

# Convergence Rates in Dynamic Network Models

**Dissertation**

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm „Mathematical Sciences“

der Georg-August University School of Science (GAUSS)

vorgelegt von

**Fabian Kück**

aus Hamburg

Göttingen, 2017

## **Betreuungsausschuss:**

Prof. Dr. Dominic Schuhmacher,  
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Anja Sturm,  
Institut für Mathematische Stochastik, Universität Göttingen

## **Mitglieder der Prüfungskommission:**

### *Referent*

Prof. Dr. Dominic Schuhmacher,  
Institut für Mathematische Stochastik, Universität Göttingen

### *Korreferentin*

Prof. Dr. Anja Sturm,  
Institut für Mathematische Stochastik, Universität Göttingen

### *Weitere Mitglieder der Prüfungskommission*

Jun.-Prof. Dr. Andrea Krajina,  
Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. David Russell Luke,  
Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Prof. Dr. Ralf Meyer  
Mathematisches Institut, Universität Göttingen

Jun.-Prof. Dr. Daniel Rudolf  
Institut für Mathematische Stochastik, Universität Göttingen

**Tag der mündlichen Prüfung:** 4. September 2017

## Abstract

The stochastic network model by Britton and Lindholm [BL10] describes a class of reasonably realistic dynamics for a complex system with an underlying network structure. In a closed social network, which is modeled by a dynamic random graph, the number of individuals evolves according to a linear birth and death process with per-capita birth rate  $\lambda$  and per-capita death rate  $\mu < \lambda$ . A random social index is assigned to each individual at birth, which controls the rate at which connections to other individuals are created. Britton and Lindholm give a somewhat rough proof for the convergence of the degree distribution in this model towards a mixed Poisson distribution. We derive a rate for this convergence giving precise 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. We treat the pure birth case and the general case separately and obtain that the degree distribution converges exponentially fast in time in terms of the total variation distance. We reveal that the degree distribution converges at least of order  $\sqrt{t}e^{-\frac{1}{2}\lambda t}$  in the pure birth case and  $t^2e^{-\frac{1}{6}(\lambda-\mu)t}$  in the general case.

We compare the model to several other network models and find further interesting results for the model. In particular, we show that the asymptotic degree distribution can exhibit power law tails, which makes it an interesting alternative to the famous preferential attachment models.

We finally add a spatial component to the model and find convergence rates for this extended model as well.

We prove several general results about linear birth and death processes along the way. Most notably, we derive the age distribution of an individual picked uniformly at random at some finite time by exploiting a bijection between the birth and death tree and a contour process.



# Acknowledgements

First and foremost, I would like to thank Dominic Schuhmacher for providing the interesting topic of this thesis and his tremendous support and understanding over the years. He guided me in many different ways. His enthusiasm has always been a great encouragement throughout my work. I am very grateful that he always took time to discuss problems with me in a very pleasant way. His suggestions and stimulating ideas helped me greatly in my research. I benefit a lot from his great mathematical intuition and deep understanding on a wide range of areas of stochastics. His valuable contributions were fundamental to the completion of this thesis.

I am very grateful that Anja Sturm agreed to be my second advisor. Furthermore, I would like to thank Andrea Krajina, Russell Luke, Ralf Meyer and Daniel Rudolf for being part of my examination committee.

I would like to express my gratitude to the current and former members of the Institute for Mathematical Stochastics who created a wonderful atmosphere.

I would like to thank Thomas Rippl for drawing my attention to the connection between birth and death and contour processes, which led to the main results of Chapter 4 of this thesis.

I am very grateful to my friends who supported me greatly in various ways in the last years, especially to those who helped me with their useful remarks concerning this thesis.

I would like to express my heartfelt thanks to my family, in particular my parents, for their constant unconditional support through the years. Finally, I would like to thank my girlfriend, Lara, for her patience and for being extraordinarily supportive.



# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| <b>2</b> | <b>Distributions related to linear birth and death processes</b>                          | <b>7</b>  |
| 2.1      | Introduction to linear birth and death processes . . . . .                                | 7         |
| 2.2      | Total population size in a linear birth and death process . . . . .                       | 9         |
| 2.3      | The number of births and deaths . . . . .   | 12        |
| <b>3</b> | <b>Bounds on the total variation distance for Poisson and mixed Poisson approximation</b> | <b>15</b> |
| 3.1      | Probability metrics and the total variation distance . . . . .                            | 15        |
| 3.1.1    | Probability metrics . . . . .   | 15        |
| 3.1.2    | The total variation distance . . . . .  | 16        |
| 3.2      | Poisson approximation . . . . .   | 20        |
| 3.3      | Mixed Poisson distributions . . . . .   | 22        |
| 3.4      | Upper bound for the total variation distance between two mixed Poisson distributions      | 24        |
| <b>4</b> | <b>The age of a randomly picked individual in a linear birth and death process</b>        | <b>27</b> |
| 4.1      | Related results . . . . .   | 27        |
| 4.1.1    | The number of individuals with age in a certain interval . . . . .                        | 27        |
| 4.1.2    | Times since the birth times . . . . .   | 28        |
| 4.2      | The age distribution in the pure birth case . . . . .                                     | 28        |
| 4.3      | The age distribution in the general case . . . . .  | 30        |
| 4.3.1    | Results . . . . .   | 30        |
| 4.3.2    | Bijection between piecewise-linear functions and trees . . . . .                          | 32        |
| 4.3.3    | Probability laws of the contour process and the corresponding random tree . .             | 34        |
| 4.3.4    | Proof of Theorem 4.3.1 . . . . .  | 35        |
| 4.3.5    | Proof of Corollary 4.3.3 . . . . .  | 42        |
| 4.3.6    | Proof of Corollary 4.3.5 . . . . .  | 43        |
| 4.3.7    | Proof of Corollary 4.3.6 . . . . .  | 44        |
| 4.3.8    | Simulations . . . . .   | 45        |
| 4.4      | Upper bound for the time since the last event . . . . .                                   | 47        |
| 4.4.1    | The main result . . . . .   | 47        |
| 4.4.2    | Comparison with results about reconstructed trees . . . . .                               | 48        |

---

|          |   |            |
|----------|---|------------|
| <b>5</b> | <b>Random networks</b>  | <b>53</b>  |
| 5.1      | Introduction to graphs . . . . .  | 53         |
| 5.2      | Heavy tailed and power law distributions . . . . .  | 55         |
| 5.3      | Static random network models . . . . .  | 56         |
| 5.3.1    | The Erdős-Rényi model and related models . . . . .  | 56         |
| 5.3.2    | The configuration model . . . . .   | 61         |
| 5.3.3    | Inhomogeneous random graphs . . . . .   | 65         |
| 5.3.4    | Further static random network models . . . . .  | 65         |
| 5.4      | Dynamic random network models . . . . .   | 65         |
| 5.4.1    | Preferential attachment . . . . .   | 65         |
| 5.4.2    | Uniform attachment . . . . .  | 69         |
| 5.4.3    | Fitness models . . . . .  | 70         |
| 5.4.4    | Node and edge deletion . . . . .  | 72         |
| 5.4.5    | A further dynamic network model . . . . .   | 73         |
| <b>6</b> | <b>The network model by Britton and Lindholm</b>  | <b>75</b>  |
| 6.1      | The model . . . . .   | 75         |
| 6.2      | Degree distribution at finite time . . . . .  | 76         |
| 6.2.1    | The number of outgoing edges in the pure birth case . . . . .   | 77         |
| 6.2.2    | The number of incoming edges in the pure birth case . . . . .   | 78         |
| 6.2.3    | Degree distribution in the pure birth case . . . . .  | 79         |
| 6.2.4    | Degree distribution in the general case . . . . .   | 80         |
| 6.3      | Bounds on the total variation distance between the finite time and asymptotic degree distribution . . . . . | 81         |
| 6.3.1    | The pure birth case . . . . .   | 81         |
| 6.3.2    | The general case . . . . .  | 86         |
| 6.3.3    | Sharpness of the convergence rates . . . . .  | 112        |
| 6.4      | Negligibility of multiple edges . . . . .   | 117        |
| 6.5      | Heavy tails and power laws of the asymptotic degree distribution . . . . .                                  | 122        |
| 6.5.1    | Heavy tails of the asymptotic degree distribution . . . . .   | 122        |
| 6.5.2    | Power law behaviour of the asymptotic degree distribution . . . . .   | 124        |
| 6.6      | Further discussion of the asymptotic degree distribution . . . . .  | 126        |
| 6.7      | The pure birth case with immortal edges . . . . .   | 127        |
| 6.7.1    | The number of outgoing edges in the pure birth case with immortal edges . . . . .                           | 127        |
| 6.7.2    | The number of incoming edges in the pure birth case with immortal edges . . . . .                           | 128        |
| 6.7.3    | The degree distribution in the pure birth case with immortal edges . . . . .                                | 128        |
| <b>7</b> | <b>A spatial network model</b>  | <b>131</b> |
| 7.1      | Model and main result . . . . .   | 131        |
| 7.2      | Proof of the main theorem . . . . .   | 133        |



|          |   |            |
|----------|---|------------|
| 7.3      | Examples . . . . .  | 136        |
| 7.3.1    | The "hardcore" case . . . . .                             | 136        |
| 7.3.2    | Edge probability depending linearly on distance . . . . . | 139        |
| 7.3.3    | Gaussian kernels . . . . .                                | 141        |
| <b>8</b> | <b>Concluding remarks</b>                                 | <b>145</b> |
| <b>A</b> | <b>Appendix</b>   | <b>147</b> |
| A.1      | R code for the introduction . . . . .                     | 147        |
| A.2      | R code for Chapter 4 . . . . .                            | 148        |
| A.3      | R code for Chapter 6 . . . . .                            | 153        |
| A.4      | R code for Chapter 7 . . . . .                            | 158        |
|          | <b>Bibliography</b>                                       | <b>161</b> |
|          | <b>List of Symbols</b>                                    | <b>166</b> |



# 1 Introduction

Network Science is a relatively young research area dealing with complex systems with an underlying network structure. In our well-connected modern world, almost every aspect of life is related to interactions that could be modeled by using tools from network science. In the recent decades, researchers have consequently examined a vast amount of different network models applied in various scientific fields. The diversity of network science makes it one of the most challenging, but also most fruitful, interdisciplinary research areas. In particular, using graphs as abstract mathematical model for real-world networks in diverse areas lead to the discovery of surprising similarities in structures from very different contexts that were thought to be unrelated. Intuitively, a graph can be thought of as a collection of points, called nodes, which may be connected to other points (and sometimes even to themselves) by lines, called edges (see Chapter 5 for a precise definition of graphs and an introduction to the subject). The nodes represent for example individuals in a population whereas the edges represent some kind of relations between the nodes. Since we are interested in complex systems, we cannot expect to be able to predict all relevant aspects with certainty. Thus we consider random graph models.

Unfortunately, there is no well-categorized catalogue of such models. However, there are well-known properties that can be observed in many real-world networks and thus in the corresponding models. Consequently, it is reasonable to search for an existing model with the desired properties in order to find a satisfying model for a given application. We state some of the most important network models and their properties now (see Chapter 5 for a more extensive introduction to random graphs).

One of the first random network models is the well-known Erdős-Rényi model, which was introduced in 1959 by Gilbert [Gil59] and named after Erdős and Rényi, who published a closely related model in the same year [ER59]. In this model, each pair of nodes is connected with some fixed probability  $p$  independently from other pairs of nodes (see also Chapter 5). The model is a static model, i.e. it describes a random network at a fixed time.

Although it shows interesting behaviour, further random network models were needed for the description of real networks since many empirical networks have important properties that are not represented by Erdős-Rényi graphs. In particular, they often have a very different degree distribution, where the degree of a node is the number of its edges. Many of the degree distributions observed in real-world networks have the so-called power law property, which means that the corresponding probability mass function decreases (approximately) at a polynomial rate. This means that high values appear with relatively high probability. Note that power law distributions do not only appear as degree distributions in random networks, but also in very different contexts (see e.g. [New13] for a great variety of examples). Examples for real-world networks having power law degree distribution are the internet, collaboration networks of scientists as well as movie actors and networks of telephone

calls, emails and human sexual contacts (see Chapter 3 of [DM03]). It is remarkable that some of those have no obvious relation to each other.

There are some generalizations of Erdős-Rényi graphs that produce power law degree distributions. For example, in the generalized random graph with  $n$  nodes introduced in [vdH16] (see also Chapter 5), each node is assigned a weight, and each pair of nodes is connected with a probability depending on the weights of those two nodes in a particular way. This model allows us to obtain any mixed Poisson distribution as asymptotic degree distribution as  $n \rightarrow \infty$ , which is a large class of distributions and, in particular, contains power law distributions (see Section 3.3 for an introduction to this class of distributions).

A further popular static random network model is the configuration model. This model gives us a random graph for any given feasible degree distribution, in particular for power law distributions. This is done by equipping each node with an appropriate number of half-edges, which are joined together in a random way (see Chapter 5 for details).

A lot of very popular classical, static network models are rather designed to obtain a snapshot of a network that depicts the properties of a complex system in a reasonable way and are often not sufficient to describe the behaviour that leads to the observation of these properties. An example is the configuration model mentioned above. Time-discrete dynamic models allow interesting insights in such underlying behaviour. The most popular class of such models are the preferential attachment models that exhibit a power law behaviour asymptotically and were popularized by Barabási and Albert [BA99]. In such models the popularity of a node is proportional to its degree.

However, “a line of research towards a naturally comprehensible explanation for the formation of power-law networks has argued that degree is not the only key factor influencing the network growth” ([TP11], Chapter 2). Therefore, “the ‘inner self’ factors such as the personality of a person in a friendship network” ([TP11], Section 2.5) or “talent” ([TP11], Section 2.7) should be considered in network models. Barabási and Albert argue that for example the extremely high degree of Google and Facebook in the World Wide Web network cannot be explained by the preferential attachment mechanism since in a preferential attachment model nodes with such a high degree are among the oldest nodes with very high probability, which is not the case for those two examples (see Section 6.1 of [Bar13]).

