{l-1}} \frac{\beta}{(l-1)\lambda} \frac{l}{Y_T}\bigg) \\ &\leq \mathbb{E}\bigg(\sum_{l=2}^{Y_T-1} \sigma_S \frac{1}{\sqrt{l-1}} \frac{\beta}{(l-1)\lambda} \frac{2(l-1)}{Y_T}\bigg) \\ &= \frac{2\beta}{\lambda}\sigma_S \mathbb{E}\bigg(\frac{1}{Y_T}\sum_{l=2}^{Y_T-1} \frac{1}{\sqrt{l-1}}\bigg) \\ &\leq \frac{2\beta}{\lambda}\sigma_S \mathbb{E}\bigg(\frac{1}{Y_T}\sum_{l=2}^{Y_T-1} \frac{2}{\sqrt{l-2}+\sqrt{l-1}}\bigg) \\ &= \frac{2\beta}{\lambda}\sigma_S \mathbb{E}\bigg(\frac{1}{Y_T}\sum_{l=2}^{Y_T-1} 2(\sqrt{l-1}-\sqrt{l-2})\bigg) \\ &= \frac{4\beta}{\lambda}\sigma_S \mathbb{E}\bigg(\frac{\sqrt{Y_T-2}}{Y_T}\bigg) \\ &\leq \frac{4\beta}{\lambda}\sigma_S \mathbb{E}\bigg(\frac{1}{\sqrt{Y_T}}\bigg).\end{aligned}$$

Since we arranged $A_{J_T}(T) \leq A_{J_\infty}$, for the third line of the right-hand side of (4.3) we obtain

$$\begin{aligned}&\mathbb{E}\bigg|\mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1\\ i\neq J_T}}^{Y_T} \sum_{l=i\vee J_T}^{Y_T-1} \frac{\mathbb{E}(S)}{l-1}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}))\bigg| \\ &= \mathbb{E}\bigg|\mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - \mathbb{E}(S)\sum_{l=J_T\vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{\substack{i=1\\ i\neq J_T}}^{l} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}))\bigg| \\ &= \mathbb{E}\bigg|\mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - \mathbb{E}(S)\sum_{l=J_T\vee 2}^{Y_T-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}))\bigg| \\ &= \mathbb{E}\bigg|\mathbb{E}(S)(1-e^{-\beta A_{J_\infty}}) - \mathbb{E}(S)(e^{-\beta A_{Y_T}(T)} - e^{-\beta A_{\max(J_T,2)}(T)}))\bigg| \\ &= \mathbb{E}\bigg(\mathbb{E}(S)(1-e^{-\beta A_{Y_T}(T)}) + \mathbb{E}(S)(e^{-\beta A_{\max(J_T,2)}(T)} - e^{-\beta A_{J_\infty}})\bigg)\end{aligned}$$

$$= \mathbb{E}(S)\mathbb{E}\left(1 - \mathbb{E}(e^{-\beta A_{Y_T}(T)}|Y_T)\right) + \mathbb{E}(S)(\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) - \mathbb{E}(e^{-\beta A_{J_\infty}})). \tag{4.6}$$

By Proposition 2.7

$$\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) = \int_0^T e^{-\beta x}\frac{\lambda e^{-\lambda x}}{1-e^{-\lambda T}}dx = \frac{\lambda}{1-e^{-\lambda T}}\int_0^T e^{-(\beta+\lambda)x}dx = \frac{\lambda}{\beta+\lambda}\frac{1-e^{-(\beta+\lambda)T}}{1-e^{-\lambda T}}, \tag{4.7}$$

whence follows that (4.6) is smaller than or equal to

$$\mathbb{E}(S)\mathbb{E}\left(\frac{\beta}{\lambda}\frac{1}{Y_t+\frac{\beta}{\lambda}}\right) + \mathbb{E}(S)\frac{\lambda}{\beta+\lambda}\left(\frac{1-e^{-(\beta+\lambda)T}}{1-e^{-\lambda T}} - 1\right) \le \frac{\beta}{\lambda}\mathbb{E}(S)\mathbb{E}\left(\frac{1}{Y_T}\right) + \mathbb{E}(S)\frac{\lambda}{\beta+\lambda}\frac{1}{e^{\lambda T}-1}.$$

Since we arranged $S_{J_T} = S_{J_\infty}$ and $A_{J_T}(T) \le A_{J_\infty}$, for the fourth line of the right-hand side of (4.3) we have

$$\mathbb{E}\left|S_{J_\infty}(1-e^{-\beta A_{J_\infty}}) - S_{J_T}(1-e^{-\beta A_{\max(J_T,2)}(T)})\right| = \mathbb{E}(S)(\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) - \mathbb{E}(e^{-\beta A_{J_\infty}})). \tag{4.8}$$

Note that the right-hand side of (4.8) is the same as the second summand of the right-hand side of (4.6), which is smaller than or equal to $\mathbb{E}(S)\frac{\lambda}{\beta+\mu}\frac{1}{e^{\lambda T}-1}$ (see above).

Altogether, we may conclude that the right-hand side of (4.3) is bounded from above by

$$\frac{2\alpha}{\lambda}\mathbb{E}(S)\mathbb{E}\left(\frac{1}{Y_T}\right) + \frac{4\alpha}{\lambda}\sigma_S\mathbb{E}\left(\frac{1}{\sqrt{Y_T}}\right) + \frac{2\alpha}{\beta}\mathbb{E}(S)\frac{\lambda}{\beta+\lambda}\frac{1}{e^{\lambda T}-1}.$$

Using the upper bounds for $\mathbb{E}(Y_T^{-1})$ and $\mathbb{E}(Y_T^{-1/2})$ from Propositions 2.3 and 2.4, we obtain the statement of the theorem. □

Proof of Theorem 4.3

In order to simplify notation, we introduce $\mathbb{E}^*(\,\cdot\,) = \mathbb{E}(\,\cdot\,|Y_T > 0)$, $\mathbb{P}^*(\,\cdot\,) = \mathbb{P}(\,\cdot\,|Y_T > 0)$ and the number of events $\mathcal{M}_T = B_T + D_T$ up to time T.

As before we use Theorem 4.5 to obtain

$$d_{TV}(\mathrm{MixPo}(\Lambda_T^*), \mathrm{MixPo}(\mathrm{M}^*)) \le \mathbb{E}^*|\Lambda_T - \mathrm{M}|.$$

In order to find an upper bound for $\mathbb{E}^*|\Lambda_T - \mathrm{M}|$, we use the triangle inequality for the absolute value after plugging in the definitions of Λ_T and M given by (3.6) and (4.2), which yields

$$\begin{aligned}\mathbb{E}^*\Bigg|&\frac{\alpha(1-e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T-1)\beta}\mathbb{1}_{\{Y_T>1\}}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}(S_i+S_{J_T})\mathbb{1}_{\{T_i^->T\}}\\ &+\frac{\alpha}{\beta}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}(S_i+S_{J_T})\mathbb{1}_{\{T_i^->T\}}\sum_{l=r(i)\vee r(J_T)}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\\ &-\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_\infty})(1-e^{-(\beta+\mu)A_{J_\infty}})\Bigg|.\end{aligned}$$

This is bounded from above by

$$\mathbb{E}^*\bigg(\frac{\alpha(1-e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T-1)\beta}\sum_{\substack{i=1\\i\neq J_T}}^{B_T} S_i \mathbb{1}_{\{T_i^->T\}}\mathbb{1}_{\{Y_T>1\}}\bigg) \tag{4.9}$$

$$+\,\mathbb{E}^*\bigg(\frac{\alpha(1-e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T-1)\beta}S_{J_T}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}\mathbb{1}_{\{T_i^->T\}}\mathbb{1}_{\{Y_T>1\}}\bigg) \tag{4.10}$$

$$+\,\mathbb{E}^*\bigg|\frac{\alpha}{\beta}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}(S_i-\mathbb{E}(S))\mathbb{1}_{\{T_i^->T\}}\sum_{l=r(i)\vee r(J_T)}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\bigg| \tag{4.11}$$

$$\begin{aligned}+\,\mathbb{E}^*\bigg|&\frac{\alpha}{\beta}\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\sum_{\substack{i=1\\i\neq J_T}}^{r^{-1}(l)}(\mathbb{E}(S)+S_{J_T})\mathbb{1}_{\{T_i^->T\}}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\\&-\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(e^{-(\beta+\mu)(T-T_{l+1})}-e^{-(\beta+\mu)(T-T_l)})\bigg|\end{aligned} \tag{4.12}$$

$$\begin{aligned}+\,\mathbb{E}^*\bigg|&\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(e^{-(\beta+\mu)(T-T_{l+1})}-e^{-(\beta+\mu)(T-T_l)})\\&-\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_\infty})(1-e^{-(\beta+\mu)A_{J_\infty}})\bigg|,\end{aligned} \tag{4.13}$$

where $r^{-1}(l)$ is the number of births that occur not later than the lth event for all $l\in\{1,\dots,\mathcal{M}_T\}$, i.e. $r^{-1}:\{1,\dots,\mathcal{M}_T\}\to\{1,\dots,B_T\}, l\mapsto\sum_{i=1}^{B_T}\mathbb{1}_{\{r(i)\le l\}}$.

In the following, we deduce upper bounds for (4.9)–(4.13).

Upper bound for (4.9) and (4.10) We treat (4.9) similarly to the corresponding expression in the pure birth case, but condition on B_T, D_T, J_T and the information which nodes survive up to time T, and obtain

$$\mathbb{E}^*\bigg(\frac{\alpha(1-e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T-1)\beta}\mathbb{1}_{\{Y_T>1\}}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}S_i\mathbb{1}_{\{T_i^->T\}}\bigg)\le\frac{\alpha}{\beta}\mathbb{E}(S)\mathbb{E}^*(1-e^{-\beta(T-T_{\mathcal{M}_T})}). \tag{4.14}$$

Since we condition on J_T, the same expression is also obtained for (4.10).

For $T\ge\frac{1}{\lambda-\mu}\log(2)$, we have that the expectation $\mathbb{E}^*(1-e^{-\beta(T-T_{\mathcal{M}_T})})$ is smaller than or equal to

$$\frac{2}{\lambda}\bigg(\lambda-\mu+\beta\Big(\log\Big(\frac{\lambda}{\lambda-\mu}\Big)+(\lambda-\mu)T\Big)\bigg)e^{-(\lambda-\mu)T} \tag{4.15}$$

by Lemma A.7. Plugging this expression in the right-hand side of (4.14) results in the following upper bound for the sum of (4.9) and (4.10):

$$\frac{4\alpha\mathbb{E}(S)}{\beta\lambda}\bigg(\lambda-\mu+\beta\Big(\log\Big(\frac{\lambda}{\lambda-\mu}\Big)+(\lambda-\mu)T\Big)\bigg)e^{-(\lambda-\mu)T}.$$

Upper bound for (4.11) Let $R_{T_l,T}$ be the number of nodes that are alive at time T_l and survive up to time T. We compute

$$\mathbb{E}^*\bigg(\bigg|\frac{\alpha}{\beta}\sum_{\substack{i=1\\i\neq J_T}}^{B_T}(S_i-\mathbb{E}(S))\mathbb{1}_{\{T_i^->T\}}\sum_{l=r(i)\vee r(J_T)}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\bigg|\bigg)$$

$$
\begin{aligned}
&\le \mathbb{E}^*\Bigg(\Bigg|\frac{\alpha}{\beta}\sum_{\substack{l=r(J_T)\\ R_{T_l,T}\ge 2}}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{R_{T_l,T}-1}{Y_{T_l}-1}\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{R_{T_l,T}-1}\sum_{\substack{i=1\\ i\neq J_T}}^{r^{-1}(l)}(S_i-\mathbb{E}(S))\mathbb{1}_{\{T_i^->T\}}\Bigg|\Bigg)\\
&\le \mathbb{E}^*\Bigg(\frac{\alpha}{\beta}\sum_{\substack{l=r(J_T)\\ R_{T_l,T}\ge 2}}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\\
&\qquad\cdot\Bigg(\frac{1}{R_{T_l,T}-1}\Bigg|\sum_{\substack{i=1\\ i\neq J_T}}^{r^{-1}(l)}(S_i-\mathbb{E}(S))\mathbb{1}_{\{T_i^->T\}}\Bigg|\;\Bigg|\;(Y_t)_{0\le t\le T},J_T\Bigg)\Bigg).
\end{aligned}
\tag{4.16}
$$

Note in the second line that we can restrict the sum to $R_{T_l,T}\ge 2$ because $R_{T_l,T}\ge 1$ by $l\ge r(J_T)$ and a summand with $R_{T_l,T}=1$ in the first line would be 0 anyway.