Consequently, 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 and stays the same for the whole lifetime. 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 [CCRM02] and [SHR13]. Depending on the application, either preferential attachment, models that combine preferential attachment as well as fitness or pure fitness models may be more suitable.

Discrete-time preferential attachment and fitness models cover a lot of aspects of real-world networks. However, the discreteness is far from being realistic. Therefore, we treat models that incor-

porate exponential waiting times, which makes it more realistic. Note that this was also done for a continuous-time preferential attachment model considered in [Rei09] (see Subsection 5.4.1 for details).

In this thesis, we focus on the time-continuous random graph model that was introduced by Britton and Lindholm [BL10], which can be seen as time-continuous pure fitness model that is particularly realistic for the possibility of nodes as well edges to die. We prove in this thesis among other things that the asymptotic degree distribution in this model can exhibit power laws such that it displays an interesting alternative mechanism for producing networks with this property (see Section 6.5).

We sketch a slightly adapted, namely a *loop-free*, version of the original *dynamic network model* by Britton and Lindholm (see [BL10] and [BLT11]) here that we consider in large parts of this thesis. We refer to Chapter 6 for a precise definition of the model.

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  $\lambda$  and dies at constant rate  $\mu$ .

We assume  $\lambda > \mu$ , so that  $Y_t \rightarrow \infty$  as  $t \rightarrow \infty$  with positive probability (see Chapter 2 for details).

We equip every node  $i$  with a positive random social index  $S_i$  representing its fitness, where the  $(S_i)_{i \in \mathbb{N}}$  are independent and identically distributed with finite expectation.

At birth every node is isolated. During its lifetime and *as long as there is at least one other node*, node  $i$  generates edges at rate  $\alpha S_i$ , which are removed after  $\text{Exp}(\beta)$  distributed times. Here  $\alpha$  and  $\beta$  are positive constants. The “second” node of each newly born edge is chosen uniformly at random from the set of all *other* living nodes.

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

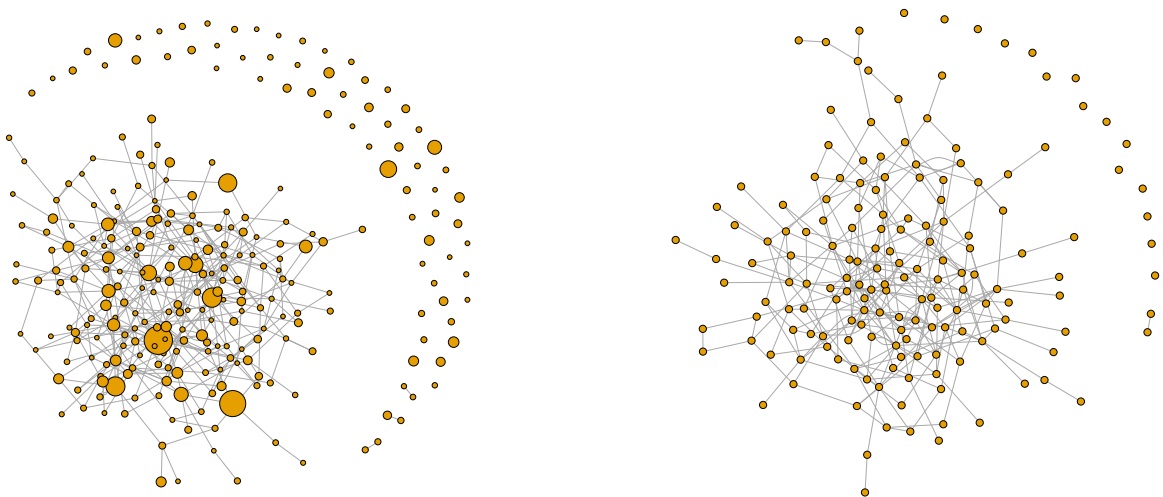


Figure 1.1: Simulated realizations of the Britton-Lindholm model with  $\alpha = 3$ ,  $\beta = \lambda = 1$ ,  $\mu = 0$  and Pareto(3, 2/3) distributed (left-hand side) and deterministic (right-hand side) social indices, respectively, for  $T = 5$ , where the size of the nodes corresponds to the social indices.

The model is illustrated in Figure 1.1.

### 1.0.1 Remark

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 further that we allow multiple edges, i.e. pairs of nodes that are connected by more than one edge (see Chapter 5 for a precise definition). Depending on the real-world application, this may be less or more realistic than the corresponding model without multiple edges. It can be shown that those edges 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 Section 6.4). This allows us to formulate the main result also for the case where we ignore multiple edges (see Corollary 1.0.3 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  $\nu_t$ . In [BL10], Britton and Lindholm give a rather heuristic argument for the weak convergence of the degree distribution in the original model towards a mixed Poisson distribution  $\nu$ . One of the results of this thesis gives a rate in total variation distance rather than a mere convergence result, where the total variation distance is a very common distance measure for probability distributions (see Section 3.1 for details). We provide a complete proof for this rate and thereby also for the weak convergence since convergence in total variation distance implies weak convergence (see Proposition 3.1.9).

The asymptotic distribution  $\nu$  is given by

$$\nu = \text{MixPo}\left(\frac{\alpha}{\beta + \mu}(S + \mathbb{E}(S))(1 - e^{-(\beta + \mu)A})\right),$$

where  $A \sim \text{Exp}(\lambda)$ ,  $S$  has the social index distribution, and  $A$  and  $S$  are independent. Here MixPo denotes the mixed Poisson distribution. Note that  $\nu$  is the same asymptotic degree distribution as in [BL10].

The following result is an immediate consequence of Theorems 6.3.2 and 6.3.4, which are proved in this thesis.

### 1.0.2 Theorem

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

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

Theorem 1.0.2 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  $\nu$  is an immediate consequence of the following corollary, which is proved in Section 6.4.

### 1.0.3 Corollary

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  $t \rightarrow \infty$ .

Note that rates for the convergence towards the degree distribution were considered for discrete-time random network models as well. Reinert [Rei09] and Peköz, Röllin and Ross [PRR13] established convergence rates for the degree distribution in preferential attachment models in terms of the total variation distance (see Chapter 5 for details).

The derivation of the convergence rates for the (loop-free version of the) Britton and Lindholm is a challenging task due to its complexity. We treat the pure birth and the general case separately because the pure birth case allows us to obtain a much better rate and present some of the main ideas of the proof much clearer.

The proof is based on a general approximation result that gives us a bound for the total variation distance between two mixed Poisson distributions. We derive this result, which might also be interesting in other contexts, in Chapter 3. We apply it to the total variation distance between the degree distribution  $\nu_t$  at time  $t$ , which we show to be a mixed Poisson distribution in Chapter 6, and the asymptotic degree distribution  $\nu$ . Let  $(\Lambda_t^*)_{t \geq 0}$  and  $M^*$  be random variables such that  $\nu_t = \text{MixPo}(\Lambda_t^*)$  and  $\nu = \text{MixPo}(M^*)$ . Then this strategy leads to  $\mathbb{E}(|\Lambda_t^* - M^*|)$  as upper bound for  $d_{TV}(\nu_t, \nu)$ .

In order to bound this expectation, we first couple  $\Lambda_t^*$  and  $M^*$  in an appropriate way. Since the distribution of  $\Lambda_t^*$  depends on the distribution of the age of a node picked uniformly at random at time  $t$ , we are interested in the distribution of the age  $A$  of an individual picked uniformly at random at time  $t$  in a linear birth and death process  $(Y_t)_{t \geq 0}$  with per-capita birth rate  $\lambda$  and per-capita death rate  $\mu$ . In the pure birth case, i.e. if  $\mu = 0$ , it is well-known that this distribution is essentially a truncated exponential with parameter  $\lambda$ , a result which follows immediately from Theorem 1 in [NR71] (see Chapter 4 for details). It is furthermore well-known that  $\mathcal{L}(A | Y_t > 0) \xrightarrow{w} \text{Exp}(\lambda)$  as  $t \rightarrow \infty$  if  $\lambda > \mu \geq 0$ , where  $\xrightarrow{w}$  denotes weak convergence. This follows e.g. from Example (6.10.14) in [Jag75].

However, somewhat surprisingly, if  $\mu > 0$ , an exact formula for the distribution of  $A$  at finite  $t$  is nowhere to be found in the literature. In this thesis, we provide such a formula for the cumulative distribution function of  $A$ , both conditionally on the number  $Y_t$  of individuals at time  $t$  and unconditionally (the conditioning on  $Y_t > 0$  being always tacitly implied). Our main proof idea relies on a bijection between Galton–Watson trees in continuous time and exploration processes, recently shown in [BPS12]. We also give upper bounds on the closeness of  $\mathcal{L}(A | Y_t > 0)$  and  $\text{Exp}(\lambda)$  if  $\lambda \geq \mu$  as well as convergence rates in various metrics. Note that these substantial new results can be of great general interest. We point out some related work in Chapter 4.

Finally, we also add a spatial component to the (loop-free version of the) Britton-Lindholm model. More precisely, we let the probability of edges depend on the positions of the nodes, which gives us a even more realistic model. For simplicity, we restrict ourselves to the pure birth case here. We derive asymptotic degree distributions and corresponding convergence rates analogously to the non-spatial case.

The rest of the thesis is organized as follows. In Chapter 2, we extensively discuss linear birth and death processes and consider distributions related to such processes since the node population in the Britton-Lindholm model is described by such a process. In particular, we deduce formulae for the expectation and the variance of the number of births, which are also of some general interest.

In Chapter 3, we first introduce probability metrics, in particular the total variation distance. Secondly, we give a short introduction to Poisson approximation in Section 3.2. In Section 3.3, we define mixed Poisson distributions and state interesting properties. Finally, we derive a universal bound for the total variation distance between two general mixed Poisson distributions (see Theorem 3.4.1), which we apply to the Britton-Lindholm model without loops in Section 6.3.

In Chapter 4, we first consider known results about concepts that are in some sense similar to “our” age distribution. Then we briefly consider the age distribution in the pure birth case in Section 4.2. In Section 4.3, we treat the general case, which is far more complicated. We introduce contour processes and derive the age distribution for a linear birth and death process using results about those processes from [BPS12]. Additionally, we deduce rates for the convergence towards the asymptotic age distribution. In Section 4.4, we compute an upper bound for the time since the last event in a linear birth and death process by applying results from the previous section and compare this bound with a similar one obtained from known results about so-called reconstructed trees.

In Chapter 5, we turn to random networks. We first introduce briefly the concept of graphs and heavy-tailed as well as power law distributions. Then we present well-known examples of static random network models in Section 5.3. Finally, we consider well-known dynamic random network models, including preferential and uniform attachment and fitness models. In particular, we summarize known results about rates for the convergence towards the asymptotic degree distribution for preferential and uniform attachment models.

In Chapter 6, we consider the (loop-free version of the) network model by Britton and Lindholm. Firstly, we give a precise definition of the model sketched above. Then we deduce the degree distribution at finite time by using results about birth and death processes in Section 6.2. In Section 6.3, we derive rates for the convergence to the asymptotic degree distribution in terms of the total variation distance treating the pure birth case and the general case separately, which leads to Theorem 1.0.2 above. At the end of Section 6.3, we also discuss the quality of the convergence rates by using simulations. In Section 6.4, we prove that multiple edges are negligible in the sense that the probability that a node picked uniformly at random has a multiple edge converges to zero exponentially in time, which allows us to formulate our main results for the case where we ignore multiple edges as mentioned above. In Section 6.5, we show that we can indeed obtain a power law distribution as asymptotic degree distribution, which motivates the consideration of the model. We briefly consider further properties of the asymptotic degree distribution in Section 6.6. Finally, we look at the case where edges cannot die and obtain similar results as before. This case turns out to be very interesting since we can rediscover asymptotic degree distributions from other well-known random network models.

Finally, we add a spatial component to the model in Chapter 7. More precisely, we let the probability of an edge depend on the positions of the nodes, which are assumed to be uniformly distributed on a hypercube of arbitrary dimension. We derive a rate of convergence towards the asymptotic degree distribution and show that multiple edges are negligible in a rather general setting. Finally, we specify how the probability of the creation of an edge depends on the distance between the corresponding nodes in three different ways and obtain explicit expressions for the degree distributions and the convergence rates.

In Chapter 8, we end this thesis with a discussion and an outlook.



# 2 Distributions related to linear birth and death processes

## 2.1 Introduction to linear birth and death processes

A linear birth and death process is a Markov process that is applied in many different areas in order to describe populations. Before we define this type of process, we recall the definition of a birth and death process.

**2.1.1 Definition** (see e.g. Section 6.2 of [All10])

A *birth and death process*  $(X_t)_{t \geq 0}$  is a homogeneous Markov process with states in (a subset of)  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$  that satisfies

$$\mathbb{P}(X_{t+\Delta t} - X_t = m | X_t = n) = \begin{cases} \lambda_n \Delta t + o(\Delta t) & \text{if } m = 1, \\ \mu_n \Delta t + o(\Delta t) & \text{if } m = -1, \\ 1 - (\lambda_n + \mu_n) \Delta t + o(\Delta t) & \text{if } m = 0, \\ o(\Delta t) & \text{otherwise} \end{cases}$$

as  $\Delta t \downarrow 0$  for  $n, m \in \mathbb{N}_0$ ,  $t \geq 0$ , where  $\lambda_n \geq 0$  and  $\mu_{n+1} \geq 0$  for all  $n \in \mathbb{N}_0$ .

We interpret  $X_t$  as population size at time  $t$ . Note that the sequences  $(\lambda_n)_{n \in \mathbb{N}_0}$  and  $(\mu_n)_{n \in \mathbb{N}}$  are often referred to as birth and death rates, respectively. In this thesis, all homogeneous Markov processes are assumed to have right-continuous trajectories.

### 2.1.2 Remark

From Section XVII.5 of [Fel67], we know that a birth and death process  $(X_t)_{t \geq 0}$  exists for all feasible sequences  $(\lambda_n)_{n \in \mathbb{N}_0}$  and  $(\mu_n)_{n \in \mathbb{N}}$ . Furthermore, the birth and death process  $(X_t)_{t \geq 0}$  is uniquely determined by the above definition in “all cases of practical significance” ([Fel67], Section XVII.5) including the following examples.

A Poisson process, which is defined in the following definition, is a birth and death process.

**2.1.3 Definition** (cf. e.g. Example 1.3 in [CM77])

A birth and death process with constant birth rates  $\lambda_n = \tilde{\lambda} > 0$  and death rates  $\mu_{n+1} = 0$  for all  $n \in \mathbb{N}_0$  started with zero individuals is called *Poisson process* of rate  $\tilde{\lambda}$ .

**2.1.4 Remark** (cf. e.g. Example 1.3 in [CM77])

For a Poisson process  $(\Xi_t)_{t \geq 0}$  of rate  $\tilde{\lambda}$ , it can be shown that the number of events, i.e. births, in a time interval  $(t, t + \Delta t]$  is  $\text{Po}(\tilde{\lambda} \Delta t)$  distributed for all  $t \geq 0$  and  $\Delta t > 0$ . In particular, we have  $\Xi_t \sim \text{Po}(\tilde{\lambda} t)$  for all  $t > 0$ .

A linear birth and death process is a birth and death process whose birth and death rates are linear in the population size, i.e. it is defined and interpreted as follows.

**2.1.5 Definition and Remark** (see e.g. Example 4.4 in [CM77] and Section 8.2 of [Bai64])

A *linear birth and death process*  $(Y_t)_{t \geq 0}$  is a birth and death process with  $\lambda_n = \lambda n$  and  $\mu_n = \mu n$ ,  $n \in \mathbb{N}_0$ , for  $\mu, \lambda \geq 0$ . Throughout this thesis,  $(Y_t)_{t \geq 0}$  always denotes such process.

A linear birth and death process can be interpreted as follows. For each  $t \geq 0$ , the random variable  $Y_t$  is the number of individuals alive in a population where each living individual gives birth to offspring according to a Poisson process of rate  $\lambda$  and lives an  $\text{Exp}(\mu)$  distributed time independently of all other individuals. Thus we call  $\lambda$  and  $\mu$  the per-capita birth and death rate, respectively.

A different interpretation is that each living individual splits into two after an  $\text{Exp}(\lambda)$  distributed time. In particular, this is reasonable if we think of unicellular organisms. However, we stick to the first interpretation throughout this thesis.

Note that a linear birth and death process is also often referred to as simple birth and death process in the older literature (see e.g. Section 8.2 of [Bai64]). Note further that linear birth and death processes belong to the class of continuous-time branching processes (see e.g. see Section III.5 of [AN72] or page 62 in [HJV05]) since they fulfill the following *branching property*.

**2.1.6 Lemma** (cf. e.g. page 105 and 106 in [AN72])

Let  $Y_u^{(i)}(t, \omega)$  denote the number of offspring of the  $i$ th of the  $Y_t(\omega)$  individuals alive at time  $t$  that are still living at time  $t + u$ , where we number the parents in an arbitrary way. Then we have

$$Y_{t+u}(\omega) = \sum_{i=1}^{Y_t(\omega)} Y_u^{(i)}(t, \omega),$$

where, given  $Y_t$ , the processes  $(Y_u^{(i)}(t))_{u \geq 0}$  are independent with  $\mathcal{L}((Y_u^{(i)}(t))_{u \geq 0} | Y_t) = \mathcal{L}((Y_u)_{u \geq 0} | Y_0 = 1)$  (and we use the usual convention that the empty sum is zero).

Finally, we consider a classical classification of linear birth and death processes.

**2.1.7 Remark** (see e.g. [AN72] or [HJV05])

For  $(Y_t)_{t \geq 0}$ , the expected number of offspring of each individual is

$$\int_0^{\infty} \lambda t \mathbb{P}L^{-1}(dt) = \int_0^{\infty} \lambda e^{-\mu t} dt = \frac{\lambda}{\mu},$$

where  $L$  is the lifetime of the individual. A branching processes is called *critical* if this expectation is equal to one, *subcritical* if it is smaller and *supercritical* if it is larger than one. Thus we call  $(Y_t)_{t \geq 0}$  *critical* if  $\lambda = \mu$ , *subcritical* if  $\lambda < \mu$  and *supercritical* if  $\lambda > \mu$ .

From now on, we assume  $Y_0 = 1$ , i.e. we begin with one individual at time 0, throughout this thesis if not stated otherwise. For a general initial population size, many results can be deduced from this case by using the branching property from Lemma 2.1.6.

## 2.2 Total population size in a linear birth and death process

In this section, we consider the distribution of the population size  $Y_t$  in the linear birth and death process  $(Y_t)_{t \geq 0}$  with per-capita birth rate  $\lambda$ , per-capita death rate  $\mu$  and initial value one, i.e.  $Y_0 = 1$ . In the network models in Chapter 6 and Chapter 7, the node process is a supercritical linear birth and death process. Since some results could also be of general interest, we also consider the case where  $\mu \geq \lambda > 0$ . Firstly, we assume that the per-capita birth rate  $\lambda$  and the per-capita death rate  $\mu$  of  $(Y_t)_{t \geq 0}$  are not equal. According to (8.15) and (8.46) in [Bai64], the one-dimensional distributions of such a process are then given by the following probability mass functions:

$$\begin{aligned} p_0(t) &= \mu \tilde{p}(t), \\ p_n(t) &= (1 - \mu \tilde{p}(t))(1 - \lambda \tilde{p}(t))(\lambda \tilde{p}(t))^{n-1}, \quad n \in \mathbb{N} = \{1, 2, 3, \dots\}, \end{aligned}$$

where

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

### 2.2.1 Remark

Note that these are just the probability mass functions of geometric distributions if  $\mu = 0$ . Since the convolution of several geometric distributions is a negative binomial distribution, this implies by Lemma 2.1.6 that the one-dimensional distributions of  $(Y_t)_{t \geq 0}$  are negative binomial distributions in the pure birth case if we consider a general initial value  $m \in \mathbb{N}$ .

For  $\lambda = \mu > 0$ , the probability mass function of the one-dimensional distributions of  $(Y_t)_{t \geq 0}$  is given by

$$\begin{aligned} p_0(t) &= \frac{\lambda t}{1 + \lambda t}, \\ p_n(t) &= \frac{(\lambda t)^{n-1}}{(1 + \lambda t)^{n+1}}, \quad n \in \mathbb{N}, \end{aligned}$$

by (8.53) in [Bai64].

### 2.2.2 Remark

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 (see Lemma 2.1.6), we have that  $p_0(t)^m$  is the probability that a linear birth and death process with a general initial value  $m \in \mathbb{N}$  goes extinct up to time  $t$ . By taking the limit  $t \rightarrow \infty$ , we obtain that the probability of eventual extinction is  $(\mu/\lambda)^m$  for  $\lambda \geq \mu$  and 1 otherwise (see also (8.59) in [Bai64]).

By elementary computations using these probability mass functions, we obtain the following proposition (cf. (8.16), (8.17), (8.48), (8.49), (8.54) and (8.55) in [Bai64]).

### 2.2.3 Proposition

(i) For  $\lambda \neq \mu$ , we have

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

(ii) For  $\lambda = \mu > 0$ , we have

$$\mathbb{E}(Y_t) = 1 \text{ and } \text{Var}(Y_t) = 2\lambda t.$$

Now we derive the expected reciprocal of the population size at time  $t$  given the population has not gone extinct until time  $t$  and state an upper bound. This result is required for the derivation of the age distribution in Chapter 4 as well as for the proofs of the convergence rates in Chapter 6 and Chapter 7.

### 2.2.4 Proposition

(i) For  $\lambda \neq \mu$ , we have

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

(ii) For  $\lambda = \mu > 0$  we have,

$$\mathbb{E}\left(\frac{1}{Y_t} \mid Y_t > 0\right) = \frac{\log(1 + \lambda t)}{\lambda t}.$$

*Proof:*

(i) We obtain for  $\lambda \neq \mu$

$$\begin{aligned} \mathbb{E}\left(\frac{1}{Y_t} \mid Y_t > 0\right) &= \frac{\sum_{n=1}^{\infty} \frac{p_n(t)}{n}}{1 - p_0(t)} \\ &= \sum_{n=1}^{\infty} \frac{1}{n} \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \mu} \left(\frac{\lambda e^{(\lambda - \mu)t} - \lambda}{\lambda e^{(\lambda - \mu)t} - \mu}\right)^{n-1} \\ &= \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \lambda} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{\lambda e^{(\lambda - \mu)t} - \lambda}{\lambda e^{(\lambda - \mu)t} - \mu}\right)^n \\ &= \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \lambda} \left(-\log\left(1 - \frac{\lambda e^{(\lambda - \mu)t} - \lambda}{\lambda e^{(\lambda - \mu)t} - \mu}\right)\right) \\ &= \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \lambda} \log\left(\frac{\lambda e^{(\lambda - \mu)t} - \mu}{\lambda - \mu}\right) \\ &\leq \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)t} - \lambda} \left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)t\right). \end{aligned}$$

(ii) Analogously, we obtain for  $\lambda = \mu > 0$

$$\mathbb{E}\left(\frac{1}{Y_t} \mid Y_t > 0\right) = \frac{\sum_{n=1}^{\infty} \frac{p_n(t)}{n}}{1 - p_0(t)} = (1 + \lambda t) \sum_{n=1}^{\infty} \frac{1}{n} \frac{(\lambda t)^{n-1}}{(1 + \lambda t)^{n+1}} = \frac{1}{\lambda t} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{\lambda t}{1 + \lambda t}\right)^n = \frac{\log(1 + \lambda t)}{\lambda t}.$$

□

For the time being, we only consider the case where  $\lambda > \mu$ . In this case, we also need an upper bound for  $\mathbb{E}(1/\sqrt{Y_t} \mid Y_t > 0)$  in Chapter 6 and Chapter 7, which is given by the following corollary.

**2.2.5 Corollary**

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

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

*Proof:* For  $\lambda > \mu$  and  $t \geq \frac{1}{\lambda - \mu} \log(2)$ , it follows from Proposition 2.2.4:

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

□

**2.2.6 Remark**

Using Jensen's inequality, we obtain an asymptotic lower bound for the conditional expectation considered in the previous corollary:

By the formula for  $p_0(t)$  given at the beginning of this section, it follows that

$$\lim_{t \rightarrow \infty} \mathbb{E}(Y_t | Y_t > 0) e^{-(\lambda - \mu)t} = \lim_{t \rightarrow \infty} \mathbb{E}(Y_t) e^{-(\lambda - \mu)t} \frac{\lambda}{\lambda - \mu} = \frac{\lambda}{\lambda - \mu} \quad (2.1)$$

for  $\lambda > \mu$ , and since the function  $x \mapsto \frac{1}{\sqrt{x}}$  is strictly convex on the positive real axis, Jensen's inequality yields

$$\mathbb{E}\left(\frac{1}{\sqrt{Y_t}} \mid Y_t > 0\right) > \frac{1}{\sqrt{\mathbb{E}(Y_t | Y_t > 0)}}. \quad (2.2)$$

Equation (2.1) implies

$$\lim_{t \rightarrow \infty} \frac{1}{\sqrt{\mathbb{E}(Y_t | Y_t > 0)}} e^{\frac{1}{2}(\lambda - \mu)t} = \sqrt{\frac{\lambda - \mu}{\lambda}}.$$

Thus we obtain by (2.2) that  $\mathbb{E}\left(\frac{1}{\sqrt{Y_t}} \mid Y_t > 0\right)$  does not converge to zero at a faster rate than  $e^{-\frac{1}{2}(\lambda - \mu)t}$  as  $t \rightarrow \infty$  for  $\lambda > \mu$ .

Now we consider the martingale property of the normed population size, which leads to useful results about the limiting behaviour.

**2.2.7 Theorem** (cf. Theorem 1 in Section III.7 of [AN72])

The process  $(Y_t e^{-(\lambda - \mu)t})_{t \geq 0}$  is a martingale with respect to the natural filtration.

*Proof:* By Lemma 2.1.6 and Proposition 2.2.3, we have for  $u < t$

$$\mathbb{E}(Y_t e^{-(\lambda-\mu)t} | Y_u) = e^{-(\lambda-\mu)t} Y_u \mathbb{E}(Y_{t-u} | Y_0 = 1) = Y_u e^{-(\lambda-\mu)t} e^{(\lambda-\mu)(t-u)} = Y_u e^{-(\lambda-\mu)u}.$$

□

The following corollary is an immediate consequence of Theorem 2.2.7.

### 2.2.8 Corollary

The process  $(Y_t)_{t \geq 0}$  is a submartingale if  $\lambda > \mu$ , a supermartingale if  $\lambda < \mu$  and a martingale if  $\lambda = \mu$  with respect to the natural filtration.

Since  $(Y_t e^{-(\lambda-\mu)t})_{t \geq 0}$  is a martingale by Theorem 2.2.7, the random variable  $\tilde{W} := \lim_{t \rightarrow \infty} Y_t e^{-(\lambda-\mu)t}$  exists almost surely by standard martingale convergence theory. In the subcritical and critical case, the probability of ultimate extinction is one (see Remark 2.2.2). Thus we have  $\tilde{W} = 0$  almost surely. However, in the supercritical case, we obtain a non-deterministic random variable. Namely, its distribution is given by  $\mathbb{P}(\tilde{W} = 0) = \mu/\lambda$  and  $\mathcal{L}(\tilde{W} | \tilde{W} > 0) = \text{Exp}(\frac{\lambda-\mu}{\lambda})$  (see e.g. page 319 in [Har50]). Note that this shows that the statement at the beginning of Section 4 of [Kei75] is wrong.

## 2.3 The number of births and deaths

We consider a linear birth and death process  $(Y_t)_{t \geq 0}$  with birth rate  $\lambda$ , death rate  $\mu$  and initial value one as before. For convenience, we assume  $\lambda \neq \mu$  in this section.