Since the sum in the last line of (4.16) has exactly $R_{T_l,T}-1$ summands of the form $S_i-\mathbb{E}(S)$ and all other summands are zero, the right-hand side of (4.16) is smaller than or equal to

$$
\mathbb{E}^*\Bigg(\Bigg|\frac{\alpha}{\beta}\sum_{\substack{l=r(J_T)\\ R_{T_l,T}\ge 2}}^{\mathcal{M}_T-1}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\frac{\sigma_S}{\sqrt{R_{T_l,T}-1}}\Bigg|\Bigg).
\tag{4.17}
$$

Since $e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)}\le\beta(T_{l+1}-T_l)$ and $R_{T_l,T}\le Y_{T_l}$, the expression (4.17) is smaller than or equal to

$$
\begin{aligned}
\mathbb{E}^*\Bigg(\alpha\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\frac{\sigma_S\mathbb{1}_{\{Y_{T_l}>1\}}}{\sqrt{Y_{T_l}-1}}(T_{l+1}-T_l)\Bigg)&\le\mathbb{E}^*\Bigg(\alpha\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\frac{\sigma_S}{\sqrt{Y_{T_l}-\frac{Y_{T_l}}{2}}}\mathbb{1}_{\{Y_{T_l}>1\}}(T_{l+1}-T_l)\Bigg)\\
&\le\sqrt{2}\alpha\sigma_S T\mathbb{E}^*\Bigg(\max_{r(J_T)\le k\le\mathcal{M}_T-1}\frac{1}{\sqrt{Y_{T_k}}}\Bigg)\\
&\le\sqrt{2}\alpha\sigma_S T\mathbb{E}^*\Bigg(\mathbb{1}_{\{\mathcal{K}(T)<r(J_T)\}}\max_{r(J_T)\le k\le\mathcal{M}_T-1}\frac{1}{\sqrt{Y_{T_k}}}\Bigg)\\
&\quad+\sqrt{2}\alpha\sigma_S T\mathbb{E}^*\Bigg(\mathbb{1}_{\{\mathcal{K}(T)\ge r(J_T)\}}\max_{r(J_T)\le k\le\mathcal{M}_T-1}\frac{1}{\sqrt{Y_{T_k}}}\Bigg),
\end{aligned}
\tag{4.18}
$$

where $\mathcal{K}(T)=\max\{k:T_k\le\frac{T}{2}\}$ (see Definition 4.4). We have

$$
\begin{aligned}
&\mathbb{E}^*\Bigg(\mathbb{1}_{\{\mathcal{K}(T)<r(J_T)\}}\max_{r(J_T)\le k\le\mathcal{M}_T-1}\frac{1}{\sqrt{Y_{T_k}}}\Bigg)\\
&\le\mathbb{E}\Bigg(\mathbb{1}_{\{\mathcal{K}(T)<r(J_T)\}}\max_{r(J_T)\le k\le\mathcal{M}_T-1}\frac{1}{\sqrt{Y_{T_k}}}\;\Bigg|\;Y_\infty>0\Bigg)\mathbb{P}^*(Y_\infty>0)+1\cdot\mathbb{P}^*(Y_\infty=0),
\end{aligned}
\tag{4.19}
$$

where $Y_\infty:=\lim_{t\to\infty}Y_t\in\{0,\infty\}$.

By Lemma A.1, we have for the second summand of the right-hand side of (4.19)

$$
\mathbb{P}^*(Y_\infty=0)=\frac{\mu}{\lambda}e^{-(\lambda-\mu)T}.
$$

The first summand of the right-hand side of (4.19) is smaller than or equal to

$$
\begin{aligned}
\mathbb{E}\Bigg(\max_{\mathcal{K}(T)<k}\frac{1}{\sqrt{Y_{T_k}}}\;\Bigg|\;Y_\infty>0\Bigg)&=\mathbb{E}\Bigg(\max_{T/2<t}\frac{1}{\sqrt{Y_t}}\;\Bigg|\;Y_\infty>0\Bigg)\\
&\le e^{-\frac{1}{6}(\lambda-\mu)T}+\mathbb{E}\Bigg(\frac{1}{Y_{T/2}}\;\Bigg|\;Y_\infty>0\Bigg)e^{\frac{1}{3}(\lambda-\mu)T}
\end{aligned}
$$

$$\leq \left(1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)e^{-\frac{1}{6}(\lambda-\mu)T} \tag{4.20}$$

for $T \geq \frac{2}{\lambda-\mu}\log(2)$ by combining Lemma A.8, where $\delta = \frac{1}{2}$ and $\gamma = \frac{1}{6}$, and Lemma A.9. Thus for $T \geq \frac{2}{\lambda-\mu}\log(2)$, the first summand of the right-hand side of (4.18) is smaller than or equal to

$$\sqrt{2}\alpha\frac{\mu}{\lambda}\sigma_S T e^{-(\lambda-\mu)T} + \sqrt{2}\alpha\sigma_S\left(1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)Te^{-\frac{1}{6}(\lambda-\mu)T}. \tag{4.21}$$

By Lemma A.6(i), we obtain that for $T \geq \frac{1}{\lambda-\mu}\log(2)$, the second summand of the right-hand side of (4.18) is bounded from above by

$$\sqrt{2}\alpha\sigma_S T\left(e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)e^{-(\lambda-\mu)T}\right). \tag{4.22}$$

Thus for $T \geq \frac{2}{\lambda-\mu}\log(2)$, the expression (4.11) is bounded from above by the sum of (4.21) and (4.22).

Upper bound for (4.12) For (4.12), we have

$$\begin{aligned}
&\mathbb{E}^*\Bigg|\frac{\alpha}{\beta}\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}\sum_{\substack{i=1\\ i\neq J_T}}^{r^{-1}(l)}(\mathbb{E}(S)+S_{J_T})\mathbb{1}_{\{T_i^->T\}}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\\
&\qquad -\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(e^{-(\beta+\mu)(T-T_{l+1})}-e^{-(\beta+\mu)(T-T_l)})\Bigg|\\
&=\mathbb{E}^*\Bigg|\frac{\alpha(\beta+\mu)}{\beta(\beta+\mu)}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}(e^{-\beta(T-T_{l+1})}-e^{-\beta(T-T_l)})\\
&\qquad -\frac{\alpha\beta}{\beta(\beta+\mu)}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(e^{-(\beta+\mu)(T-T_{l+1})}-e^{-(\beta+\mu)(T-T_l)})\Bigg|
\end{aligned} \tag{4.23}$$

since the second sum on the left-hand side of (4.23) has exactly $R_{T_l,T}-1$ non-zero summands. The right-hand side of (4.23) is equal to

$$\begin{aligned}
&\mathbb{E}^*\Bigg|\frac{\alpha}{\beta(\beta+\mu)}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\Bigg((\beta+\mu)\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\beta(T-T_{l+1})}(1-e^{-\beta(T_{l+1}-T_l)})\\
&\qquad -\beta e^{-(\beta+\mu)(T-T_{l+1})}(1-e^{-(\beta+\mu)(T_{l+1}-T_l)})\Bigg)\Bigg|.
\end{aligned} \tag{4.24}$$

Now we use

$$1-e^{-\beta(T_{l+1}-T_l)} = \beta(T_{l+1}-T_l) - \sum_{k=2}^{\infty}\frac{(-\beta(T_{l+1}-T_l))^k}{k!}$$

and

$$1-e^{-(\beta+\mu)(T_{l+1}-T_l)} = (\beta+\mu)(T_{l+1}-T_l) - \sum_{k=2}^{\infty}\frac{(-(\beta+\mu)(T_{l+1}-T_l))^k}{k!}$$

to obtain that (4.24) is smaller than or equal to

$$\mathbb{E}^*\Bigg|\frac{\alpha(\mathbb{E}(S)+S_{J_T})}{\beta(\beta+\mu)}\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(\beta+\mu)\beta\left(\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\right)e^{-\beta(T-T_{l+1})}(T_{l+1}-T_l)\Bigg|$$

$$
\begin{aligned}
&+ \mathbb{E}^*\bigg|\frac{\alpha}{\beta(\beta+\mu)}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\bigg(\beta e^{-(\beta+\mu)(T-T_{l+1})}\sum_{k=2}^{\infty}\frac{(-(\beta+\mu)(T_{l+1}-T_l))^k}{k!} \\
&\qquad\qquad - (\beta+\mu)\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\beta(T-T_{l+1})}\sum_{k=2}^{\infty}\frac{(-\beta(T_{l+1}-T_l))^k}{k!}\bigg)\bigg| \\
&\le \mathbb{E}^*\bigg(\alpha(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\infty}\mathbb{1}_{\{T_{l+1}\le T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\bigg) \\
&\quad + \mathbb{E}^*\bigg(\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\sum_{k=2}^{\infty}\frac{(-(\beta+\mu)(T_{l+1}-T_l))^k}{k!}\bigg) \\
&\quad + \mathbb{E}^*\bigg(\frac{\alpha}{\beta}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\sum_{k=2}^{\infty}\frac{(-\beta(T_{l+1}-T_l))^k}{k!}\bigg)\bigg). \qquad (4.25)
\end{aligned}
$$

Firstly, we consider the first summand of the right-hand side of (4.25). Since the social index S_{J_T} is independent of all other random variables appearing in (4.25), this summand is equal to

$$
2\alpha\mathbb{E}(S)\mathbb{E}^*\bigg(\sum_{l=r(J_T)}^{\infty}\mathbb{1}_{\{T_{l+1}\le T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\bigg).
$$

We derive

$$
\begin{aligned}
&\mathbb{E}^*\bigg(\sum_{l=r(J_T)}^{\infty}\mathbb{1}_{\{T_{l+1}<T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\bigg) \\
&\le \mathbb{E}^*\bigg(\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{1}_{\{T_{l+1}<T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\bigg) \\
&\quad + T\mathbb{P}^*(\mathcal{K}(T)\ge r(J_T)). \qquad (4.26)
\end{aligned}
$$

For $\frac{1}{\lambda-\mu}\log(2)$, the second summand of the right-hand side of (4.26) is smaller than or equal to

$$
Te^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\bigg(\log\Big(\frac{\lambda}{\lambda-\mu}\Big)+(\lambda-\mu)T\bigg)Te^{-(\lambda-\mu)T}
$$

by Lemma A.6(i). The first summand of the right-hand side of (4.26) is equal to

$$
\mathbb{E}^*\bigg(\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\bigg(\mathbb{1}_{\{T_{l+1}<T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\;\bigg|\;\mathcal{K}(T),Y_T>0\bigg)\bigg).
$$