Let  $\mathfrak{B}_t$  be the *number of births* up to time  $t$ . Then  $(\mathfrak{B}_t)_{t \geq 0}$  is in general no Markov process since the number of births depends on the population size. However, the two-dimensional process  $(\mathfrak{B}_t, Y_t)_{t \geq 0}$  is always a Markov process (see Example 6.4 in [CM77]). Let  $\mathcal{G}(w, z; t) := \mathbb{E}(w^{\mathfrak{B}_t} z^{Y_t})$  be the joint probability generating function. Then a partial differential equation can be derived for  $\mathcal{G}$  (see (39) on page 265 in [CM77] or (43) in [Ken49]), which can be used to compute  $\mathcal{G}$  (see (40) on page 266 in [CM77]). Moreover, the partial differential equation for  $\mathcal{G}$  leads to the following partial differential equation for the joint cumulant generating function  $K(\theta_w, \theta_z; t) = \log(\mathcal{G}(e^{\theta_w}, e^{\theta_z}; t))$  (cf. page 271 in [CM77]):

$$\frac{\partial K}{\partial t} = (\lambda e^{\theta_w + \theta_z} - (\lambda + \mu) + \mu e^{-\theta_z}) \frac{\partial K}{\partial \theta_z}. \quad (2.3)$$

Since  $\mathcal{G}$  is an analytic function,  $K$  is also analytic (as a composition of analytic functions). Thus for  $(\theta_w, \theta_z)$  from a ball around 0, we can write  $K$  in the form

$$K(\theta_w, \theta_z; t) = \sum_{i,j=0}^{\infty} \varkappa_{ij}(t) \frac{\theta_w^i \theta_z^j}{i!j!}, \quad (2.4)$$

where  $\varkappa_{ij}$  denote the corresponding cumulants. By inserting (2.4) into (2.3), we obtain

$$\sum_{i,j=0}^{\infty} \frac{d}{dt} \varkappa_{ij}(t) \frac{\theta_w^i \theta_z^j}{i!j!} = \left( \lambda \sum_{k=0}^{\infty} \frac{\theta_w^k}{k!} \sum_{l=0}^{\infty} \frac{\theta_z^l}{l!} - (\lambda + \mu) + \mu \sum_{k=0}^{\infty} \frac{(-\theta_z)^k}{k!} \right) \sum_{i,j=0}^{\infty} \varkappa_{i,j+1}(t) \frac{\theta_w^i \theta_z^j}{i!j!}.$$

By equating coefficients, this yields the following differential equations for the cumulants:

$$\frac{d}{dt} \varkappa_{01}(t) = (\lambda - \mu) \varkappa_{01}(t) \quad (2.5)$$

$$\frac{1}{2} \frac{d}{dt} \varkappa_{02}(t) = (\lambda - \mu) \varkappa_{02}(t) + \frac{\lambda + \mu}{2} \varkappa_{01}(t) \quad (2.6)$$

$$\frac{d}{dt} \varkappa_{10}(t) = \lambda \varkappa_{01}(t) \quad (2.7)$$

$$\frac{d}{dt} \varkappa_{11}(t) = (\lambda - \mu) \varkappa_{11}(t) + \lambda \varkappa_{02}(t) + \lambda \varkappa_{01}(t) \quad (2.8)$$

$$\frac{1}{2} \frac{d}{dt} \varkappa_{20}(t) = \lambda \varkappa_{11}(t) + \frac{1}{2} \lambda \varkappa_{01}(t). \quad (2.9)$$

Note that  $\varkappa_{10}(t) = \mathbb{E}(\mathfrak{B}_t)$ ,  $\varkappa_{01}(t) = \mathbb{E}(Y_t)$ ,  $\varkappa_{20}(t) = \text{Var}(\mathfrak{B}_t)$ ,  $\varkappa_{02}(t) = \text{Var}(Y_t)$  and  $\varkappa_{11}(t) = \text{Cov}(\mathfrak{B}_t, Y_t)$  (cf. e.g. Subsection 5.1.2 of [CJ10]). Thus from (2.5) and (2.6) follows Proposition 2.2.3. The differential equation (2.7) yields

$$\mathbb{E}(\mathfrak{B}_t) = \frac{\lambda}{\lambda - \mu} e^{(\lambda - \mu)t} - \frac{\mu}{\lambda - \mu}$$

if we assume  $\mathfrak{B}_0 = 1$ . The differential equation (2.8) implies

$$\text{Cov}(\mathfrak{B}_t, Y_t) = \frac{\lambda(\lambda + \mu)}{(\lambda - \mu)^2} e^{2(\lambda - \mu)t} - \frac{2\lambda\mu}{\lambda - \mu} t e^{(\lambda - \mu)t} - \frac{\lambda^2}{(\lambda - \mu)^2} e^{(\lambda - \mu)t}.$$

From (2.9) follows

$$\frac{d}{dt} \varkappa_{20}(t) = \frac{2\lambda^2(\lambda + \mu)}{(\lambda - \mu)^2} e^{2(\lambda - \mu)t} - \frac{4\lambda^2\mu}{\lambda - \mu} t e^{(\lambda - \mu)t} - \frac{2\lambda^3}{(\lambda - \mu)^2} e^{(\lambda - \mu)t} + \lambda e^{(\lambda - \mu)t},$$

which implies

$$\text{Var}(\mathfrak{B}_t) = \frac{\lambda^2(\lambda + \mu)}{(\lambda - \mu)^3} e^{2(\lambda - \mu)t} - \frac{4\lambda^2\mu}{(\lambda - \mu)^2} t e^{(\lambda - \mu)t} + \left( \frac{2\lambda^2\mu}{(\lambda - \mu)^3} - \frac{\lambda(\lambda + \mu)}{(\lambda - \mu)^2} \right) e^{(\lambda - \mu)t}.$$

We summarize these results in the following proposition.

### 2.3.1 Proposition

If we assume  $\mathfrak{B}_0 = 1$ , we have

$$\mathbb{E}(\mathfrak{B}_t) = \frac{\lambda}{\lambda - \mu} e^{(\lambda - \mu)t} - \frac{\mu}{\lambda - \mu}, \quad (2.10)$$

$$\text{Cov}(\mathfrak{B}_t, Y_t) = \frac{\lambda(\lambda + \mu)}{(\lambda - \mu)^2} e^{2(\lambda - \mu)t} - \frac{2\lambda\mu}{\lambda - \mu} t e^{(\lambda - \mu)t} - \frac{\lambda^2}{(\lambda - \mu)^2} e^{(\lambda - \mu)t},$$

$$\text{Var}(\mathfrak{B}_t) = \frac{\lambda^2(\lambda + \mu)}{(\lambda - \mu)^3} e^{2(\lambda - \mu)t} - \frac{4\lambda^2\mu}{(\lambda - \mu)^2} t e^{(\lambda - \mu)t} + \left( \frac{2\lambda^2\mu}{(\lambda - \mu)^3} - \frac{\lambda(\lambda + \mu)}{(\lambda - \mu)^2} \right) e^{(\lambda - \mu)t}.$$

### 2.3.2 Remark

Formula (2.10) can also be computed by using the differential equation

$$\frac{d}{dt} \mathbb{E}(\mathfrak{B}_t) = \lambda \mathbb{E}(Y_t),$$

which can be derived by conditioning on  $Y_t$  (see e.g. Section 2 of [CS12] or proof of Lemma 1 in Section 4 of [BL09]). Let  $\mathfrak{D}_t$  be the *number of deaths* up to time  $t$ . Then we have the analogue differential equation

$$\frac{d}{dt}\mathbb{E}(\mathfrak{D}_t) = \mu\mathbb{E}(Y_t),$$

which implies

$$\mathbb{E}(\mathfrak{D}_t) = \frac{\mu}{\lambda - \mu}e^{(\lambda - \mu)t} - \frac{\mu}{\lambda - \mu}$$

if we assume  $\mathfrak{D}_0 = 0$ .

### 2.3.3 Remark

Note that  $\mathbb{E}(Y_t) = \mathbb{E}(\mathfrak{B}_t) - \mathbb{E}(\mathfrak{D}_t)$  and  $\frac{\mathbb{E}(\mathfrak{B}_t)-1}{\mathbb{E}(\mathfrak{D}_t)} = \frac{\lambda}{\mu} = \frac{\lambda}{\lambda + \mu}(\frac{\mu}{\lambda + \mu})^{-1}$  is the ratio of the probabilities of a birth and a death at each event time. Furthermore,  $\mathbb{E}(\mathfrak{B}_t) + \mathbb{E}(\mathfrak{D}_t) = \frac{\lambda + \mu}{\lambda - \mu}e^{(\lambda - \mu)t} - \frac{2\mu}{\lambda - \mu}$  is the expected number of events up to time  $t$ .



# 3 Bounds on the total variation distance for Poisson and mixed Poisson approximation

## 3.1 Probability metrics and the total variation distance

Before we introduce the total variation distance in Subsection 3.1.2, which is crucial in this thesis, we briefly present the general concept of probability metrics.

### 3.1.1 Probability metrics

In order to be able to describe the speed of convergence of a sequence of distributions towards an asymptotic distribution, we need to quantify the distance between distributions. Therefore, we consider metrics on a subset  $\mathfrak{M}$  of the set  $\mathfrak{P}(\mathcal{X})$  of probability measures on a measurable space  $(\mathcal{X}, \mathcal{B})$ , i.e. a map  $d : \mathfrak{M} \times \mathfrak{M} \rightarrow [0, \infty)$  such that for all  $\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3 \in \mathfrak{M}$

$$(i) \quad d(\mathbb{P}_1, \mathbb{P}_2) = 0 \Leftrightarrow \mathbb{P}_1 = \mathbb{P}_2$$

$$(ii) \quad d(\mathbb{P}_1, \mathbb{P}_2) = d(\mathbb{P}_2, \mathbb{P}_1)$$

$$(iii) \quad d(\mathbb{P}_1, \mathbb{P}_2) \leq d(\mathbb{P}_1, \mathbb{P}_3) + d(\mathbb{P}_3, \mathbb{P}_2).$$

**3.1.1 Remark** (see e.g. the introduction of [Zol84] and the first two chapters of [RKSF13])

Note that the term probability metric is often used for a more general concept. Firstly, it is often allowed to be a semimetric on  $\mathfrak{M}$  only, i.e. a map  $d : \mathfrak{M} \times \mathfrak{M} \rightarrow [0, \infty)$  that fulfills the symmetry (ii), the triangle inequality (iii) and  $\mathbb{P}_1 = \mathbb{P}_2 \Rightarrow d(\mathbb{P}_1, \mathbb{P}_2) = 0$  instead of (i) for all  $\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3 \in \mathfrak{M}$ . Secondly, it is common to define it on the space of random variables instead of (a subset of)  $\mathfrak{P}(\mathcal{X})$  and allow so-called compound probability metrics, i.e. semimetrics that do not only depend on the distributions of the random variables. The probability metrics that are not compound probability metrics are called simple and are identified with the corresponding semimetrics on  $\mathfrak{M}$ . We only consider simple probability metrics here that fulfill (i), i.e. that are metrics on  $\mathfrak{M}$ .

Many well-known probability metrics  $d$  can be written in the form

$$d(\mathbb{P}_1, \mathbb{P}_2) = \sup_{f \in \mathfrak{F}} \left| \int_{\mathcal{X}} f d\mathbb{P}_1 - \int_{\mathcal{X}} f d\mathbb{P}_2 \right| \quad (3.1)$$

for  $\mathbb{P}_1, \mathbb{P}_2 \in \mathfrak{M}$  and a set  $\mathfrak{F}$  of measurable functions  $f : \mathcal{X} \rightarrow \mathbb{R}$  on a separable measurable space such that  $f$  is  $\tilde{\mathbb{P}}$ -integrable for any  $\tilde{\mathbb{P}} \in \mathfrak{M}$  (see e.g. Section 2 of [Zol84] and Section 4.1 of [RKSF13]). An example is the total variation distance, which is introduced in the next subsection. Here we give two other very well-known examples.

**3.1.2 Example** (see e.g. Section 2.2 of [RKSF13])

Let  $\mathcal{X} = \mathbb{R}$ ,  $\mathfrak{M} = \mathfrak{P}(\mathcal{X})$  and  $\mathfrak{F} = \{\mathbb{1}_{(-\infty, x]} : x \in \mathbb{R}\}$ . Then the probability metric  $d$  defined by (3.1) is the *Kolmogorov distance*, which we denote by  $d_K$ . We have

$$d_K(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{x \in \mathbb{R}} |F(x) - G(x)|$$

for random variables  $X$  and  $Y$  on  $(\mathcal{X}, \mathcal{B})$  having cumulative distribution functions  $F$  and  $G$ , respectively.

**3.1.3 Example** (see e.g. Appendix A1 in [BHJ92])

Consider a separable metric space  $(\mathcal{X}, \tilde{d})$  equipped with its Borel  $\sigma$ -algebra and the set  $\mathfrak{M}$  of probability measures that fulfill  $\int_{\mathcal{X}} \tilde{d}(x, x_0) d\tilde{\mathbb{P}}(x) < \infty$  for some, and therefore for every,  $x_0 \in \mathcal{X}$ . If  $\mathcal{X} = \mathbb{R}$  and the metric  $\tilde{d}$  is the usual one that is induced by the absolute value, the latter means that  $\int_{\mathcal{X}} x d\tilde{\mathbb{P}}(x) < \infty$ , i.e.  $\tilde{\mathbb{P}}$  has finite expectation. Let  $\mathfrak{F} = \{f : \mathcal{X} \rightarrow \mathbb{R} : |f(x) - f(y)| \leq \tilde{d}(x, y)\}$  be the set of Lipschitz continuous functions with Lipschitz constant 1. Then the probability metric defined by (3.1) is called *Wasserstein distance* (also known as Dudley, Fortet-Mourier and Kontorovich distance), which we denote by  $d_W$ . For  $\mathbb{P}_1, \mathbb{P}_2 \in \mathfrak{M}$ , it can be shown that

$$d_W(\mathbb{P}_1, \mathbb{P}_2) = \inf_{X \sim \mathbb{P}_1, Y \sim \mathbb{P}_2} \mathbb{E}(\tilde{d}(X, Y)).$$

Furthermore, for  $\mathcal{X} = \mathbb{R}$  and the metric  $\tilde{d}$  induced by the absolute value, it can be deduced that

$$d_W(\mathcal{L}(X), \mathcal{L}(Y)) = \int_{-\infty}^{\infty} |\mathbb{P}(X \leq x) - \mathbb{P}(Y \leq y)| dx$$

for real-valued random variables  $X$  and  $Y$ .

**3.1.2 The total variation distance****3.1.4 Definition and Remark** (see e.g. Appendix A1 in [BHJ92])

Let  $\mathfrak{F} = \{\mathbb{1}_B : B \in \mathcal{B}\}$  and  $\mathfrak{M} = \mathfrak{P}(\mathcal{X})$ . Then the probability metric defined by (3.1) is called *total variation distance* and is denoted by  $d_{TV}$ . Thus we have

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{B \in \mathcal{B}} |\mathbb{P}_1(B) - \mathbb{P}_2(B)|.$$

It can be shown by a simple approximation argument that we can also take the set of functions  $\mathfrak{F} = \{f : \mathcal{X} \rightarrow [0, 1] \text{ measurable}\}$  in (3.1) in order to obtain the total variation distance  $d_{TV}$  (see e.g. Appendix A1 in [BHJ92]). We now present further results that can be found in Appendix A1 in [BHJ92].

**3.1.5 Proposition**

Let  $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}(\mathcal{X})$ .

- (i) There exists an event  $B_+ \in \mathcal{B}$  such that  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{P}_1(B_+) - \mathbb{P}_2(B_+)$ .

(ii) Let  $\mathbb{P}_1$  and  $\mathbb{P}_2$  both be absolutely continuous with respect to some measure  $\zeta$  on  $(\mathcal{X}, \mathcal{B})$  and have densities  $f_1$  and  $f_2$  with respect to  $\zeta$ , respectively. Then we have

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \frac{1}{2} \int_{\mathcal{X}} |f_1 - f_2| d\zeta = 1 - \int_{\mathcal{X}} \min(f_1, f_2) d\zeta.$$

*Proof* [cf. proof of Lemma 3.3.1. in [Rei12], proof of Proposition 4.2 and page 51 in [LPW09]]:

Let  $\mathbb{P}_1$  and  $\mathbb{P}_2$  both be absolutely continuous with respect to some measure  $\zeta$  on  $(\mathcal{X}, \mathcal{B})$  and have densities  $f_1$  and  $f_2$  with respect to  $\zeta$ . Note that the choice  $\zeta = \mathbb{P}_1 + \mathbb{P}_2$  is always possible. Furthermore, let  $B_{>} := \{x : f_1(x) > f_2(x)\}$  and let  $B \in \mathcal{B}$ . Then

$$\begin{aligned} \mathbb{P}_1(B) - \mathbb{P}_2(B) &= \int_B (f_1(x) - f_2(x)) \zeta(dx) = \int_{B \cap B_{>}} (f_1(x) - f_2(x)) \zeta(dx) + \int_{B \cap B_{>}^c} (f_1(x) - f_2(x)) \zeta(dx) \\ &\leq \int_{B \cap B_{>}} (f_1(x) - f_2(x)) \zeta(dx) \leq \int_{B_{>}} (f_1(x) - f_2(x)) \zeta(dx) \\ &= \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}) \end{aligned} \tag{3.2}$$

since  $f_1(x) - f_2(x)$  is negative or zero for all  $x \in B_{>}^c$  and positive for all  $x \in B_{>}$ .

Analogously to (3.2), we obtain

$$\begin{aligned} \mathbb{P}_2(B) - \mathbb{P}_1(B) &= \int_B (f_2(x) - f_1(x)) \zeta(dx) = \int_{B \cap B_{>}} (f_2(x) - f_1(x)) \zeta(dx) + \int_{B \cap B_{>}^c} (f_2(x) - f_1(x)) \zeta(dx) \\ &\leq \int_{B \cap B_{>}^c} (f_2(x) - f_1(x)) \zeta(dx) \leq \int_{B_{>}^c} (f_2(x) - f_1(x)) \zeta(dx) \\ &= \mathbb{P}_2(B_{>}^c) - \mathbb{P}_1(B_{>}^c). \end{aligned} \tag{3.3}$$

Note that

$$\mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}) - (\mathbb{P}_2(B_{>}^c) - \mathbb{P}_1(B_{>}^c)) = \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}) - (1 - \mathbb{P}_2(B_{>}) - (1 - \mathbb{P}_1(B_{>}))) = 0.$$

Thus the right-hand sides of (3.2) and (3.3) are equal and we obtain

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}) = \mathbb{P}_2(B_{>}^c) - \mathbb{P}_1(B_{>}^c) \tag{3.4}$$

by the definition of the total variation distance since  $B \in \mathcal{B}$  is an arbitrary event. This yields  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{P}_1(B_+) - \mathbb{P}_2(B_+)$  with  $B_+ = B_{>} \in \mathcal{B}$ .

Furthermore, (3.4) implies

$$\begin{aligned} \int_{\mathcal{X}} |f_1 - f_2| d\zeta &= \int_{B_{>}} |f_1 - f_2| d\zeta + \int_{B_{>}^c} |f_1 - f_2| d\zeta \\ &= \int_{B_{>}} (f_1 - f_2) d\zeta + \int_{B_{>}^c} (f_2 - f_1) d\zeta \\ &= \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}) + \mathbb{P}_2(B_{>}^c) - \mathbb{P}_1(B_{>}^c) = 2d_{TV}(\mathbb{P}_1, \mathbb{P}_2). \end{aligned}$$

Dividing by 2 yields the first equality of (ii).

For the second equality of (ii), we note that

$$\begin{aligned}
 1 &= \int_{\mathcal{X}} f_1 d\zeta = \int_{B_{>}} f_2 + f_1 - f_2 d\zeta + \int_{B_{>}^c} f_1 d\zeta \\
 &= \int_{B_{>}} \min(f_1, f_2) + (f_1 - f_2) d\zeta + \int_{B_{>}^c} \min(f_1, f_2) d\zeta \\
 &= \int_{\mathcal{X}} \min(f_1, f_2) d\zeta + \int_{B_{>}} f_1 - f_2 d\zeta \\
 &= \int_{\mathcal{X}} \min(f_1, f_2) d\zeta + \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>}). \tag{3.5}
 \end{aligned}$$

Recall that  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{P}_1(B_{>}) - \mathbb{P}_2(B_{>})$  by (3.4). Thus (3.5) implies

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = 1 - \int_{\mathcal{X}} \min(f_1, f_2) d\zeta,$$

which completes the proof.  $\square$

### 3.1.6 Corollary

Let  $\mathbb{P}_1$  and  $\mathbb{P}_2$  be two probability measures on  $\mathcal{X} = \mathbb{N}_0$  with probability mass functions  $f_1$  and  $f_2$ , respectively. Then the corresponding total variation distance is given by

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \frac{1}{2} \sum_{k=0}^{\infty} |f_1(k) - f_2(k)|.$$

*Proof:* The statement follows immediately from Theorem 3.1.5 (ii).  $\square$

### 3.1.7 Remark (cf. e.g. Appendix A1 in [BHJ92])

It can be shown that  $d_{TV}$  is a complete metric on the space of all probability measures on  $(\mathcal{X}, \mathcal{B})$ . Moreover, it is easy to see that  $0 \leq d_{TV} \leq 1$  and that  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = 1$  if and only if  $\mathbb{P}_1$  and  $\mathbb{P}_2$  are mutually singular, i.e. there exists a set  $B \in \mathcal{B}$  such that  $\mathbb{P}_1(B) = 0$  and  $\mathbb{P}_2(B^c) = 0$ .

We now show that the total variation distance can be expressed as a minimal coupling.

### 3.1.8 Theorem

For all probability measures  $\mathbb{P}_1, \mathbb{P}_2$  on  $(\mathcal{X}, \mathcal{B})$ , we have

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \min \mathbb{P}(X_1 \neq X_2),$$

where we take the minimum over all random variables  $X_1$  and  $X_2$  that are defined on a common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  and satisfy  $X_1 \sim \mathbb{P}_1$  and  $X_2 \sim \mathbb{P}_2$ .

*Proof* [cf. e.g. page 254 in [BHJ92] and the proof of Proposition 4.7 in [LPW09]]:

Let  $X_1 \sim \mathbb{P}_1, X_2 \sim \mathbb{P}_2$  be random variables on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . Then we have for all probability measures  $\mathbb{P}_1, \mathbb{P}_2$  on  $(\mathcal{X}, \mathcal{B})$ , an arbitrary event  $B \in \mathcal{B}$  and  $k, l \in \{1, 2\}$  with  $k \neq l$

$$\mathbb{P}_k(B) - \mathbb{P}_l(B) = \mathbb{P}(X_k \in B) - \mathbb{P}(X_l \in B) \leq \mathbb{P}(X_k \in B, X_l \notin B) \leq \mathbb{P}(X_1 \neq X_2).$$

Thus  $|\mathbb{P}_1(B) - \mathbb{P}_2(B)| \leq \mathbb{P}(X_1 \neq X_2)$  for all  $B \in \mathcal{B}$  and Definition and Remark 3.1.4 implies

$$d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \sup_{B \in \mathcal{B}} |\mathbb{P}_1(B) - \mathbb{P}_2(B)| \leq \mathbb{P}(X_1 \neq X_2).$$

On the other hand, we can construct  $X_1 \sim \mathbb{P}_1, X_2 \sim \mathbb{P}_2$  with  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) \geq \mathbb{P}(X_1 \neq X_2)$  as follows.

Let  $\delta := d_{TV}(\mathbb{P}_1, \mathbb{P}_2)$ .

For  $\delta = 0$ , we then have  $|\mathbb{P}_1(B) - \mathbb{P}_2(B)| = 0$  for all  $B \in \mathcal{B}$ , i.e.  $\mathbb{P}_1 = \mathbb{P}_2$ . Let  $X_1 \sim P_1$  and  $X_2 := X_1 \sim P_2$ . Then we obtain  $\mathbb{P}(X_1 \neq X_2) = 0 = d_{TV}(\mathbb{P}_1, \mathbb{P}_2)$ .

For  $\delta = 1$ , we obviously have  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = 1 \geq \min \mathbb{P}(X_1 \neq X_2)$  since  $\mathbb{P}$  is a probability measure.

Finally, we assume that  $\delta \in (0, 1)$  and give a construction for  $X_1 \sim \mathbb{P}_1, X_2 \sim \mathbb{P}_2$  such that  $d_{TV}(\mathbb{P}_1, \mathbb{P}_2) = \mathbb{P}(X_1 \neq X_2)$ . Let  $\zeta$  be a measure on  $(\mathcal{X}, \mathcal{B})$  such that  $\mathbb{P}_1$  and  $\mathbb{P}_2$  are absolutely continuous with respect to  $\zeta$ , and let  $\mathbb{P}_1$  and  $\mathbb{P}_2$  have densities  $f_1$  and  $f_2$  with respect to  $\zeta$ , respectively. From Proposition 3.1.5 and its proof, we can conclude that  $\frac{1}{1-\delta} \min(f_1, f_2)$ ,  $\frac{1}{\delta}(f_1 - f_2)^+$  and  $\frac{1}{\delta}(f_1 - f_2)^-$  are *probability* densities with respect to  $\zeta$ , where we use the common notation  $f^+ := \max(0, f)$  for the positive part and  $f^- := -\min(0, f)$  for the negative part of a real-valued function  $f$ . Let  $Z, \tilde{X}_1$  and  $\tilde{X}_2$  be  $\mathcal{X}$ -valued random variables on the probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  such that their distributions have densities  $\frac{1}{1-\delta} \min(f_1, f_2)$ ,  $\frac{1}{\delta}(f_1 - f_2)^+$  and  $\frac{1}{\delta}(f_1 - f_2)^-$  with respect to  $\zeta$ , respectively. Let  $I$  be a random variable on  $(\Omega, \mathcal{A}, \mathbb{P})$  that satisfies  $I \sim \text{Be}(1 - \delta)$  and is independent of  $(Z, \tilde{X}_1, \tilde{X}_2)$ . Moreover, let

$$(X_1, X_2) = \begin{cases} (Z, Z) & \text{if } I = 1, \\ (\tilde{X}_1, \tilde{X}_2) & \text{if } I = 0. \end{cases}$$

Then we obtain

$$\mathbb{P}(X_1 \neq X_2) = \mathbb{P}(I = 0) = \delta$$

because we either have  $(f_1 - f_2)^+(x) = 0$  or  $(f_1 - f_2)^-(x) = 0$  for every  $x \in \mathcal{X}$ .

Furthermore, we have  $X_1 \sim \mathbb{P}_1$  since

$$\begin{aligned} \mathbb{P}(X_1 \in B) &= \mathbb{P}(Z \in B)(1 - \delta) + \mathbb{P}(\tilde{X}_1 \in B)\delta \\ &= \int_B \min(f_1, f_2) d\zeta + \int_B (f_1 - f_2)^+ d\zeta \\ &= \int_{B \cap B_{>}} \min(f_1, f_2) d\zeta + \int_{B \cap B_{\leq}} \min(f_1, f_2) d\zeta + \int_{B \cap B_{>}} (f_1 - f_2)^+ d\zeta + \int_{B \cap B_{\leq}} (f_1 - f_2)^+ d\zeta \\ &= \int_{B \cap B_{>}} f_2 d\zeta + \int_{B \cap B_{\leq}} f_1 d\zeta + \int_{B \cap B_{>}} f_1 - f_2 d\zeta \\ &= \int_B f_1 d\zeta, \end{aligned}$$

where  $B_{>} := \{x | f_1(x) > f_2(x)\}$ . Analogously, we obtain  $X_2 \sim \mathbb{P}_2$  since

$$\begin{aligned}
 \mathbb{P}(X_2 \in B) &= \mathbb{P}(Z \in B)(1 - \delta) + \mathbb{P}(\tilde{X}_2 \in B)\delta \\
 &= \int_B \min(f_1, f_2) d\zeta + \int_B (f_1 - f_2)^- d\zeta \\
 &= \int_{B \cap B_{>}} \min(f_1, f_2) d\zeta + \int_{B \cap B_{>}^c} \min(f_1, f_2) d\zeta + \int_{B \cap B_{>}} (f_1 - f_2)^- d\zeta + \int_{B \cap B_{>}^c} (f_1 - f_2)^- d\zeta \\
 &= \int_{B \cap B_{>}} f_2 d\zeta + \int_{B \cap B_{>}^c} f_1 d\zeta - \int_{B \cap B_{>}^c} f_1 - f_2 d\zeta \\
 &= \int_B f_2 d\zeta.
 \end{aligned}$$

□

Finally, we treat the convergence with respect to the total variation distance.