For the inner expectation, we have

$$
\begin{aligned}
&\mathbb{E}\bigg(\mathbb{1}_{\{T_{l+1}<T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\;\bigg|\;\mathcal{K}(T),Y_T>0\bigg) \\
&\le \mathbb{E}\bigg(\mathbb{1}_{\{T_{l+1}<T\}}\bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\bigg|(T_{l+1}-T_l)\;\bigg|\;\mathcal{K}(T),Y_{T_l}>0\bigg)\frac{\mathbb{P}(Y_{T_l}>0|\mathcal{K}(T))}{\mathbb{P}(Y_T>0|\mathcal{K}(T))}. \qquad (4.27)
\end{aligned}
$$

For the fraction, we obtain for $l\ge\mathcal{K}(T)$

$$
\frac{\mathbb{P}(Y_{T_l}>0|\mathcal{K}(T))}{\mathbb{P}(Y_T>0|\mathcal{K}(T))}\le\frac{\mathbb{P}(Y_{T_{\mathcal{K}(T)}}>0|\mathcal{K}(T))}{\mathbb{E}(\mathbb{P}(Y_T>0|\mathcal{K}(T),Y_{T_{\mathcal{K}(T)}})|\mathcal{K}(T))}. \qquad (4.28)
$$

Note that, given $\mathcal{K}(T)$ and $Y_{T_{\mathcal{K}(T)}}$, we know that exactly $Y_{T_{\mathcal{K}(T)}}$ nodes are alive at time $\frac{T}{2}$. Thus due to the Markov property for $(Y_t)_{t\geq 0}$ and the formula for the extinction probability of a linear birth and death process with a general initial value given in Remark 2.1, the conditional probability $\mathbb{P}(Y_T > 0|\mathcal{K}(T), Y_{T_{\mathcal{K}(T)}})$ is equal to $p_0(\frac{T}{2})^{Y_{T_{\mathcal{K}(T)}}}$, which implies that (4.28) is equal to

$$\frac{\mathbb{P}(Y_{T_{\mathcal{K}(T)}} > 0|\mathcal{K}(T))}{\mathbb{E}(1-p_0(\frac{1}{2}T)^{Y_{T_{\mathcal{K}(T)}}}|\mathcal{K}(T))} \leq \frac{\mathbb{P}(Y_{T_{\mathcal{K}(T)}} > 0|\mathcal{K}(T))}{(1-p_0(\frac{1}{2}T))\mathbb{P}(Y_{T_{\mathcal{K}(T)}} > 0|\mathcal{K}(T))} = \frac{\lambda e^{\frac{1}{2}(\lambda-\mu)T}-\mu}{(\lambda-\mu)e^{\frac{1}{2}(\lambda-\mu)T}} \leq \frac{\lambda}{\lambda-\mu}.$$

Thus the right-hand side of (4.27) is smaller than or equal to

$$\mathbb{E}\Bigg(\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}\Bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\Bigg|(T_{l+1}-T_l)\;\Bigg|\,Y_{T_l},T_l,\mathcal{K}(T)\Bigg)\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\cdot\frac{\lambda}{\lambda-\mu}. \tag{4.29}$$

By the Markov property of $(Y_t)_{t\geq 0}$, we may omit the conditioning on $\mathcal{K}(T)$ if we condition on Y_{T_l} and T_l for $l \geq \mathcal{K}(T)+1$. Conditioning in addition on T_{l+1}, we see that (4.29) is equal to

$$\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)\mathbb{E}\Bigg(\Bigg|\frac{R_{T_l,T}-1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-e^{-\mu(T-T_{l+1})}\Bigg|\;\Bigg|\,Y_{T_l},T_l,T_{l+1}\Bigg)\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\cdot\frac{\lambda}{\lambda-\mu}. \tag{4.30}$$

By Lemma A.11, summing over l and taking the expectation yields the following upper bound for the first summand of the right-hand side of (4.26):

$$\mathbb{E}^*\Bigg(\sum_{l=\mathcal{K}(T)+1}^{\infty}\sqrt{6}\frac{\lambda}{\lambda-\mu}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg). \tag{4.31}$$

Obviously, (4.31) is equal to

$$\begin{aligned}&\sqrt{6}\frac{\lambda}{\lambda-\mu}\mathbb{E}^*\Bigg(\mathbb{1}_{\{Y_\infty>0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg)\\&+\sqrt{6}\frac{\lambda}{\lambda-\mu}\mathbb{E}^*\Bigg(\mathbb{1}_{\{Y_\infty=0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg).\end{aligned}$$

We may conclude that for $T \geq \frac{1}{\lambda-\mu}\log(2)$, the first summand of (4.25) is bounded from above by

$$\begin{aligned}2\alpha\mathbb{E}(S)\Bigg(&\sqrt{6}\frac{\lambda}{\lambda-\mu}\mathbb{E}^*\Bigg(\mathbb{1}_{\{Y_\infty>0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg)\\&+\sqrt{6}\frac{\lambda}{\lambda-\mu}\mathbb{E}^*\Bigg(\mathbb{1}_{\{Y_\infty=0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg)\\&+Te^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\Bigg(\log\Bigg(\frac{\lambda}{\lambda-\mu}\Bigg)+(\lambda-\mu)T\Bigg)Te^{-(\lambda-\mu)T}\Bigg).\end{aligned} \tag{4.32}$$

For the first outer expectation in (4.32), we have

$$\mathbb{E}^*\Bigg(\mathbb{1}_{\{Y_\infty>0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\Bigg(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\;\Bigg|\,\mathcal{K}(T),Y_{T_l}>0\Bigg)\Bigg)$$

$$
\begin{aligned}
&\leq \mathbb{E}\biggl(\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\biggl(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\biggm|\mathcal{K}(T)\biggr)\biggm|Y_\infty>0\biggr)\\
&\leq \mathbb{E}\biggl(\max_{\mathcal{K}(T)\leq k}Y_{T_k}^{-\frac{1}{2}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)\biggm|Y_\infty>0\biggr)\\
&\leq \frac{T}{2}\mathbb{E}\biggl(\max_{\mathcal{K}(T)\leq k}Y_{T_k}^{-\frac{1}{2}}\biggm|Y_\infty>0\biggr).
\end{aligned}
$$

By inequality (4.20), we obtain that for $T\geq\frac{2}{\lambda-\mu}\log(2)$, this expression is smaller than or equal to

$$
\frac{T}{2}\biggl(1+2\log\Bigl(\frac{\lambda}{\lambda-\mu}\Bigr)+(\lambda-\mu)T\biggr)e^{-\frac{1}{6}(\lambda-\mu)T}.
$$

For the second outer expectation in (4.32), we obtain

$$
\begin{aligned}
&\mathbb{E}\biggl(\mathbb{1}_{\{Y_\infty=0\}}\sum_{l=\mathcal{K}(T)+1}^{\infty}\mathbb{E}\biggl(\mathbb{1}_{\{T_{l+1}<T\}}(T_{l+1}-T_l)Y_{T_l}^{-\frac{1}{2}}\biggm|\mathcal{K}(T),Y_{T_l}>0\biggr)\biggm|Y_T>0\biggr)\\
&\leq\frac{T}{2}\mathbb{P}(Y_\infty=0|Y_T>0)=\frac{T}{2}\frac{\mu}{\lambda}e^{-(\lambda-\mu)T},
\end{aligned}
$$

where the last equality follows from Lemma A.1.

In conclusion, for $T\geq\frac{2}{\lambda-\mu}\log(2)$, the first summand of the right-hand side of (4.25) is bounded from above by

$$
\begin{aligned}
&\sqrt{6}\alpha\frac{\lambda}{\lambda-\mu}\mathbb{E}(S)T\Biggl(\biggl(1+2\log\Bigl(\frac{\lambda}{\lambda-\mu}\Bigr)+(\lambda-\mu)T\biggr)e^{-\frac{1}{6}(\lambda-\mu)T}+e^{-(\lambda-\mu)T}\Biggr)\\
&+2\alpha\mathbb{E}(S)\Biggl(Te^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\biggl(\log\Bigl(\frac{\lambda}{\lambda-\mu}\Bigr)+(\lambda-\mu)T\biggr)Te^{-(\lambda-\mu)T}\Biggr).
\end{aligned}
$$

Since the social index S_{J_T} is independent of all other random variables appearing in (4.25), the second summand of the right-hand side of (4.25) is smaller than or equal to

$$
\frac{2\alpha}{\beta+\mu}\mathbb{E}(S)\mathbb{E}^*\biggl(\sum_{l=r(J_T)}^{\mathcal{M}_T-1}\sum_{k=2}^{\infty}\frac{(-(\beta+\mu)(T_{l+1}-T_l))^k}{k!}\biggr).
$$

Since $\sum_{k=2}^{\infty}\frac{(-x)^k}{k!}\leq\frac{x^2}{2}$ for $x\geq0$, this expression is bounded from above by

$$
\frac{\alpha}{\beta+\mu}\mathbb{E}(S)\mathbb{E}^*\biggl(\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(\beta+\mu)^2(T_{l+1}-T_l)^2\biggr)\biggr). \tag{4.33}
$$

For $T\geq(\frac{2}{\lambda-\mu}\log(2)\vee\frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$, Lemma A.12 implies that (4.33) and hence also the second summand of the right-hand side of (4.25) is smaller than or equal to

$$
\begin{aligned}
\alpha(\beta+\mu)\mathbb{E}(S)\Biggl(&\frac{\mu}{\lambda}\frac{T^2}{4}e^{-(\lambda-\mu)T}+\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}T^2e^{-(\lambda+\mu)T}+T^2e^{-\frac{1}{2}\lambda T}\\
&+\frac{2(\lambda-\mu)}{\lambda}\biggl(\log\Bigl(\frac{\lambda}{\lambda-\mu}\Bigr)+(\lambda-\mu)T\biggr)T^2e^{-(\lambda-\mu)T}\\
&+\Bigl(\frac{3}{4}T^2+\frac{T}{2(\lambda+\mu)}\Bigr)\biggl(1+2\log\Bigl(\frac{\lambda}{\lambda-\mu}\Bigr)+(\lambda-\mu)T\biggr)e^{-\frac{1}{4}(\lambda-\mu)T}\Biggr).
\end{aligned}
$$

For the third summand of the right-hand side of (4.25), we obviously obtain the same upper bound except that the factor $\beta+\mu$ is replaced by β.

We may conclude that for $T \geq (\frac{2}{\lambda-\mu}\log(2) \vee \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$, the expression (4.12) is smaller than or equal to

$$\begin{aligned}
&\sqrt{6}\alpha\frac{\lambda}{\lambda-\mu}\mathbb{E}(S)T\left(\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-\frac{1}{6}(\lambda-\mu)T}+e^{-(\lambda-\mu)T}\right)\\
&+2\alpha\mathbb{E}(S)\left(Te^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)Te^{-(\lambda-\mu)T}\right)\\
&+\alpha(2\beta+\mu)\mathbb{E}(S)\left(\frac{\mu}{\lambda}\frac{T^2}{4}e^{-(\lambda-\mu)T}+\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}T^2e^{-(\lambda+\mu)T}+T^2e^{-\frac{1}{2}\lambda T}\right.\\
&\qquad+\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)T^2e^{-(\lambda-\mu)T}\\
&\qquad\left.+\left(\frac{3}{4}T^2+\frac{T}{2(\lambda+\mu)}\right)\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-\frac{1}{4}(\lambda-\mu)T}\right).
\end{aligned}$$

Upper bound for (4.13) Recall that we arranged $S_{J_T}=S_{J_\infty}$. Since the sum in (4.13) telescopes, we obtain

$$\begin{aligned}
&\mathbb{E}^*\left(\left|\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_T})\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(e^{-(\beta+\mu)(T-T_{l+1})}-e^{-(\beta+\mu)(T-T_l)})\right.\right.\\
&\qquad\left.\left.-\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_\infty})(1-e^{-(\beta+\mu)A_{J_\infty}})\right|\right)\\
&=\mathbb{E}^*\left(\left|\frac{\alpha}{\beta+\mu}(\mathbb{E}(S)+S_{J_\infty})\left(e^{-(\beta+\mu)(T-T_{\mathcal{M}_T})}-e^{-(\beta+\mu)(T-T_{r(J_T)})}-(1-e^{-(\beta+\mu)A_{J_\infty}})\right)\right|\right)\\
&\leq\frac{2\alpha}{\beta+\mu}\mathbb{E}(S)\left(\mathbb{E}^*\left(1-e^{-(\beta+\mu)(T-T_{\mathcal{M}_T})}\right)+\mathbb{E}^*\left(\left|e^{-(\beta+\mu)A_{J_T}(T)}-e^{-(\beta+\mu)A_{J_\infty}}\right|\right)\right),
\end{aligned}\tag{4.34}$$

where the last inequality holds since S_{J_∞} is independent of all other random variables appearing in (4.34).