### 3.1.9 Proposition

Let  $\mathbb{P}_1, \mathbb{P}_2, \dots \in \mathcal{P}(\mathcal{X})$  and  $\mathbb{P}_\infty \in \mathcal{P}(\mathcal{X})$ . Then we have

$$d_{TV}(\mathbb{P}_n, \mathbb{P}_\infty) \xrightarrow{n \rightarrow \infty} 0 \implies \mathbb{P}_n \xrightarrow{w} \mathbb{P}_\infty,$$

where  $\xrightarrow{w}$  denotes *weak convergence*.

*Proof:*  $d_{TV}(\mathbb{P}_n, \mathbb{P}_\infty) \xrightarrow{n \rightarrow \infty} 0$  implies  $\mathbb{P}_n(B) \xrightarrow{n \rightarrow \infty} \mathbb{P}_\infty(B)$  for all  $B \in \mathcal{B}$  by the characterisation of the total variation distance given in Definition and Remark 3.1.4. Thus the claim follows with the Portmanteau Theorem. □

### 3.1.10 Remark

If  $\mathcal{X}$  is countable, the converse of Proposition 3.1.9 holds also, i.e.  $\mathbb{P}_n \xrightarrow{w} \mathbb{P}_\infty \implies d_{TV}(\mathbb{P}_n, \mathbb{P}_\infty) \xrightarrow{n \rightarrow \infty} 0$ . However, this is not true in general.

## 3.2 Poisson approximation

In this section, we briefly introduce the concept of Poisson approximation and refer to [BHJ92] for a lot of further results. The first part of this section is based on the introduction in [BHJ92].

In the remaining of this chapter, we assume that all random variables are defined on an underlying common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

The Poisson limit theorem, also known as law of small number, is one of the famous limit theorems in probability theory. Its simplest version says that the  $\text{Bin}(n, p(n))$  distribution converges to the  $\text{Po}(\tilde{\lambda})$  distribution as  $n \rightarrow \infty$  if  $p(n) = \tilde{\lambda}/n$  and was proved by Poisson in 1837. Thus for small  $p$  and large  $n$ , it is reasonable to approximate the  $\text{Bin}(n, p)$  distribution by a Poisson distribution. In other words, the Poisson distribution can be a useful approximation if rare events are considered. One of

the first remarkable applications of this result was the analysis of the numbers of Prussian soldiers killed each year as a result of being kicked by horses by von Bortkewitsch in 1898.

However, the Poisson limit theorem does not provide any information about the quality of the approximation for particular values of  $n$  and  $p$ . That is where Poisson approximation comes into play. One of the first results states that the total variation distance  $d_{TV}(\text{Bin}(n, p(n)), \text{Po}(np(n)))$  is at most of order  $\max(np(n)^2, p(n))$ . If  $I_1^{(n)}, \dots, I_n^{(n)}$  are independent  $\text{Be}(p(n))$  distributed random variables, this is an approximation result for the distribution of the sum  $\sum_{k=1}^n I_k^{(n)} \sim \text{Bin}(n, p(n))$ . If the random variables  $I_1^{(n)}, \dots, I_n^{(n)}$  are still assumed to be independent and Bernoulli but not necessarily identically distributed, i.e.  $I_k^{(n)} \sim \text{Be}(p_k(n))$  for  $k \in \{1, \dots, n\}$ , an upper bound for  $d_{TV}(\mathcal{L}(\sum_{k=1}^n I_k^{(n)}), \text{Po}(\tilde{\lambda}))$  can be derived using Stein's method, where  $\tilde{\lambda} = \sum_{k=1}^n p_k(n)$  (see the introduction in [BHJ92]). Namely, it can be shown that

$$d_{TV}\left(\mathcal{L}\left(\sum_{k=1}^n I_k^{(n)}\right), \text{Po}(\tilde{\lambda})\right) \leq \min\left(1, \frac{1}{\tilde{\lambda}}\right) \sum_{k=1}^n p_k(n)^2. \quad (3.6)$$

Note that the right-hand side of (3.6) is smaller than or equal to  $\max_{k \in \{1, \dots, n\}} p_k(n)$ . Thus we obtain a good approximation as long as all  $p_k(n)$  are small, no matter how large  $n$  is. This justifies the expression law of small number.

There are also other types of distributions that can be approximated by Poisson distributions in a reasonable way. Of course, a Poisson distribution is a good approximation for another Poisson distribution if the parameters are close to each other. The following theorem makes this statement more precise.

### 3.2.1 Theorem (cf. page 3 in [Roo03])

Let  $\tilde{\lambda} > 0$  and  $\tilde{\mu} > 0$ . Then we have

$$d_{TV}(\text{Po}(\tilde{\lambda}), \text{Po}(\tilde{\mu})) \leq \min\left(\sqrt{\frac{2}{e}}|\sqrt{\tilde{\lambda}} - \sqrt{\tilde{\mu}}|, |\tilde{\lambda} - \tilde{\mu}|\right).$$

We only prove  $d_{TV}(\text{Po}(\tilde{\lambda}), \text{Po}(\tilde{\mu})) \leq |\tilde{\lambda} - \tilde{\mu}|$  here since this is the statement that is needed for Chapter 6 and Chapter 7. The inequality  $d_{TV}(\text{Po}(\tilde{\lambda}), \text{Po}(\tilde{\mu})) \leq \sqrt{2/e}|\sqrt{\tilde{\lambda}} - \sqrt{\tilde{\mu}}|$  follows from the more general Theorem 3.3.9, which is proved in [Roo03]. A direct prove of  $d_{TV}(\text{Po}(\tilde{\lambda}), \text{Po}(\tilde{\mu})) \leq |\sqrt{\tilde{\lambda}} - \sqrt{\tilde{\mu}}|$  is presented in [Yan91].

A proof that is in some sense similar to ours can be found on page 260 in [Fre74]. Let  $(\Xi_t)_{t \geq 0}$  be a Poisson process of rate 1 on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$  (cf. Definition 2.1.3), and let  $X_1 := \Xi_{\tilde{\lambda}}$  and  $X_2 := \Xi_{\tilde{\mu}}$ . Then we have  $X_1 \sim \text{Po}(\tilde{\lambda})$  and  $X_2 \sim \text{Po}(\tilde{\mu})$  by Remark 2.1.4. Thus we obtain by Theorem 3.1.8 that

$$d_{TV}(\text{Po}(\tilde{\lambda}), \text{Po}(\tilde{\mu})) \leq \mathbb{P}(X_1 \neq X_2). \quad (3.7)$$

By the definition of  $X_1$  and  $X_2$ , we have  $X_1(\omega) \neq X_2(\omega)$  if and only if  $(\Xi_t(\omega))_{t \geq 0}$  jumps in the interval  $(\min(\tilde{\lambda}, \tilde{\mu}), \max(\tilde{\lambda}, \tilde{\mu})]$ , i.e.  $\Xi_{\max(\tilde{\lambda}, \tilde{\mu})}(\omega) - \Xi_{\min(\tilde{\lambda}, \tilde{\mu})}(\omega) > 0$ . Thus the right-hand side of (3.7) is equal to  $\mathbb{P}(\Xi_{\max(\tilde{\lambda}, \tilde{\mu})} - \Xi_{\min(\tilde{\lambda}, \tilde{\mu})} > 0)$ , and since  $(\Xi_t)_{t \geq 0}$  is a Poisson process of rate 1, the random variable  $\Xi_{\max(\tilde{\lambda}, \tilde{\mu})} - \Xi_{\min(\tilde{\lambda}, \tilde{\mu})}$  is Poisson distributed with parameter  $\max(\tilde{\lambda}, \tilde{\mu}) - \min(\tilde{\lambda}, \tilde{\mu}) = |\tilde{\lambda} - \tilde{\mu}|$  by Remark 2.1.4. Thus the right-hand side of (3.7) is equal to  $1 - e^{-|\tilde{\lambda} - \tilde{\mu}|}$ , which is bounded from above by  $|\tilde{\lambda} - \tilde{\mu}|$ .  $\square$

### 3.3 Mixed Poisson distributions

The degree distributions of the network models in Chapter 6 and Chapter 7 are mixed Poisson distributions, which are introduced in this section and appear in many different contexts (see [KX05]). We briefly state some interesting properties of these distributions and refer to [KX05] for an extensive survey.

**3.3.1 Definition** (cf. Definition 6.7 in [vdH16] and Definition 5 in [KX05])

Let  $\Lambda$  be a non-negative random variable. A random variable  $X$  has a *mixed Poisson distribution* with *mixing distribution*  $\mathbb{P}^\Lambda$ , denoted by  $\text{MixPo}(\Lambda)$ , if and only if for all  $k \in \mathbb{N}_0$

$$\mathbb{P}(X = k) = \int_0^\infty \frac{\tilde{\lambda}^k}{k!} e^{-\tilde{\lambda}} \mathbb{P}^\Lambda(d\tilde{\lambda}),$$

where  $\mathbb{P}^\Lambda$  denotes the distribution of  $\Lambda$ .

Note that a mixed Poisson distribution can be interpreted as Poisson distribution with a random parameter  $\Lambda$ . Note further that Definition 3.3.1 defines a rich family of probability distributions on  $\mathbb{N}_0$  (see [KX05]). However, not every probability distribution on  $\mathbb{N}_0$  is a mixed Poisson distribution (cf. Exercise 6.9 in [vdH16]).

A very well-known example of a mixed Poisson distribution is the negative binomial distribution, which is obtained by using a gamma distribution as mixing distribution (see e.g. [KX05]). The Yule-Simon distribution defined in Definition 5.4.3 below is a mixed Poisson distribution shifted by one (cf. also Remark 6.7.2(i) below).

The following proposition states that the mixing distribution is *identifiable*. It was first stated in [Fel43] and is discussed in Subsection 3.4 of [KX05].

**3.3.2 Proposition** (see Section 2 of [Fel43])

Let  $M_1$  and  $M_2$  be two non-negative random variables. Then we have  $\text{MixPo}(M_1) = \text{MixPo}(M_2)$  only if  $\mathcal{L}(M_1) = \mathcal{L}(M_2)$ .

Now we consider the tail behaviour of mixed Poisson distributions. We begin with a rather simple result from [vdH16].

**3.3.3 Proposition** (cf. Exercise 6.12 in [vdH16])

Let  $\Lambda$  be a non-negative random variable, and let  $F$  be the cumulative distribution function of  $\mathbb{P}^\Lambda$ . Suppose that there exist real constants  $c_1 < c_2$  such that

$$c_1 x^{1-\theta} \leq 1 - F(x) \leq c_2 x^{1-\theta}$$

for all  $x \in \mathbb{R}$  and some  $\theta > 1$ , and let  $G$  be the cumulative distribution function of the  $\text{MixPo}(\Lambda)$  distribution. Then there exist real constants  $\tilde{c}_1 < \tilde{c}_2$  such that

$$\tilde{c}_1 m^{1-\theta} \leq 1 - G(m) \leq \tilde{c}_2 m^{1-\theta}$$

for all  $m \in \mathbb{N}_0$ .



Before we consider further results about the tails of mixed Poisson distributions, we state two definitions.

**3.3.4 Definition** (cf. Section 2 of [Per98])

For functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  and  $g : \mathbb{R} \rightarrow \mathbb{R}$ , we write  $f \sim g$  and say that  $f$  and  $g$  are *tail-equivalent* if and only if  $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 1$ . For functions  $f : \mathbb{N}_0 \rightarrow \mathbb{R}$  and  $g : \mathbb{N}_0 \rightarrow \mathbb{R}$ , the notation  $f \sim g$  is defined analogously and corresponds to the usual concept of asymptotic equivalence for sequences.

**3.3.5 Definition** (cf. Section 1.4 of [vdH16])

A function  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  is called *slowly varying* if and only if

$$\lim_{x \rightarrow \infty} \frac{\varphi(cx)}{\varphi(x)} = 1$$

for all  $c > 0$ . For functions  $\varphi : \mathbb{N}_0 \rightarrow \mathbb{R}$  slowly varying is defined analogously.

**3.3.6 Remark**

An example of a slowly varying function is  $x \mapsto \log(x)$  since

$$\lim_{x \rightarrow \infty} \frac{\log(cx)}{\log(x)} = \lim_{x \rightarrow \infty} \frac{\log(c) + \log(x)}{\log(x)} = 1.$$

On the other hand, for any  $\varepsilon > 0$ , the function  $x \mapsto x^\varepsilon$  is not slowly varying because

$$\lim_{x \rightarrow \infty} \frac{(cx)^\varepsilon}{x^\varepsilon} = c^\varepsilon$$

and  $c^\varepsilon \neq 1$  for  $c \neq 1$ .

**3.3.7 Theorem** (cf. Theorem 2 in [Per98])

Let  $\Lambda$  be a non-negative random variable and assume that  $\mathbb{P}^\Lambda$  has density  $f$  with respect to the Lebesgue measure. Furthermore, let  $F$  and  $G$  denote the cumulative distribution functions of  $\mathbb{P}^\Lambda$  and  $\text{MixPo}(\Lambda)$ , respectively. Assume that the so-called *first von Mises condition* is fulfilled, i.e.

$$\lim_{x \rightarrow \infty} \frac{xf(x)}{F(x)} = c$$

for some  $c > 0$ . Then we have  $F \sim G$ .

If the first von Mises condition from Theorem 3.3.7 cannot be shown, the following theorem can often still be applied in order to describe the tail behaviour of the mixed Poisson distribution.