By using Lemma A.7 again, the first conditional expectation on the right-hand side of (4.34) can be bounded from above by

$$\frac{2}{\lambda}\left(\lambda-\mu+(\beta+\mu)\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)\right)e^{-(\lambda-\mu)T}$$

if $T \geq \frac{1}{\lambda-\mu}\log(2)$.

Corollary 2.10 implies that the second conditional expectation on the right-hand side of (4.34) is bounded from above by

$$\left(\frac{2\lambda}{\beta+\mu+\lambda}+\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)\right)e^{-(\lambda-\mu)T}$$

if $T \geq \frac{1}{\lambda-\mu}\log(2)$.

Altogether, we obtain that the right-hand side of (4.34) is smaller than or equal to

$$\frac{4\alpha}{\beta+\mu}\mathbb{E}(S)\left(\frac{\lambda-\mu}{\lambda}+\frac{\beta+\lambda}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)+\frac{\lambda}{\beta+\mu+\lambda}\right)e^{-(\lambda-\mu)T}$$

if $T \geq \frac{1}{\lambda-\mu}\log(2)$.

Conclusion Combining the upper bounds obtained for (4.9)–(4.13), we have

$$
\begin{aligned}
&d_{TV}(\mathrm{MixPo}(\Lambda_T^*), \mathrm{MixPo}(\mathrm{M}^*)) \\
&\le \frac{4\alpha\mathbb{E}(S)}{\beta\lambda}\left(\lambda-\mu+\beta\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)\right)e^{-(\lambda-\mu)T} \\
&+\sqrt{2}\alpha\frac{\mu}{\lambda}\sigma_S T e^{-(\lambda-\mu)T}+\sqrt{2}\alpha\sigma_S\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)Te^{-\frac{1}{6}(\lambda-\mu)T} \\
&+\sqrt{2}\alpha\sigma_S T\left(e^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-(\lambda-\mu)T}\right) \\
&\sqrt{6}\alpha\frac{\lambda}{\lambda-\mu}\mathbb{E}(S)T\left(\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-\frac{1}{6}(\lambda-\mu)T}+e^{-(\lambda-\mu)T}\right) \\
&+2\alpha\mathbb{E}(S)\left(Te^{-\frac{1}{2}\lambda T}+\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)Te^{-(\lambda-\mu)T}\right) \\
&+\alpha(2\beta+\mu)\mathbb{E}(S)\left(\left(\frac{3}{4}T^2+\frac{T}{2(\lambda+\mu)}\right)\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-\frac{1}{4}(\lambda-\mu)T}+\frac{T^2}{4}\frac{\mu}{\lambda}e^{-(\lambda-\mu)T}\right) \\
&+\alpha(2\beta+\mu)\mathbb{E}(S)T^2\left(\frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-(\lambda-\mu)T}+\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}e^{-(\lambda+\mu)T}+e^{-\frac{1}{2}\lambda T}\right) \\
&+\frac{4\alpha}{\beta+\mu}\mathbb{E}(S)\left(\frac{\lambda-\mu}{\lambda}+\frac{\beta+\lambda}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)+\frac{\lambda}{\beta+\mu+\lambda}\right)e^{-(\lambda-\mu)T}. \qquad (4.35)
\end{aligned}
$$

for $T \ge (\frac{2}{\lambda-\mu}\log(2) \vee \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$. The upper bound claimed is now obtained by elementary computations; see the last part of the proof of [KS15, Theorem 5.4] for details. □

Appendix A1: Lemmas for the proof of Theorem 4.3

Some lemmas of general interest

Recall that we set $\mathbb{E}^*(\,\cdot\,) = \mathbb{E}(\,\cdot\,|Y_T > 0)$ and $\mathbb{P}^*(\,\cdot\,) = \mathbb{P}(\,\cdot\,|Y_T > 0)$. We start with a number of results about the linear birth and death process $(Y_t)_{t\ge 0}$ that could well be useful in other situations. We first compute the extinction probability given the process has survived up to time T. We write $Y_\infty = \lim_{t\to\infty} Y_t \in \{0,\infty\}$ a.s.

Lemma A.1. *For the conditioned extinction probability given $Y_T > 0$, we have*

$$\mathbb{P}^*(Y_\infty = 0) = \frac{\mu}{\lambda}e^{-(\lambda-\mu)T}.$$

Proof: By conditioning on the population size, we obtain

$$\mathbb{P}^*(Y_\infty = 0) = \mathbb{E}^*(\mathbb{P}(Y_\infty = 0|Y_T)) = \mathbb{E}^*\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right) \qquad (A.1)$$

since $(\mu/\lambda)^m$ is the extinction probability of a linear birth and death process with birth rate λ, death rate μ and initial value m (see Remark 2.1).

On the one hand, we have

$$\mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right) = \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\,\middle|\, Y_T > 0\right)\mathbb{P}(Y_T > 0) + \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\,\middle|\, Y_T = 0\right)\mathbb{P}(Y_T = 0)$$

$$= \mathbb{E}^*\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right)(1-p_0(T)) + p_0(T).$$

On the other hand, we can make use of the known formula for the probability generating function of Y_T (see III.5 in [AN72]) to obtain

$$\mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right) = \frac{\mu}{\lambda}.$$

This yields

$$\mathbb{E}^*\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right) = \left(\frac{\mu}{\lambda} - p_0(T)\right)(1-p_0(T))^{-1} = \frac{\mu(\lambda-\mu)}{\lambda(\lambda e^{(\lambda-\mu)T}-\mu)} \frac{\lambda e^{(\lambda-\mu)T}-\mu}{(\lambda-\mu)e^{(\lambda-\mu)T}} \leq \frac{\mu}{\lambda} e^{-(\lambda-\mu)}.$$

□

For the probability of $Y_T = 1$ conditioned on $Y_T > 0$, we have the following lemma.

Lemma A.2 (used for Lemma A.7)**.** *We have for* $T \geq \frac{1}{\lambda-\mu}\log(2)$

$$\mathbb{P}^*(Y_T = 1) \leq \frac{2(\lambda-\mu)}{\lambda} e^{-(\lambda-\mu)T}.$$

Proof: Using the probability mass functions p_n and the function $\tilde{p}$ from Section 2, for $T \geq \frac{1}{\lambda-\mu}\log(2)$, we obtain

$$\mathbb{P}^*(Y_T = 1) = \frac{p_1(T)}{1-p_0(T)} = 1 - \lambda\tilde{p}(T) = \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)T}-\mu} \leq \frac{2(\lambda-\mu)}{\lambda e^{(\lambda-\mu)T}}.$$

□

We next consider sub- and supermartingale properties of conditioned processes.

Lemma A.3 (used for Lemma A.4)**.**

(i) $(Y_t)_{t\geq 0}$ *conditioned on* $Y_\infty = 0$ *is a supermartingale.*
(ii) $(Y_t)_{t\geq 0}$ *conditioned on* $Y_\infty > 0$ *is a submartingale.*

Proof:

(i) Consider a subcritical linear birth and death process $(\tilde{Y}_t)_{t\geq 0}$ with *birth rate* μ, *death rate* λ and initial value one. Then $(\tilde{Y}_t)_{t\geq 0}$ has the same law as $(Y_t)_{t\geq 0}$ conditioned on $Y_\infty = 0$ (see e.g. page 78 in [Lam08]). It is well-known that a subcritical linear birth and death process is a supermartingale, which yields the result.

(ii) Consider a process $(\hat{Y}_t)_{t\geq 0}$ that has the law of $(Y_t)_{t\geq 0}$ conditioned on $Y_\infty > 0$. Note that $(\hat{Y}_t)_{t\geq 0}$ inherits the Markov property from $(Y_t)_{t\geq 0}$. Furthermore, it is well-known that $(Y_t)_{t\geq 0}$ is a submartingale. Thus we obtain for $t > s \geq 0$ and $y_s \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}(\hat{Y}_t | \hat{Y}_s = y_s) &= \mathbb{E}(Y_t | Y_s = y_s, Y_\infty > 0) \\
&= \frac{\mathbb{E}(\mathbb{1}_{\{Y_\infty > 0\}} Y_t | Y_s = y_s)}{\mathbb{P}(Y_\infty > 0 | Y_s = y_s)} \\
&= \frac{\mathbb{E}(Y_t | Y_s = y_s) - \mathbb{E}(\mathbb{1}_{\{Y_\infty = 0\}} Y_t | Y_s = y_s)}{\mathbb{P}(Y_\infty > 0 | Y_s = y_s)} \\
&\geq \frac{y_s - \mathbb{P}(Y_\infty = 0 | Y_s = y_s)\mathbb{E}(\tilde{Y}_t | \tilde{Y}_s = y_s)}{\mathbb{P}(Y_\infty > 0 | Y_s = y_s)}
\end{aligned}$$

$$\geq \frac{y_s - \mathbb{P}(Y_\infty = 0 | Y_s = y_s) y_s}{\mathbb{P}(Y_\infty > 0 | Y_s = y_s)}$$
$$= y_s,$$

where $(\tilde{Y}_t)_{t\geq 0}$ is the supermartingale from (i), which also inherits the Markov property from $(Y_t)_{t\geq 0}$. Thus $(Y_t)_{t\geq 0}$ conditioned on ultimate survival is a submartingale. □

The following is a simple consequence.

Lemma A.4 (used for Lemma A.8)**.** *$(Y_t^{-1})_{k\in\mathbb{N}}$ conditioned on $Y_\infty > 0$ is a supermartingale.*

Proof: In general, for a submartingale $(Z_t)_{t\geq 0}$ with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ with $Z_t \geq 1$ for all $t \geq 0$, we have for $t > s \geq 0$

$$\mathbb{E}\left(\frac{1}{Z_t} - \frac{1}{Z_s} \,\middle|\, \mathcal{F}_s\right) = \mathbb{E}\left(\frac{Z_s - Z_t}{Z_s Z_t} \,\middle|\, \mathcal{F}_s\right) \leq \mathbb{E}(Z_s - Z_t | \mathcal{F}_s) \leq 0.$$

Thus $(Z_t^{-1})_{t\geq 0}$ is a supermartingale. Consequently, Lemma A.3 implies that $(Y_t^{-1})_{t\geq 0}$ conditioned on ultimate survival is a supermartingale. □

In order to cope with dependencies, we already introduced the random variable $\mathcal{K}(T)$ in Definition 4.4. For the same reason, we define the deterministic number $\kappa(T)$, which depends on T in such a way that the probability for more than $\kappa(T)$ events up to time T decreases exponentially in T (see Lemma A.6(ii) below). The quantity $\kappa(T)$ is solely used in Lemma A.6 and in the proof of Lemma A.12 below. Essentially, we substitute the number of events $B_T + D_T$ up to time T by $\kappa(T)$ since it is difficult to treat the dependencies between the number of events up to time T and the event times.

Notation A.5. *Let $\kappa(T) := \lfloor e^{\frac{3}{2}(\lambda+\mu)T} \rfloor$.*

Lemma A.6 below states essentially that, for T large, it is unlikely that the randomly picked individual was born before $\frac{T}{2}$ or that more than $e^{\frac{3}{2}(\lambda+\mu)T}$ events have occurred by time T.

Lemma A.6.

(i) For the probability that fewer than $\mathcal{K}(T)$ events have occurred up to the birth time of the randomly picked node J_T was born, we have

$$\mathbb{P}^*(r(J_T) \leq \mathcal{K}(T)) = \mathbb{P}^*\left(T_{r(J_T)} \leq \frac{T}{2}\right) \leq e^{-\frac{1}{2}\lambda T} + \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda}\left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right).$$

(ii) For $T \geq \frac{2(\log(4\lambda) - \log(\lambda-\mu))}{\lambda+\mu}$, we have

$$\mathbb{P}^*(\kappa(T) \leq \mathcal{M}_T) \leq \frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4} e^{-(\lambda+\mu)T},$$

where $\mathcal{M}_T = B_T + D_T$ is the number of events up to time T as before.

Proof:

(i) We have

$$\mathbb{P}^*(r(J_T) \leq \mathcal{K}(T)) = \mathbb{P}^*(T_{r(J_T)} \leq T_{\mathcal{K}(T)}) = 1 - \mathbb{P}^*\left(T - T_{r(J_T)} \leq \frac{T}{2}\right)$$

Note that $\mathbb{P}^*(T - T_{r(J_T)} \leq T/2)$ is given by Corollary 2.7. Thus the result follows from this corollary by elementary computation.

(ii) Using Proposition 2.5, we compute

$$\begin{aligned}\kappa(T) - 2\mathbb{E}(B_T) &\geq e^{\frac{3}{2}(\lambda+\mu)T} - 1 - \frac{2\lambda}{\lambda-\mu}e^{(\lambda-\mu)T} + \frac{2\mu}{\lambda-\mu} \\ &= e^{(\lambda-\mu)T}\left(e^{2\mu T}e^{\frac{1}{2}(\lambda+\mu)T} - \frac{2\lambda}{\lambda-\mu}\right) + \frac{3\mu-\lambda}{\lambda-\mu}. \end{aligned} \tag{A.2}$$

Since $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$, we have

$$\frac{1}{2}e^{2\mu T}e^{\frac{1}{2}(\lambda+\mu)T} \geq \frac{2\lambda}{\lambda-\mu}. \tag{A.3}$$

Thus for $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$, the right-hand side of (A.2) is larger than or equal to

$$e^{(\lambda-\mu)T}\frac{1}{2}e^{2\mu T}e^{\frac{1}{2}(\lambda+\mu)T} + \frac{3\mu-\lambda}{\lambda-\mu}. \tag{A.4}$$

Since for $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$, we have

$$e^{(\lambda-\mu)T} \geq 1 + (\lambda-\mu)T \geq 1 + 2\log\left(\frac{4\lambda}{\lambda+\mu}\right) \geq 1 + 2\log(4) \geq 2, \tag{A.5}$$

inequality (A.3) implies that for $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda-\mu}$, the expression (A.4) is bounded from below by

$$e^{(\lambda-\mu)T}\frac{3}{8}e^{2\mu T}e^{\frac{1}{2}(\lambda+\mu)T} + \frac{3\mu}{\lambda-\mu} \geq \frac{3}{8}e^{\frac{3}{2}(\lambda+\mu)T}.$$

Thus for $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$, we have

$$\kappa(T) - 2\mathbb{E}(B_T) \geq \frac{3}{8}e^{\frac{3}{2}(\lambda+\mu)T} > 0. \tag{A.6}$$

Consequently, we can apply Chebyshev's inequality and obtain

$$\begin{aligned}\mathbb{P}(\kappa(T) \leq \mathcal{M}_T | Y_T > 0) &\leq \frac{\mathbb{P}(\kappa(T) \leq \mathcal{M}_T)}{\mathbb{P}(Y_T > 0)} \leq \frac{\mathbb{P}(\kappa(T) \leq 2B_T)}{\mathbb{P}(Y_T > 0)} \\ &= \frac{1}{\mathbb{P}(Y_T > 0)}\mathbb{P}(\kappa(T) - 2\mathbb{E}(B_T) \leq 2B_T - 2\mathbb{E}(B_T)) \\ &\leq \frac{1}{\mathbb{P}(Y_T > 0)}\frac{4\mathrm{Var}(B_T)}{(\kappa(T) - 2\mathbb{E}(B_T))^2}, \end{aligned} \tag{A.7}$$

where $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$. By $\mathbb{P}(Y_T > 0) \geq \mathbb{P}(Y_\infty > 0) = \frac{\lambda-\mu}{\lambda}$, Proposition 2.5 and (A.6), for $T \geq \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu}$, the right-hand side of (A.7) is smaller than or equal to

$$\frac{256}{9}\frac{\lambda}{\lambda-\mu}\frac{\mathrm{Var}(B_T)}{e^{3(\lambda+\mu)T}} \leq \frac{30\lambda}{\lambda-\mu}\left(\frac{\lambda^2(\lambda+\mu)}{(\lambda-\mu)^3}e^{-(\lambda+\mu)T} + \frac{2\lambda^2\mu}{(\lambda-\mu)^3}e^{-2(\lambda+\mu)T}\right). \tag{A.8}$$

Since $e^{-(\lambda+\mu)T} \leq \frac{1}{2}$ by (A.5), the right-hand side of (A.8) is smaller than or equal to

$$\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}e^{-(\lambda+\mu)T}.$$

□

Further lemmas

In what follows we give some more specialized results that are used in the proof of Theorem 4.3.

The following lemma states that the time $T - T_{\mathcal{M}_T}$ since the last event becomes small quickly and is proved using Theorem 2.11.

Lemma A.7. *For $T \geq \frac{1}{\lambda-\mu}\log(2)$ and $c > 0$, we have*

$$\mathbb{E}^*(1 - e^{-c(T-T_{\mathcal{M}_T})}) \leq \frac{2(\lambda-\mu)}{\lambda e^{(\lambda-\mu)T}} + \frac{2c}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)e^{-(\lambda-\mu)T}. \tag{A.9}$$

Proof: Let X have the cumulative distribution function

$$G(t) = \mathbb{1}_{\{Y_T>1\}}(1 - e^{-(Y_T-1)(\lambda-\mu)t}) + \mathbb{1}_{\{Y_T\leq 1\}}\mathbb{1}_{\{t\geq T\}}.$$

Conditionally on Y_T, we then have $T - T_{\mathcal{M}_T} \leq_{st} X$ by Theorem 2.11, which implies

$$\begin{aligned}\mathbb{E}^*(1 - e^{-c(T-T_{\mathcal{M}_T})}) &\leq \mathbb{E}^*(1-e^{-cX}) \\ &= \mathbb{E}^*\Big((1-e^{-cT})\mathbb{P}(X=T|Y_T) + \mathbb{E}(1-e^{-cX}|X<T,Y_T)\mathbb{P}(X<T|Y_T)\Big). \end{aligned}\tag{A.10}$$

Since $\mathbb{P}(X = T|Y_T = 1) = 1$ and $\mathcal{L}(X|Y_T) = \mathrm{Exp}((Y_T - 1)\lambda)$ on $\{Y_T \geq 2\}$, we obtain that the right-hand side of (A.10) is smaller than or equal to

$$\mathbb{E}^*\left(\mathbb{1}_{\{Y_T=1\}} + \frac{c}{c+(Y_T-1)\lambda}\mathbb{1}_{\{Y_T>1\}}\right) \leq \mathbb{P}^*(Y_T=1) + \frac{c}{\lambda}\mathbb{E}^*\left(\frac{1}{Y_T-1}\mathbb{1}_{\{Y_T>1\}}\right).$$

For the conditional probability of $Y_T = 1$, by Lemma A.2, we have for $T \geq \frac{1}{\lambda-\mu}\log(2)$

$$\mathbb{P}^*(Y_T = 1) \leq \frac{2(\lambda-\mu)}{\lambda e^{(\lambda-\mu)T}}.$$

Furthermore, we have

$$\begin{aligned}\mathbb{E}^*\left(\frac{1}{Y_T-1}\mathbb{1}_{\{Y_T>1\}}\right) &= \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)T}-\mu}\sum_{n=1}^{\infty}\frac{1}{n}\left(\frac{\lambda e^{(\lambda-\mu)T}-\lambda}{\lambda e^{(\lambda-\mu)T}-\mu}\right)^n \\ &= \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)T}-\mu}\left(-\log\left(1-\frac{\lambda e^{(\lambda-\mu)T}-\lambda}{\lambda e^{(\lambda-\mu)T}-\mu}\right)\right) \\ &= \frac{\lambda-\mu}{\lambda e^{(\lambda-\mu)T}-\mu}\log\left(\frac{\lambda e^{(\lambda-\mu)T}-\mu}{\lambda-\mu}\right) \\ &\leq \frac{2}{e^{(\lambda-\mu)T}}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)\end{aligned}$$

for $T \geq \frac{1}{\lambda-\mu}\log(2)$, which completes the proof. □

The next result follows from Lemma A.4 and a well-known inequality for supermartingales.

Lemma A.8. *For all $\delta, \gamma > 0$, we have*

$$\mathbb{E}\left(\frac{1}{\min\limits_{T/2\leq t} Y_t^\delta}\,\middle|\, Y_\infty > 0\right) \leq e^{-\gamma(\lambda-\mu)T} + \mathbb{E}\left(\frac{1}{Y_{T/2}}\,\middle|\, Y_\infty > 0\right)e^{\frac{\gamma}{\delta}(\lambda-\mu)T}.$$

Proof: Writing $Z_{T/2} = \max_{T/2 \le t} Y_t^{-1}$, we obtain

$$\begin{aligned}
\mathbb{E}\big(Z_{T/2}^{\delta} \mid Y_\infty > 0\big) &\le \mathbb{E}\Big(\mathbb{E}\big(\mathbb{1}_{\{Z_{T/2}^{\delta} \le e^{-\gamma(\lambda-\mu)T}\}} Z_{T/2}^{\delta} \mid Y_{T/2}\big) \Big| Y_\infty > 0\Big) \\
&\quad + \mathbb{E}\Big(\mathbb{E}\big(\mathbb{1}_{\{Z_{T/2}^{\delta} > e^{-\gamma(\lambda-\mu)T}\}} Z_{T/2}^{\delta} \mid Y_{T/2}\big) \Big| Y_\infty > 0\Big) \\
&\le e^{-\gamma(\lambda-\mu)T} + \mathbb{E}\Big(\mathbb{P}\big(Z_{T/2} > e^{-\frac{\gamma}{\delta}(\lambda-\mu)T} \mid Y_{T/2}\big) \Big| Y_\infty > 0\Big) \\
&\le e^{-\gamma(\lambda-\mu)T} + \mathbb{E}\big(Y_{T/2}^{-1} \mid Y_\infty > 0\big) e^{\frac{\gamma}{\delta}(\lambda-\mu)T},
\end{aligned}$$

where the last line follows from Lemma A.4 and the the inequality (2) in Theorem 6.14 on page 99 in [Yeh95]. □

We use the following lemma to bound the conditional expectation on the right hand side of Lemma A.8 from above.