**3.3.8 Theorem** (cf. Theorem 2.1 in [Wil90])

Let

$$g(x) = \varphi(x)x^{-\theta}e^{-bx},$$

where  $\theta > 1$ ,  $b \geq 0$  and the function  $x \mapsto \varphi(x)$  is slowly varying. Furthermore, let  $\Lambda$  be a non-negative random variable and assume that  $\mathbb{P}^\Lambda$  has density  $f$  with respect to the Lebesgue or counting measure such that  $f \sim g$ . Then for  $X \sim \text{MixPo}(\Lambda)$ , we have

$$\mathbb{P}(X = k) \sim \frac{\varphi(k)}{(1+b)^{1-\theta}} \left( \frac{1}{1+b} \right)^k k^{-\theta}.$$

Finally, we return to Poisson approximation. The following theorem states an upper bound for the total variation distance between a Poisson and a mixed Poisson distribution, which can be used in order to approximate the mixed Poisson distribution by a Poisson distribution in special cases.

**3.3.9 Theorem** (see Theorem 1 in [Roo03])

Let  $\Lambda$  be a non-negative random variable, and let  $\tilde{\lambda} > 0$ . Then we have

$$d_{TV}(\text{MixPo}(\Lambda), \text{Po}(\tilde{\lambda})) \leq \min\left(\sqrt{\frac{2}{e}}\mathbb{E}(|\sqrt{\Lambda} - \sqrt{\tilde{\lambda}}|), \mathbb{E}(|\Lambda - \tilde{\lambda}|)\right).$$

### 3.4 Upper bound for the total variation distance between two mixed Poisson distributions

Theorem 3.2.1 gives us an upper bound for the total variation distance between two Poisson distributions. By conditioning, we can generalize this result to two mixed Poisson distributions.

**3.4.1 Theorem**

Let  $\Lambda$  and  $M$  be non-negative random variables. Then we have for the total variation distance between the mixed Poisson distributions  $\text{MixPo}(\Lambda)$  and  $\text{MixPo}(M)$ :

$$d_{TV}(\text{MixPo}(\Lambda), \text{MixPo}(M)) \leq \mathbb{E}\left(\min\left(\sqrt{\frac{2}{e}}|\sqrt{\Lambda} - \sqrt{M}|, |\Lambda - M|\right)\right).$$

*Proof:* Let  $X_1$  and  $X_2$  be  $\text{MixPo}(\Lambda)$  and  $\text{MixPo}(M)$  distributed, respectively. Then it follows

$$\begin{aligned} d_{TV}(\text{MixPo}(\Lambda), \text{MixPo}(M)) &= \sup_{B \subset \mathbb{N}_0} |\mathbb{P}(X_1 \in B) - \mathbb{P}(X_2 \in B)| \\ &= \sup_{B \subset \mathbb{N}_0} |\mathbb{E}(\mathbb{P}(X_1 \in B|\Lambda)) - \mathbb{E}(\mathbb{P}(X_2 \in B|M))| \\ &\leq \sup_{B \subset \mathbb{N}_0} \mathbb{E}(|\mathbb{P}(X_1 \in B|\Lambda) - \mathbb{P}(X_2 \in B|M)|) \\ &\leq \mathbb{E}\left(\sup_{B \subset \mathbb{N}_0} |\mathbb{P}(X_1 \in B|\Lambda) - \mathbb{P}(X_2 \in B|M)|\right) \\ &= \mathbb{E}(d_{TV}(\mathcal{L}(X_1|\Lambda), \mathcal{L}(X_2|M))) \\ &\leq \mathbb{E}\left(\min\left(\sqrt{\frac{2}{e}}|\sqrt{\Lambda} - \sqrt{M}|, |\Lambda - M|\right)\right), \end{aligned}$$

where the last line follows from Theorem 3.2.1. □

**3.4.2 Remark**

Note that the mixed Poisson distribution depends on the parameter random variable only via its distribution; therefore Theorem 3.4.1 yields

$$d_{TV}(\text{MixPo}(\Lambda), \text{MixPo}(M)) \leq \inf_{\substack{\tilde{\Lambda}: \tilde{\Lambda} \stackrel{\mathcal{D}}{=} \Lambda \\ \tilde{M}: \tilde{M} \stackrel{\mathcal{D}}{=} M}} \mathbb{E}\left(\min\left(\sqrt{\frac{2}{e}}|\tilde{\Lambda}^{\frac{1}{2}} - \tilde{M}^{\frac{1}{2}}|, |\tilde{\Lambda} - \tilde{M}|\right)\right) \leq d_W(\mathcal{L}(\Lambda), \mathcal{L}(M)).$$

Note further that Theorem 3.4.1 immediately implies that

$$d_{TV}(\text{MixPo}(\Lambda), \text{Po}(\tilde{\lambda})) \leq \mathbb{E} \left( \min \left( \sqrt{\frac{2}{e}} |\sqrt{\Lambda} - \sqrt{\tilde{\lambda}}|, |\Lambda - \tilde{\lambda}| \right) \right)$$

for a non-negative random variable  $\Lambda$  and  $\tilde{\lambda} > 0$ , which gives an upper bound for the total variation distance  $d_{TV}(\text{MixPo}(\Lambda), \text{Po}(\tilde{\lambda}))$  that is smaller than the one from Theorem 3.3.9 in some cases.



# 4 The age of a randomly picked individual in a linear birth and death process

We are interested in the age of an individual picked uniformly at random at a fixed time  $T > 0$  in the linear birth and death process  $(Y_t)_{t \geq 0}$  (given  $Y_T > 0$ ). We briefly call the distribution of this age *the age distribution of  $(Y_t)_{t \geq 0}$  at time  $T$* . However, before we consider this distribution, we turn to related known results. Firstly, we present well-known results about another concept for more general branching processes often referred to as “age distribution” in the literature, which can for example be found in [Har02] and [Ken49]. Namely, we consider the number of individuals with age in a certain interval and the ratio between this number and the total population size. Secondly, we consider the times since the birth times. Note that the latter does not lead to the age distribution since the birth times depend on the survival of the corresponding individuals up to time  $T$ .

## 4.1 Related results

### 4.1.1 The number of individuals with age in a certain interval

We consider the individuals that are not older than a fixed non-negative real number  $a$ .

#### 4.1.1 Definition

Let  $Y_T^a$  be the number of individuals whose age is smaller than or equal to  $a$  at time  $T$ .

The following theorem states that the expected value of  $Y_T^a$  is equal to the product of the expected population size and the cumulative distribution function of the  $\text{Exp}(\lambda)$  distribution at  $a$  for  $a < T$ .

#### 4.1.2 Theorem (see e.g. page 259 in [Ken49])

For  $a < T$ , we have

$$\mathbb{E}(Y_T^a) = e^{(\lambda - \mu)T} (1 - e^{-\lambda a}) = \mathbb{E}(Y_T) (1 - e^{-\lambda a}).$$

Note that an expression for  $\text{Var}(Y_T^a)$  can also be found in [Ken49].

The following theorem reveals that, given ultimate survival of  $(Y_t)_{t \geq 0}$ , the percentage of individuals that are not older than a fixed number  $a$  converges almost surely to a deterministic number.

#### 4.1.3 Theorem (see e.g. the corollary after Theorem 25.1 in Chapter VI of [Har02] or Example (6.10.14) in [Jag75])

Given  $\lim_{t \rightarrow \infty} Y_t = \infty$ , we have

$$\frac{Y_T^a}{Y_T} \rightarrow 1 - e^{-\lambda a}$$

almost surely as  $T$  goes to infinity.

#### 4.1.4 Remark

By Theorem 4.1.3, we have for the age  $A_{J_T}$  of an individual picked uniformly at random among all living individuals at time  $T$  and  $a \geq 0$ ,

$$\mathbb{P}(A_{J_T} \leq a | Y_T > 0) = \mathbb{E}\left(\frac{Y_T^a}{Y_T} \mid Y_T > 0\right) \rightarrow 1 - e^{-\lambda a}$$

as  $T \rightarrow \infty$  since  $\mathbb{P}(Y_T \rightarrow \infty | Y_T > 0) \rightarrow 1$  as  $T \rightarrow \infty$  and  $Y_T^a/Y_T \leq 1$  (cf. e.g. Example (6.10.14) in [Jag75]). Thus, given  $Y_T > 0$ , the age distribution, as defined in the beginning of this chapter, converges weakly to the  $\text{Exp}(\lambda)$  distribution. Note that this is a well-known result. However, to the best of our knowledge, an explicit formula for the expectation  $\mathbb{E}(Y_T^a/Y_T | Y_T > 0)$  cannot be found in the literature.

Note further that asymptotic age distributions have a long history. They were already considered by Euler in the 18th century as pointed out for example in Section 6.2 of [HJV05].

#### 4.1.2 Times since the birth times

The following theorem leads to the distribution of the birth times of the individuals that have been born up to time  $T$ . It is a special case of Theorem 1 in [SKBD13].

#### 4.1.5 Theorem (cf. Theorem 1 in [SKBD13])

Let  $X_2^+, \dots, X_{\mathfrak{B}_T}^+$  and  $X_1^-, \dots, X_{\mathfrak{D}_T}^-$  be the unordered (i.e. randomly permuted) non-zero birth and death times of the linear birth and death process  $(Y_t)_{t \geq 0}$  up to time  $T$ , respectively. Then the joint distribution of  $\mathfrak{B}_T, \mathfrak{D}_T, X_2^+, \dots, X_{\mathfrak{B}_T}^+$  and  $X_1^-, \dots, X_{\mathfrak{D}_T}^-$  is given by the density

$$f(b_T, d_T, \tilde{x}_2^+, \dots, \tilde{x}_{b_T}^+, \tilde{x}_1^-, \dots, \tilde{x}_{d_T}^-) = e^{(\lambda+\mu)T} \prod_{i=2}^{b_T} \lambda e^{-(\lambda+\mu)(T-x_i^+)} \prod_{k=1}^{d_T} \frac{\mu}{e^{-(\lambda+\mu)(T-x_k^-)}}$$

for all admissible  $b_T, d_T, \tilde{x}_2^+, \dots, \tilde{x}_{b_T}^+$  and  $\tilde{x}_1^-, \dots, \tilde{x}_{d_T}^-$ .

The following corollary is an immediate consequence of Theorem 4.1.5.

#### 4.1.6 Corollary

The unordered times since the non-zero birth times are independent and identically distributed. Their distribution is the  $\text{Exp}(\lambda + \mu)$  distribution truncated at  $T$ .

#### 4.1.7 Remark

Consider the case where we pick a birth time uniformly at random among all birth times that are not larger than  $T$ . Then the distribution of the time since this birth time converges to the  $\text{Exp}(\lambda + \mu)$  distribution as  $T \rightarrow \infty$  by Corollary 4.1.6. Note that this distribution is stochastically dominated by the asymptotic age distribution, which is the  $\text{Exp}(\lambda)$  distribution (see Remark 4.1.4).

## 4.2 The age distribution in the pure birth case

Recall that we call the distribution of the age of an individual picked uniformly at random at a fixed time  $T > 0$  (given  $Y_T > 0$ ) the age distribution. In this section, we consider the pure birth case and

show that the age distribution can be easily derived from known results in this case. The following proposition can be easily deduced from Theorem 1 in [NR71].

#### 4.2.1 Proposition

Let  $\mu = 0$ . The ages of the individuals at time  $T$  that have been born later than the first individual are independent and identically distributed and their distribution is the  $\text{Exp}(\lambda)$  distribution truncated at  $T$ . Consequently, the distribution of the age  $\hat{A}$  of an individual picked uniformly at random from all individuals at time  $T$  that were not alive at time 0 is the  $\text{Exp}(\lambda)$  distribution truncated at  $T$ .

*Proof:* Let  $Y_T = n$  be given. From Theorem 1 in [NR71] follows that the  $n - 1$  positive birth times in the time interval  $(0, T]$  are distributed as the order statistics of independent and identically distributed random variables  $\hat{T}_2, \dots, \hat{T}_n$  that have the cumulative distribution function

$$\mathbb{P}(\hat{T}_j \leq u) = \frac{1 - e^{-\lambda u}}{1 - e^{-\lambda T}}.$$

This implies that the ages  $T - \hat{T}_2, \dots, T - \hat{T}_n$  are independent and identically distributed and have the cumulative distribution function

$$\mathbb{P}(T - \hat{T}_j \leq u) = 1 - \mathbb{P}(\hat{T}_j \leq T - u) = \frac{1 - e^{-\lambda T} - 1 + e^{-\lambda(T-u)}}{1 - e^{-\lambda T}} = \frac{1 - e^{-\lambda u}}{1 - e^{-\lambda T}},$$

which is the cumulative distribution function of the  $\text{Exp}(\lambda)$  distribution truncated at  $T$ . Since this distribution does not depend on  $n$ , the conditional distribution given  $Y_T = n$  is equal to the unconditional one.  $\square$

#### 4.2.2 Remark

Note that Proposition 4.2.1 can be obtained as a special case of Corollary 4.1.6. However, we use the result from [NR71] since this is a more natural way to prove Proposition 4.2.1.

Before we turn to the general case, we give the following corollary.

#### 4.2.3 Corollary

Let  $\mu = 0$ , and let  $\tilde{A}$  be the age of an (arbitrary) individual at time  $T$  and  $0 < b < T$ . Then the conditional distribution of  $\tilde{A} - b$  given  $\tilde{A} \geq b$  is the  $\text{Exp}(\lambda)$  distribution truncated at  $T - b$ .