Lemma A.9. *We have for* $T \ge \frac{2}{\lambda-\mu}\log(2)$

$$\mathbb{E}\left(\frac{1}{Y_{T/2}} \,\middle|\, Y_\infty > 0\right) \le 2\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + \frac{1}{2}(\lambda-\mu)T\right) e^{-\frac{1}{2}(\lambda-\mu)T}.$$

Proof: For $T \ge \frac{2}{\lambda-\mu}\log(2)$, we obtain

$$\begin{aligned}
\mathbb{E}\big(Y_{T/2}^{-1} \mid Y_\infty > 0\big) &= \frac{\mathbb{E}\big(Y_{T/2}^{-1}\mathbb{1}_{\{Y_\infty>0\}}\big)}{\mathbb{P}(Y_\infty > 0)} \le \frac{\lambda}{\lambda-\mu}\mathbb{E}\big(Y_{T/2}^{-1}\mathbb{1}_{\{Y_{T/2}>0\}}\big) \le \frac{\lambda}{\lambda-\mu}\mathbb{E}\big(Y_{T/2}^{-1} \mid Y_{T/2} > 0\big) \\
&\le \frac{\lambda}{\lambda e^{\frac{1}{2}(\lambda-\mu)T} - \lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + \frac{1}{2}(\lambda-\mu)T\right) \\
&\le \frac{1}{\frac{1}{2}e^{\frac{1}{2}(\lambda-\mu)T} + \frac{1}{2}e^{\frac{1}{2}(\lambda-\mu)T} - 1}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + \frac{1}{2}(\lambda-\mu)T\right) \\
&\le 2\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + \frac{1}{2}(\lambda-\mu)T\right) e^{-\frac{1}{2}(\lambda-\mu)T},
\end{aligned}$$

where the second line follows from Proposition 2.3 and the last inequality holds for $T \ge \frac{2}{\lambda-\mu}\log(2)$. □

Recall that $R_{T_l,T}$ is the number of nodes that are alive at time T_l and survive up to time T. The following lemma gives us the first two conditional moments of $R_{T_l,T}$.

Lemma A.10 (used for Lemma A.11)**.** *We have*

$$\begin{aligned}
&\mathbb{E}(R_{T_l,T} | Y_{T_l}, T_l, T_{l+1}) = (Y_{T_l} - 1)e^{-\mu(T-T_{l+1})} \\
&\mathbb{E}(R_{T_l,T}^2 | Y_{T_l}, T_l, T_{l+1}) = Y_{T_l}e^{-\mu(T-T_{l+1})} - Y_{T_l}e^{-2\mu(T-T_{l+1})} + Y_{T_l}^2 e^{-2\mu(T-T_{l+1})} \\
&\qquad - \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} - 1 | Y_{T_l}, T_l, T_{l+1})\Big(2(Y_{T_l} - 1)e^{-2\mu(T-T_{l+1})} + e^{-\mu(T-T_{l+1})}\Big).
\end{aligned}$$

Proof: Firstly, we determine the conditional expectation of $R_{T_l,T}$:

$$\begin{aligned}
\mathbb{E}(R_{T_l,T} | Y_{T_l}, T_l, T_{l+1}) &= p^{+}\mathbb{E}(R_{T_l,T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\
&\quad + p^{-}\mathbb{E}(R_{T_l,T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1), \qquad \text{(A.11)}
\end{aligned}$$

where $p^{+} := \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} + 1 | Y_{T_l}, T_l, T_{l+1})$ and $p^{-} := \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} - 1 | Y_{T_l}, T_l, T_{l+1})$.

With

$$\begin{aligned}\mathbb{E}(R_{T_l,T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1) &= \mathbb{E}(R_{T_{l+1},T}-\mathbb{1}_{\{T^-_{r^{-1}(l+1)}>T\}}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &= (Y_{T_l}+1)e^{-\mu(T-T_{l+1})}-e^{-\mu(T-T_{l+1})}=Y_{T_l}e^{-\mu(T-T_{l+1})} \qquad \text{(A.12)}\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}(R_{T_l,T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}-1) &= \mathbb{E}(R_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}-1).\\ &= (Y_{T_l}-1)e^{-\mu(T-T_{l+1})}.\end{aligned}$$

Equation (A.11) implies

$$\mathbb{E}(R_{T_l,T}|Y_{T_l},T_l,T_{l+1})=Y_{T_l}e^{-\mu(T-T_{l+1})}-e^{-\mu(T-T_{l+1})}p^-.$$

Secondly, we compute the conditional second moment of $R_{T_l,T}$:

$$\begin{aligned}\mathbb{E}(R^2_{T_l,T}|Y_{T_l},T_l,T_{l+1}) &= p^+\mathbb{E}(R^2_{T_l,T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &\quad+p^-\mathbb{E}(R^2_{T_l,T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}-1). \qquad \text{(A.13)}\end{aligned}$$

We treat the summands separately again. For the case where a birth occurs at time T_{l+1}, we have

$$\begin{aligned}&\mathbb{E}(R^2_{T_l,T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &=\mathbb{E}((R_{T_{l+1},T}-\mathbb{1}_{\{T^-_{r^{-1}(l+1)}>T\}})^2|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &=\mathbb{E}\Big(R^2_{T_{l+1},T}-2R_{T_{l+1},T}\mathbb{1}_{\{T^-_{r^{-1}(l+1)}>T\}}+\mathbb{1}_{\{T^-_{r^{-1}(l+1)}>T\}}\;\Big|\;Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1,Y_{T_l}>0\Big)\end{aligned} \qquad \text{(A.14)}$$

and further

$$\begin{aligned}\mathbb{E}(R^2_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1) &= \mathrm{Var}(R_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &\quad+(\mathbb{E}(R_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1))^2. \qquad \text{(A.15)}\end{aligned}$$

Given $Y_{T_{l+1}}$, let $\mathcal{L}_{l+1}$ be the set of the $Y_{T_{l+1}}$ nodes living at time T_{l+1}. Then by independence of various death times, we obtain for the conditional variance

$$\begin{aligned}&\mathrm{Var}(R_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)\\ &=\mathrm{Var}\Big(\sum_{j\in\mathcal{L}_{l+1}}\mathbb{1}_{\{T^-_j>T\}}\;\Big|\;Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1,Y_{T_l}>0\Big)\\ &=(Y_{T_l}+1)(e^{-\mu(T-T_{l+1})}-e^{-2\mu(T-T_{l+1})}).\end{aligned}$$

For the second summand of (A.15), we obtain:

$$\mathbb{E}(R_{T_{l+1},T}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)=(Y_{T_l}+1)e^{-\mu(T-T_{l+1})}.$$

Thus (A.15) is equal to

$$(Y_{T_l}+1)\Big(e^{-\mu(T-T_{l+1})}+Y_{T_l}e^{-2\mu(T-T_{l+1})}\Big).$$

For the remaining parts of (A.14), we have

$$\mathbb{E}(R_{T_{l+1},T}\mathbb{1}_{\{T^-_{r^{-1}(l+1)}>T\}}|Y_{T_l},T_l,T_{l+1},Y_{T_{l+1}}=Y_{T_l}+1)$$

$$
\begin{aligned}
&= \mathbb{P}(T^{-}_{r^{-1}(l+1)} > T | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\
&\quad \cdot \mathbb{E}(R_{T_{l+1},T} | T^{-}_{r^{-1}(l+1)} > T, Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\
&= e^{-\mu(T-T_{l+1})}(1 + Y_{T_l} e^{-\mu(T-T_{l+1})})
\end{aligned}
$$

and

$$
\mathbb{E}(\mathbb{1}_{\{T^{-}_{r^{-1}(l+1)} > T\}} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) = e^{-\mu(T-T_{l+1})}.
$$

Thus (A.14) implies

$$
\begin{aligned}
&\mathbb{E}(R^2_{T_l,T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\
&= (Y_{T_l} + 1)\Big(e^{-\mu(T-T_{l+1}} + Y_{T_l} e^{-2\mu(T-T_{l+1})}\Big) - 2\Big(e^{-\mu(T-T_{l+1}} + Y_{T_l} e^{-2\mu(T-T_{l+1})}\Big) + e^{-\mu(T-T_{l+1})} \\
&= Y_{T_l} e^{-\mu(T-T_{l+1})} - Y_{T_l} e^{-2\mu(T-T_{l+1})} + Y^2_{T_l} e^{-2\mu(T-T_{l+1})}.
\end{aligned}
$$

For the case where a death occurs at time T_{l+1}, we have

$$
\begin{aligned}
&\mathbb{E}(R^2_{T_l,T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1) \\
&= \mathrm{Var}(R_{T_{l+1},T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1) + (\mathbb{E}(R_{T_{l+1},T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1))^2 \\
&= (Y_{T_l} - 1)(e^{-\mu(T-T_{l+1})} - e^{-2\mu(T-T_{l+1})}) + (Y_{T_l} - 1)^2 e^{-2\mu(T-T_{l+1})} \\
&= (Y_{T_l} - 1)e^{-\mu(T-T_{l+1})} + (Y^2_{T_l} - 3Y_{T_l} + 2)e^{-2\mu(T-T_{l+1})}.
\end{aligned}
$$

Thus from (A.13) follows

$$
\begin{aligned}
\mathbb{E}(R^2_{T_l,T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_l} > 0) = {} & Y_{T_l} e^{-\mu(T-T_{l+1})} - Y_{T_l} e^{-2\mu(T-T_{l+1})} + Y^2_{T_l} e^{-2\mu(T-T_{l+1})} \\
& + p^{-}\Big(2(1 - Y_{T_l})e^{-2\mu(T-T_{l+1})} - e^{-\mu(T-T_{l+1})}\Big).
\end{aligned}
$$

□

Knowing the conditional moments of $R_{T_l,T}$, we can find an upper bound for a more complex conditional expectation involving $R_{T_l,T}$ that appears in the proof of the main theorem. The following lemma allows us to control the proportion of nodes surviving up to time T of the nodes alive at time T_l (if we reduce both numbers by one).

Lemma A.11. *For $l \in \mathbb{N}$, we have*

$$
\mathbb{E}\left(\left| \frac{R_{T_l,T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| \;\middle|\; Y_{T_l}, T_l, T_{l+1}\right) \le \left(\frac{6}{Y_{T_l}}\right)^{\frac{1}{2}}.
$$

Proof: By Jensen's inequality, we obtain

$$
\begin{aligned}
&\mathbb{E}\left(\left| \frac{R_{T_l,T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| \;\middle|\; Y_{T_l}, T_l, T_{l+1}\right) \\
&\le \left(\mathbb{E}\left(\left(\frac{R_{T_l,T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})}\right)^2 \;\middle|\; Y_{T_l}, T_l, T_{l+1}\right)\right)^{\frac{1}{2}} \\
&= \left(\mathbb{1}_{\{Y_{T_l} > 1\}} \frac{\mathbb{E}((R_{T_l,T} - 1)^2 | Y_{T_l}, T_l, T_{l+1})}{(Y_{T_l} - 1)^2} - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} \mathbb{E}(R_{T_l,T} - 1 | Y_{T_l}, T_l, T_{l+1})\right. \\
&\quad \left. + e^{-2\mu(T-T_{l+1})}\right)^{\frac{1}{2}}.
\end{aligned} \tag{A.16}
$$

Lemma A.10 implies

$$\frac{2}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\mu(T-T_{l+1})}\mathbb{E}(R_{T_l,T}-1|Y_{T_l},T_l,T_{l+1}) = 2\cdot\mathbb{1}_{\{Y_{T_l}>1\}}e^{-2\mu(T-T_{l+1})} - \frac{2}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\mu(T-T_{l+1})}. \tag{A.17}$$

and

$$\begin{aligned}
&\frac{1}{(Y_{T_l}-1)^2}\mathbb{1}_{\{Y_{T_l}>1\}}\mathbb{E}((R_{T_l,T}-1)^2|Y_{T_l},T_l,T_{l+1})\\
&=\frac{1}{(Y_{T_l}-1)^2}\mathbb{1}_{\{Y_{T_l}>1\}}\Big(\mathbb{E}(R^2_{T_l,T}|Y_{T_l},T_l,T_{l+1})-2\mathbb{E}(R_{T_l,T}|Y_{T_l},T_l,T_{l+1})+1\Big)\\
&=\frac{1}{(Y_{T_l}-1)^2}\mathbb{1}_{\{Y_{T_l}>1\}}\Big(Y_{T_l}e^{-\mu(T-T_{l+1})}-Y_{T_l}e^{-2\mu(T-T_{l+1})}+Y_{T_l}^2e^{-2\mu(T-T_{l+1})}\\
&\qquad-\mathbb{P}(Y_{T_{l+1}}=Y_{T_l}-1|Y_{T_l},T_l,T_{l+1})\Big(2(Y_{T_l}-1)e^{-2\mu(T-T_{l+1})}+e^{-\mu(T-T_{l+1})}\Big)\Big)\\
&\quad-\frac{2}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\mu(T-T_{l+1})}+\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{(Y_{T_l}-1)^2}\\
&\le\frac{1}{(Y_{T_l}-1)^2}\mathbb{1}_{\{Y_{T_l}>1\}}\Big(Y_{T_l}e^{-\mu(T-T_{l+1})}+Y_{T_l}(Y_{T_l}-1)e^{-2\mu(T-T_{l+1})}\Big)\\
&\quad-\frac{2}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\mu(T-T_{l+1})}+\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{(Y_{T_l}-1)^2}\\
&=\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{Y_{T_l}-1}e^{-\mu(T-T_{l+1})}+\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{(Y_{T_l}-1)^2}e^{-\mu(T-T_{l+1})}+\frac{Y_{T_l}}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-2\mu(T-T_{l+1})}\\
&\quad-\frac{2}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-\mu(T-T_{l+1})}+\frac{\mathbb{1}_{\{Y_{T_l}>1\}}}{(Y_{T_l}-1)^2}\\
&\le\mathbb{1}_{\{Y_{T_l}>1\}}e^{-2\mu(T-T_{l+1})}+\frac{1}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}e^{-2\mu(T-T_{l+1})}+\frac{1}{(Y_{T_l}-1)^2}\mathbb{1}_{\{Y_{T_l}>1\}}\\
&\le\Big(\frac{1}{Y_{T_l}-1}+e^{-2\mu(T-T_{l+1})}\Big)\mathbb{1}_{\{Y_{T_l}>1\}}.
\end{aligned}\tag{A.18}$$

By (A.17) and (A.18), we can bound the right-hand side of (A.16) from above by

$$\begin{aligned}
\Big(\frac{3}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}-\mathbb{1}_{\{Y_{T_l}>1\}}e^{-2\mu(T-T_{l+1})}+e^{-2\mu(T-T_{l+1})}\Big)^{\frac12}&\le\Big(\frac{3}{Y_{T_l}-1}\mathbb{1}_{\{Y_{T_l}>1\}}+\mathbb{1}_{\{Y_{T_l}=1\}}\Big)^{\frac12}\\
&\le\Big(\frac{6}{Y_{T_l}}\Big)^{\frac12}.
\end{aligned}\tag{A.19}$$

□

For the conditional expectation of the sum of the squared inter-event times since the birth of the randomly picked node, we have the following lemma, which is proved by using Lemmas A.6, A.8 and A.9.

Lemma A.12. *For* $T\ge(\frac{2}{\lambda-\mu}\log(2)\vee\frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$, *we have*

$$\begin{aligned}
\mathbb{E}^*\Big(\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(T_{l+1}-T_l)^2\Big)&\le\frac{\mu}{\lambda}\frac{T^2}{4}e^{-(\lambda-\mu)T}+60T^2\frac{\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}e^{-(\lambda+\mu)T}+T^2e^{-\frac12\lambda T}\\
&\quad+\frac{2(\lambda-\mu)}{\lambda}\Big(\log\Big(\frac{\lambda}{\lambda-\mu}\Big)+(\lambda-\mu)T\Big)T^2e^{-(\lambda-\mu)T}
\end{aligned}$$

$$+\left(\frac{3}{4}T^2+\frac{T}{2(\lambda+\mu)}\right)\left(1+2\log\left(\frac{\lambda}{\lambda-\mu}\right)+(\lambda-\mu)T\right)e^{-\frac{1}{4}(\lambda-\mu)T}.$$

Proof: For the left-hand side, we deduce

$$\begin{aligned}
&\mathbb{E}\bigg(\sum_{l=r(J_T)}^{\mathcal{M}_T-1}(T_{l+1}-T_l)^2 \,\bigg|\, Y_T>0\bigg)\\
&\le \mathbb{E}\bigg(\mathbb{1}_{\{T_{\kappa(T)}>T\}}\mathbb{1}_{\{T_{\mathcal{K}(T)}<T_{r(J_T)}\}}\sum_{l=1}^{\mathcal{M}_T-1}\mathbb{1}_{\{T_l\ge T_{r(J_T)}\}}(T_{l+1}-T_l)\max_{r(J_T)\le j\le \mathcal{M}_T-1}(T_{j+1}-T_j)\,\bigg|\, Y_T>0\bigg)\\
&\quad+\mathbb{E}(\mathbb{1}_{\{T_{\kappa(T)}\le T\}}|Y_T>0)T^2+\mathbb{E}(\mathbb{1}_{T_{\mathcal{K}(T)}\ge T_{r(J_T)}\}}|Y_T>0)T^2. \qquad\qquad (A.20)
\end{aligned}$$

Note that, given $(Y_{T_k})_{k\in\mathbb{N}}$ and $\mathcal{K}(T)$, the inter-event times $T_{j+1}-T_j$ are $\mathrm{Exp}((\lambda+\mu)Y_{T_j})$ distributed and independent for $j>\mathcal{K}(T)$. In order to derive an upper bound for the first conditional expectation on the right-hand side of (A.20), we introduce a sequence of random variables $(U_j)_{j\in\mathbb{N}}$ such that, given $(Y_{T_k})_{k\in\mathbb{N}}$ and $\mathcal{K}(T)$, $U_j\sim\mathrm{Exp}((\lambda+\mu)\min_{\mathcal{K}(T)<k}Y_{T_k})$ i.i.d. Then, given $(Y_{T_k})_{k\in\mathbb{N}}$ and $\mathcal{K}(T)$, we have $T_{j+1}-T_j\le_{st}U_j$ for $\mathcal{K}(T)<j\le\kappa(T)$ and obtain

$$\begin{aligned}
&\mathbb{E}\bigg(\mathbb{1}_{\{T_{\kappa(T)}>T\}}\mathbb{1}_{\{T_{\mathcal{K}(T)}<T_{r(J_T)}\}}\sum_{l=1}^{\mathcal{M}_T-1}\mathbb{1}_{\{T_l\ge T_{r(J_T)}\}}(T_{l+1}-T_l)\max_{r(J_T)\le j\le \mathcal{M}_T-1}(T_{j+1}-T_j)\,\bigg|\, Y_T>0\bigg)\\
&\le\frac{T}{2}\mathbb{E}\bigg(\mathbb{1}_{\{T_{\kappa(T)}>T\}}\mathbb{1}_{\{T_{\mathcal{K}(T)}<T_{r(J_T)}\}}\max_{r(J_T)\le j\le \mathcal{M}_T-1}(T_{j+1}-T_j)\,\bigg|\, Y_\infty>0\bigg)+\frac{T^2}{4}\mathbb{P}(Y_\infty=0|Y_T>0). \\
&\qquad\qquad (A.21)
\end{aligned}$$

By Lemma A.1, the second summand is equal to

$$\frac{T^2}{4}\frac{\mu}{\lambda}e^{-(\lambda-\mu)T}.$$

The first summand of (A.21) is bounded from above by

$$\begin{aligned}
&\frac{T}{2}\mathbb{E}\bigg(\mathbb{E}\bigg(\mathbb{1}_{\{T_{\kappa(T)}>T\}}\mathbb{1}_{\{T_{\mathcal{K}(T)}<T_{r(J_T)}\}}\max_{\mathcal{K}(T)<j\le\kappa(T)}(T_{j+1}-T_j)\,\bigg|\,\mathcal{K}(T),(Y_{T_k})_{k\ge1}\bigg)\,\bigg|\,Y_\infty>0\bigg)\\
&\le\frac{T}{2}\mathbb{E}\bigg(\mathbb{E}\bigg(\max_{1\le j\le\kappa(T)}U_j\,\bigg|\,\mathcal{K}(T),(Y_{T_k})_{k\ge1}\bigg)\,\bigg|\,Y_\infty>0\bigg)\\
&=\frac{T}{2}\mathbb{E}\bigg(\frac{1}{(\lambda+\mu)\min\limits_{\mathcal{K}(T)<k}Y_{T_k}}\sum_{l=1}^{\kappa(T)}\frac{1}{l}\,\bigg|\,Y_\infty>0\bigg),
\end{aligned}$$

where the last equality follows from from the formula for the expectation of the maximum of i.i.d. exponentially distributed random variables (see e.g. the introduction of [Eis08]).

Using the well-known upper bound for the harmonic sum yields

$$\begin{aligned}
&\frac{T}{2}\mathbb{E}\bigg(\frac{1}{(\lambda+\mu)\min\limits_{\mathcal{K}(T)<k}Y_{T_k}}\sum_{l=1}^{\kappa(T)}\frac{1}{l}\,\bigg|\,Y_\infty>0\bigg)\\
&\le\frac{T}{2(\lambda+\mu)}\mathbb{E}\bigg(\frac{1}{\min\limits_{\mathcal{K}(T)<k}Y_{T_k}}\,\bigg|\,Y_\infty>0\bigg)(\log(\kappa(T))+1)\\
&\le\frac{T}{2(\lambda+\mu)}\mathbb{E}\bigg(\frac{1}{\min\limits_{\mathcal{K}(T)<k}Y_{T_k}}\,\bigg|\,Y_\infty>0\bigg)\left(\frac{3}{2}(\lambda+\mu)T+1\right). \qquad\qquad (A.22)
\end{aligned}$$

For the conditional expectation in (A.22), we obtain for $T \geq \frac{2\log(2)}{\lambda-\mu}$

$$\mathbb{E}\left(\frac{1}{\min\limits_{\mathcal{K}(T)<k} Y_{T_k}} \,\middle|\, Y_\infty > 0\right) = \mathbb{E}\left(\frac{1}{\min\limits_{T/2\leq t} Y_t} \,\middle|\, Y_\infty > 0\right)$$
$$\leq \left(1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)e^{-\frac{1}{4}(\lambda-\mu)T}. \tag{A.23}$$

where the last line follows from Lemma A.8 with $\delta = 1$ and $\gamma = \frac{1}{4}$ and Lemma A.9.

Thus we can conclude that the first summand of the right-hand side of (A.21) is smaller than or equal to

$$\left(\frac{3}{4}T^2 + \frac{T}{2(\lambda+\mu)}\right)\left(1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)e^{-\frac{1}{4}(\lambda-\mu)T}$$

for sufficiently large T.

For the last line of (A.20), we can use the upper bounds from Lemma A.6 and obtain that for $T \geq (\frac{1}{\lambda-\mu}\log(2) \vee \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$, it is smaller than or equal to

$$\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}T^2e^{-(\lambda+\mu)T} + T^2e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda-\mu)}{\lambda}\left(\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T\right)T^2e^{-(\lambda-\mu)T}. \tag{A.24}$$

Altogether, we obtain the statement of the lemma. □

Appendix A2: Negligibility of multiple edges

In this section, we show that multiple edges are negligible (Lemma A.13 below) and use this result to prove Corollary 1.4 from the introduction, which states our main result for the case where multiple edges are ignored.

Lemma A.13. *The probability that an individual picked uniformly at random at time T has at least one multiple edge given the number of nodes is positive at time T is of the order $O(T^2e^{-\frac{1}{6}(\lambda-\mu)T})$ as $T \to \infty$.*

Proof. Let N_T denote the degree of the randomly picked node J_T at time T, i.e. the number of edges that are incident to the node picked uniformly at random at time T. Let $\rho_1 < \ldots < \rho_{N_T}$ be the birth times of these edges. Condition on J_T, N_T, $\rho_1 < \ldots < \rho_{N_T}$ and $(Y_t)_{0\leq t\leq T}$. Then the probability that we do not create a multiple edge at time ϱ_i that stays a multiple edge up to time T is

$$\left(1 - \frac{i-1}{Y_{\rho_i}-1}\right)\mathbb{1}_{\{Y_{\rho_i}\geq i+1\}}.$$

Thus we obtain for the probability that J_T has at least one multiple edge

$$\mathbb{E}\left(1 - \left(1 - \frac{1}{Y_{\rho_2}-1}\right)\mathbb{1}_{\{Y_{\rho_2}\geq 3\}} \cdot \left(1 - \frac{2}{Y_{\rho_3}-1}\right)\mathbb{1}_{\{Y_{\rho_3}\geq 4\}} \cdot \ldots \cdot \left(1 - \frac{N_T-1}{Y_{\rho_{N_T}}-1}\right)\mathbb{1}_{\{Y_{\rho_{N_T}}\geq N_T+1\}} \,\middle|\, Y_T > 0\right)$$
$$\leq \mathbb{E}\left(1 - \left(1 - \frac{N_T-1}{\min\limits_{T-A_{J_T}\leq t\leq T} Y_t - 1}\right)^{N_T-1}\mathbb{1}_{\{\min_{T-A_{J_T}\leq t\leq T} Y_t - 1 \geq N_T\}} \,\middle|\, Y_T > 0\right)$$
$$\leq \mathbb{P}(N_T \geq e^{\frac{1}{12}(\lambda-\mu)T}|Y_T > 0) + \mathbb{P}\left(\min_{T-A_{J_T}\leq t\leq T} Y_t - 1 \leq e^{\frac{1}{3}(\lambda-\mu)T} \,\middle|\, Y_T > 0\right)$$
$$+ 1 - \left(1 - \frac{e^{\frac{1}{12}(\lambda-\mu)T} - 1}{e^{\frac{1}{3}(\lambda-\mu)T}}\right)^{e^{\frac{1}{12}(\lambda-\mu)T}-1}. \tag{A.25}$$

Writing N_∞ for a random variable having the asymptotic degree distribution MixPo(M*) with M∗ defined at the beginning of Subsection 4.2, we obtain by conditioning on M* that the second moment $\mathbb{E}(N_\infty^2)$ is equal to

$$\frac{2\alpha E(S)}{\lambda+\beta+\mu}+\frac{2\alpha\mathbb{E}((S+\mathbb{E}(S))^2)}{(\lambda+\beta+\mu)(\lambda+2(\beta+\mu))} \text{ (cf. Subsection 3.3 in [BL10]).}$$

Thus Theorem 4.3 and the Markov inequality imply

$$\begin{aligned}\mathbb{P}(N_T \geq e^{\frac{1}{12}(\lambda-\mu)T}\,|\,Y_T>0) &\leq \mathbb{E}(N_\infty^2)T^{\frac{1}{4}}e^{-\frac{1}{8}(\lambda-\mu)T}+O(T^2e^{-\frac{1}{6}(\lambda-\mu)T})\\ &\leq \left(\frac{2\alpha E(S)}{\lambda+\beta+\mu}+\frac{2\alpha\mathbb{E}((S+\mathbb{E}(S))^2)}{(\lambda+\beta+\mu)(\lambda+2(\beta+\mu))}\right)e^{-\frac{1}{6}(\lambda-\mu)T}+O(T^2e^{-\frac{1}{6}(\lambda-\mu)T})\\ &= O(T^2e^{-\frac{1}{6}(\lambda-\mu)T}).\end{aligned}$$

For the second summand of the right-hand side of (A.25), we obtain

$$\begin{aligned}\mathbb{P}\left(\min_{T-A_{J_T}\leq t\leq T} Y_t-1\leq e^{\frac{1}{3}(\lambda-\mu)T}\,\middle|\,Y_T>0\right) &\leq \mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1}\,\middle|\,Y_T>0\right)\\ &\quad+\mathbb{P}\left(A_{J_T}<\frac{T}{2}\,\middle|\,Y_T>0\right).\end{aligned}\tag{A.26}$$

By Lemma A.6, the second summand of the right-hand side is of the order $O(e^{-\frac{1}{2}(\lambda-\mu)T})$ as $T\to\infty$. With $Y_\infty=\lim_{t\to\infty}Y_t\in\{0,\infty\}$, we have

$$\begin{aligned}\mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1},Y_\infty>0\,\middle|\,Y_T>0\right) &= \frac{\mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1},Y_\infty>0\right)}{\mathbb{P}(Y_T>0)}\\ &\leq \frac{\mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1},Y_\infty>0\right)}{\mathbb{P}(Y_\infty>0)}\\ &\leq \mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1}\,\middle|\,Y_\infty>0\right).\end{aligned}$$

Thus the first summand of the right-hand side of (A.26) is smaller than or equal to

$$\mathbb{P}\left(\max_{T/2\leq t\leq T}\frac{1}{Y_t}>\frac{1}{e^{\frac{1}{3}(\lambda-\mu)T}+1}\,\middle|\,Y_\infty>0\right)+\mathbb{P}(Y_\infty=0|Y_T>0)\tag{A.27}$$

The second summand is smaller that or equal to $\frac{\mu}{\lambda}e^{-(\lambda-\mu)T}$ by Lemma A.1. By Lemma A.4 and the inequality (2) in Theorem 6.14 on page 99 in [Yeh95], the first summand of (A.27) is bounded from above by

$$\mathbb{E}\left(\frac{1}{Y_{T/2}}\,\middle|\,Y_\infty>0\right)(e^{\frac{1}{3}(\lambda-\mu)T}+1)=O(Te^{-\frac{1}{2}(\lambda-\mu)T})(e^{\frac{1}{3}(\lambda-\mu)T}+1)=O(Te^{-\frac{1}{6}(\lambda-\mu)T}),$$

where the first equality follows from Lemma A.9.

What remains is the third summand on the right-hand side of (A.25), which is smaller than or equal to

$$1-\left(1-\frac{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil}{e^{\frac{1}{3}(\lambda-\mu)T}}\right)^{\lceil e^{(\lambda-\mu)T/12}-1\rceil},\tag{A.28}$$

where $\lceil\cdot\rceil$ denotes the ceiling function. Let $T\geq\frac{12\log(2)}{\lambda-\mu}$. Then (A.28) is bounded from above by

$$1-\left(1-\frac{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^4}\right)^{\lceil e^{(\lambda-\mu)T/12}-1\rceil}=1-\left(1-\frac{1}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^3}\right)^{\lceil e^{(\lambda-\mu)T/12}-1\rceil}$$

$$\leq 1-\exp\left(-\frac{1}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^2}\right)+\left|\exp\left(-\frac{1}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^2}\right)-\left(1-\frac{1}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^3}\right)^{\lceil e^{(\lambda-\mu)T/12}-1\rceil}\right|. \tag{A.29}$$

By elementary calculations (see the proof of AE06, Theorem III.6.23) we obtain for all $n \in \mathbb{N}$ and $x \in (0, 1/n)$

$$\left|e^{-x}-\left(1-\frac{x}{n}\right)^n\right| \leq \frac{e}{2}\frac{x^2}{n}.$$

Thus the right-hand side of (A.29) is bounded from above by

$$\frac{1}{\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^2}+\frac{e}{2\lceil e^{\frac{1}{12}(\lambda-\mu)T}-1\rceil^3} \leq \frac{4}{e^{\frac{1}{6}(\lambda-\mu)T}}+\frac{4e}{e^{\frac{1}{4}(\lambda-\mu)T}}=O(e^{-\frac{1}{6}(\lambda-\mu)T}).$$

Altogether, we obtain that the probability that J_T has at least one multiple edge is of the order $O(T^2e^{-\frac{1}{6}(\lambda-\mu)T})$. □

Proof of Corollary 1.4

Recall that $\tilde{\nu}_t$ denotes the distribution of the number of neighbours, ν_t the degree distribution at time t and ν the asymptotic degree distribution. Lemma A.13 implies that $d_{TV}(\tilde{\nu}_t, \nu_t) = O(t^2e^{-\frac{1}{6}(\lambda-\mu)t})$ as $t \to \infty$. From Theorem 1.2, we know that $d_{TV}(\nu_t, \nu) = O\big(t^2e^{-\frac{1}{6}(\lambda-\mu)t}\big)$ as $t \to \infty$. Thus the triangle inequality yields the desired result. □

References

[AE06] Herbert Amann and Joachim Escher. *Analysis I*. Birkhäuser, Basel, 2005.

[AN72] Krishna B Athreya and Peter E Ney. *Branching processes*. Springer, Berlin, Heidelberg, New York, 1972.

[BA99] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[Bai64] Norman TJ Bailey. *The elements of stochastic processes with applications to the natural sciences*. Wiley, New York, 1964.

[BB01] Ginestra Bianconi, A-L Barabási,. *Competition and multiscaling in evolving networks*. *Europhysics Letters*, 54(4), 2001.

[BL10] Tom Britton and Mathias Lindholm. Dynamic random networks in dynamic populations. *Journal of Statistical Physics*, 139(3):518–535, 2010.

[BLT11] Tom Britton, Mathias Lindholm, and Tatyana Turova. A dynamic network in a dynamic population: asymptotic properties. *Journal of Applied Probability*, 48(4):1163–1178, 2011.

[CCDLRM02] Guido Caldarelli, Andrea Capocci, Paolo De Los Rios, Miguel A. Munoz Scale-free networks from varying vertex intrinsic fitness *Physical review letters*, 89(25), 2002.

[CM77] David R Cox and Hilton D Miller. *The theory of stochastic processes*, volume 134. CRC Press, Boca Raton, Florida, 1977.

[Eis08] Bennett Eisenberg. On the expectation of the maximum of IID geometric random variables. *Statistics & Probability Letters*, 78(2):135–143, 2008.

[Har50] Theodore Edward Harris. Some mathematical models for branching processes. Technical report, No. P-152, RAND Corporation, Santa Monica, 1950.

[KS15] Fabian Kück and Dominic Schuhmacher. Convergence Rates for the Degree Distribution in a Dynamic Network Model. Technical Report, 2015. http://arxiv.org/abs/1509.04918.

[KS17] Fabian Kück and Dominic Schuhmacher. On the age of a randomly picked individual in a linear birth and death process. *submitted*, 2017.

[Lam08] Amaury Lambert. Population dynamics and random genealogies. *Stochastic Models*, 24(S1):45–163, 2008.

[NR71] Marcel F Neuts and Sidney I Resnick. On the times of births in a linear birthprocess. *Journal of the Australian Mathematical Society*, 12(04):473–475, 1971.

[PRR13] Erol A. Peköz, Adrian Röllin, and Nathan Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632, 2013.

[SHR13] Erol A. Peköz, Adrian Röllin, and Nathan Ross. Network growth model with intrinsic vertex fitness. *Physical Review E*, 88(1), 2013.

[vdH15] Remco van der Hofstad. *Random Graphs and Complex Networks. Vol. I.* Lecture Notes, 2015. http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf

[Yan91] Nikos Yannaros. Poisson approximation for random sums of Bernoulli random variables. *Statistics & Probability Letters*, 11(2):161–165, 1991.

[Yeh95] James Yeh. *Martingales and stochastic analysis*, Vol. 1. World Scientific, 1995.