*Proof:* Let  $Z$  be  $\text{Exp}(\lambda)$  distributed. Then for the age  $\tilde{A}$  and  $z + b \leq T$ , Proposition 4.2.1 implies

$$\begin{aligned} \mathbb{P}(\tilde{A} - b \leq z | \tilde{A} \geq b) &= \mathbb{P}(Z \leq z + b | b \leq Z \leq T) = \frac{\mathbb{P}(b \leq Z \leq z + b)}{\mathbb{P}(b \leq Z \leq T)} = \frac{e^{-\lambda b} - e^{-\lambda(z+b)}}{e^{-\lambda b} - e^{-\lambda T}} \\ &= \frac{1 - e^{-\lambda z}}{1 - e^{-\lambda(T-b)}} = \mathbb{P}(Z \leq z | Z \leq T - b), \end{aligned}$$

and  $z \mapsto \mathbb{P}(Z \leq z | Z \leq T - b)$  is the cumulative distribution function of the  $\text{Exp}(\lambda)$  distribution truncated at  $T - b$ .  $\square$

#### 4.2.4 Remark

From the proof of Proposition 4.2.1, we know that we obtain the same distribution in Proposition 4.2.1

and Corollary 4.2.3 if we condition on the population size  $Y_T$ . Note that this implies that the age distribution of an individual picked uniformly at random conditioned on  $Y_T$  is a mixture of the  $\text{Exp}(\lambda)$  distribution truncated at  $T$  and the Dirac measure in  $T$ . The unconditional age distribution can be easily deduced from the conditional one by applying Proposition 2.2.4. The result is omitted here because the cumulative distribution function of the age distribution for arbitrary  $\mu \geq 0$  is stated in Corollary 4.3.2 below.

### 4.3 The age distribution in the general case

In the previous section, we only considered the case where individuals cannot die. In the general case with per-capita birth rate  $\lambda > 0$  and per-capita death rate  $\mu \geq 0$ , the situation is far more complicated. However, we finally obtain the age distribution for any  $\mu \geq 0$  in this section. For convenience, we assume  $\mu > 0$  here although all results also hold for  $\mu = 0$ , as can be verified from the previous subsection.

This section is organized as follows. We state our main results in Subsection 4.3.1. In Subsection 4.3.2 and Subsection 4.3.3, we introduce a bijection between trees and piecewise-linear functions and the contour process, respectively, which are used in the proof of the main result of this section in Subsection 4.3.4. The proofs of the additional results stated in Subsection 4.3.1 are given in Subsection 4.3.5, Subsection 4.3.6 and Subsection 4.3.7. Finally, we confirm the main results of this section using simulations in Subsection 4.3.8.

#### 4.3.1 Results

The following theorem gives us the age distribution at some time  $T$  conditioned on the population size.

##### 4.3.1 Theorem

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

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} \left( 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right) + \frac{1}{y_T} \left( \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0,T)}(t) + \mathbb{1}_{\{T\}}(t) \right)$$

for  $t \in [0, T]$  if  $\lambda \neq \mu$  and by

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} \left( 1 - \frac{e^{-\lambda t}(T - t)}{T} \right) + \frac{1}{y_T} \left( 1 - e^{-\lambda t}(1 + \lambda t) \right) \mathbb{1}_{[0,T)}(t) + \mathbb{1}_{\{T\}}(t)$$

for  $t \in [0, T]$  if  $\lambda = \mu > 0$ .

Before we prove this theorem, we state two corollaries. The first one reveals the unconditional age distribution of an individual picked uniformly at random.



### 4.3.2 Corollary

The cumulative distribution function  $F$  of the age distribution of  $(Y_t)_{t \geq 0}$  at time  $T$  is given by

$$F(t) = \left(1 - \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \lambda} \log\left(\frac{\lambda e^{(\lambda - \mu)T} - \mu}{\lambda - \mu}\right)\right) \left(1 - \frac{e^{-\lambda t} - e^{-(\lambda - \mu)T} e^{-\mu t}}{1 - e^{-(\lambda - \mu)T}}\right) \\ + \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \lambda} \log\left(\frac{\lambda e^{(\lambda - \mu)T} - \mu}{\lambda - \mu}\right) \left(\frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0, T)}(t) + \mathbb{1}_{\{T\}}(t)\right)$$

for  $t \in [0, T]$  if  $\lambda \neq \mu$  and by

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

for  $t \in [0, T]$  if  $\lambda = \mu > 0$ .

*Proof:* The expression for the cumulative distribution function follows directly from Theorem 4.3.1 and Proposition 2.2.4.  $\square$

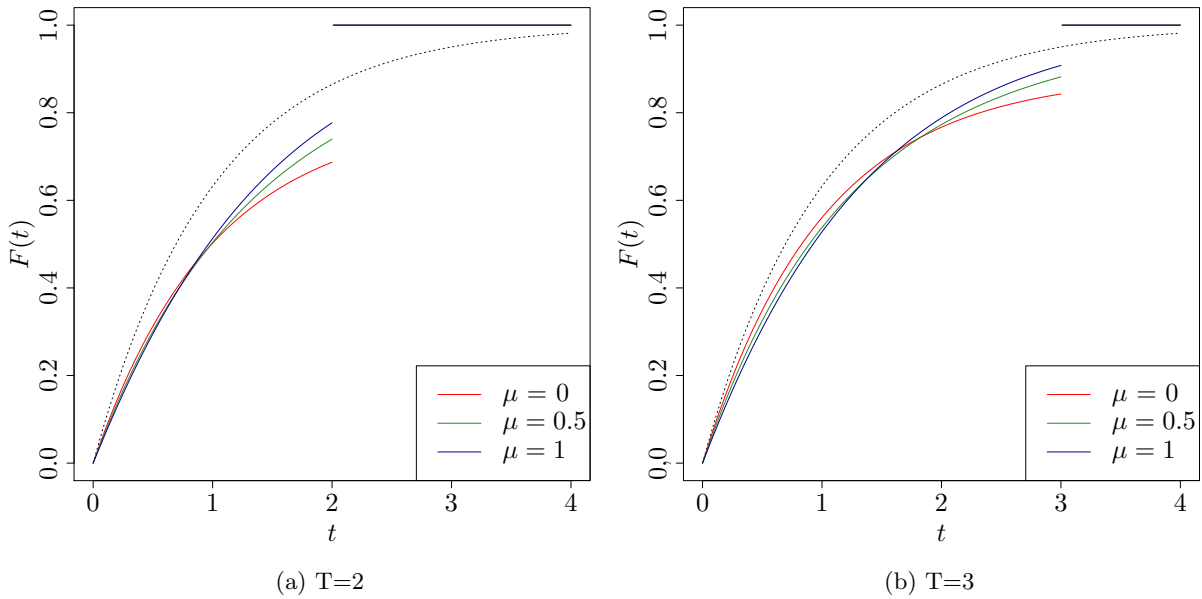


Figure 4.1: Cumulative distribution functions of the age distribution (solid lines) and the asymptotic age distribution (dotted line) for  $\lambda = 1$ .

The cumulative distribution function of the age distribution is illustrated in Figure 4.1. Note that Corollary 4.3.2 immediately implies that the age distribution converges weakly to the  $\text{Exp}(\lambda)$  distribution if  $\lambda \geq \mu$  and to a mixture of the  $\text{Exp}(\lambda)$  and the  $\text{Exp}(\mu)$  distribution if  $\lambda < \mu$ . The following corollary states that the age distribution converges exponentially fast to the  $\text{Exp}(\lambda)$  distribution in a certain sense for  $\lambda \geq \mu$ .

### 4.3.3 Corollary

Let  $c > 0$ , and let  $A_{J_T}$  denote the age of an individual picked uniformly at random at time  $T$  in the

linear birth and death process  $(Y_t)_{t \geq 0}$ . Then there exist random variables  $A_*$  and  $Z_*$  on  $(\Omega, \mathcal{A}, P)$  such that  $\mathcal{L}(A_* | Y_T > 0) = \mathcal{L}(A_{J_T} | Y_T > 0)$  and  $\mathcal{L}(Z_* | Y_T > 0) = \text{Exp}(\lambda)$  and

$$\begin{aligned} \mathbb{E}\left(\left|e^{-cA_*} - e^{-cZ_*}\right| \mid Y_T > 0\right) &\leq \frac{\lambda}{\lambda + c} \frac{1}{e^{(\lambda-\mu)T} - 1} + \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right) \\ &= O(Te^{-(\lambda-\mu)T}) \end{aligned}$$

if  $\lambda > \mu$  and

$$\mathbb{E}\left(\left|e^{-cA_*} - e^{-cZ_*}\right| \mid Y_T > 0\right) \leq \frac{c}{(\lambda + c)^2 T} + \frac{\log(1 + \lambda T)}{\lambda T} = O(\log(T)/T)$$

if  $\lambda = \mu > 0$ .

#### 4.3.4 Remark

Note that the upper bounds from Corollary 4.3.3 are obviously also upper bounds for the absolute difference  $|\mathbb{E}(e^{-cA_*} | Y_T > 0) - \mathbb{E}(e^{-cZ_*} | Y_T > 0)|$  of the corresponding Laplace transforms.

By adapting the proof of Corollary 4.3.3, we can obtain a similar bound for the Wasserstein distance from Example 3.1.3.

#### 4.3.5 Corollary

Let  $A_{J_T}$  be defined as in Corollary 4.3.3. Then we have, as  $T \rightarrow \infty$ ,

$$d_W(\mathcal{L}(A_{J_T} | Y_T > 0), \text{Exp}(\lambda)) = \mathbb{E}\left(\min_{\substack{\tilde{A}_* \sim \mathcal{L}(A_{J_T} | Y_T > 0) \\ \tilde{Z}_* \sim \text{Exp}(\lambda)}} |\tilde{A}_* - \tilde{Z}_*|\right) = \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu \\ O(\log(T)/T) & \text{if } \lambda = \mu > 0. \end{cases}$$

To round out the picture, we also provide a convergence rate in terms of the total variation metric for  $\lambda \geq \mu$ .

#### 4.3.6 Corollary

Let  $A_{J_T}$  be defined as in Corollary 4.3.3. Then we have, as  $T \rightarrow \infty$ ,

$$d_{TV}(\mathcal{L}(A_{J_T} | Y_T > 0), \text{Exp}(\lambda)) = \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu \\ O(\log(T)/T) & \text{if } \lambda = \mu > 0. \end{cases}$$

### 4.3.2 Bijection between piecewise-linear functions and trees

In order to prove Theorem 4.3.1 and Corollary 4.3.3, we use results from [BPS12] about contour processes (also known as exploration processes), which we introduce now.

#### 4.3.7 Definition

We consider *piecewise-linear functions*  $h : [0, b] \rightarrow [0, \infty)$  for  $0 < b < \infty$  with the following properties:

- The derivative is either 1 or -1 at all points where it exists.<sup>1</sup>

<sup>1</sup> Note that the formulation in [BPS12] is slightly more general. However, we do not need this more general setting for our purposes.

- $h(x) > 0$  if and only if  $x \in (0, b)$ .
- All non-zero local minima are at distinct heights.

We identify each such function  $h : [0, b] \rightarrow [0, \infty)$  with the function  $\tilde{h} : [0, \infty) \rightarrow [0, \infty)$  that fulfills

$$\tilde{h}(x) := \begin{cases} h(x) & \text{if } 0 \leq x \leq b, \\ 0 & \text{if } x > b. \end{cases}$$

Let  $\mathcal{H}$  denote the set of all such functions.

Now we introduce a *bijection*  $\phi$  between  $\mathcal{H}$  and the set  $\mathcal{T}$  of finite rooted binary *trees*. If not stated otherwise, we draw trees in a left-aligned way in this section, i.e. each inner node has a vertical line going in and out, and one horizontal line going to the right (see Figure 4.2). Horizontal line segments are for constructive purposes only, whereas the vertical line segments represent actual edges and the corresponding edge lengths in the tree. Further below we will identify such trees with realizations of birth and death processes that also include information about the lifetimes of the individuals represented by the lengths of the vertical lines.

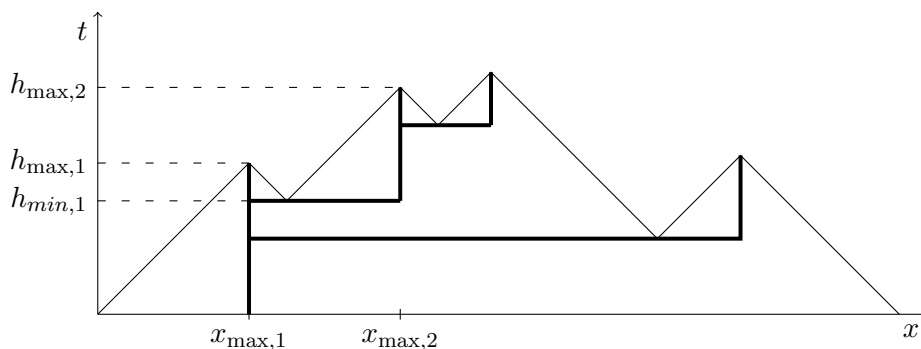


Figure 4.2: A piecewise linear function and the corresponding (left-aligned) tree

Starting from a piecewise linear function  $h \in \mathcal{H}$ , we define  $\phi(h)$  by drawing a tree under the graph of  $h$  as follows (see also Figure 4.2):

Let  $(x_{\max,1}, h_{\max,1})$  be the leftmost local maximum of  $h$ . Then we draw a (vertical) line from  $(x_{\max,1}, 0)$  to  $(x_{\max,1}, h_{\max,1})$ . Let  $(x_{\max,2}, h_{\max,2})$  be the second leftmost local maximum and  $(x_{\min,1}, h_{\min,1})$  be the leftmost non-zero local minimum if they exist. In this case, we add a (vertical) line from  $(x_{\max,2}, h_{\min,1})$  to  $(x_{\max,2}, h_{\max,2})$  and a (horizontal) line from  $(x_{\max,2}, h_{\min,1})$  connecting the vertical line horizontally to the rest of the tree. We continue this procedure until we have explored all non-zero local extrema (see Figure 4.2).

It is clear that  $\phi$  is a bijection since we obtain the inverse of  $\phi$  by assigning to any (left-aligned) tree, where the horizontal position of the root and the horizontal line lengths have been suitably adapted, the function that starts at  $(0, 0)$  and “turns” at every leaf and every time it hits a horizontal line until it hits the  $x$ -axis again.

### 4.3.3 Probability laws of the contour process and the corresponding random tree

Now we define probability measures on  $\mathcal{H}$  and  $\mathcal{T}$ . For the time being, let  $0 < \lambda \leq \mu$ . We define the *contour process* as a stochastic process  $(H_x)_{x \geq 0}$  whose trajectories belong to  $\mathcal{H}$ . Let  $(U_k)_{k \in \mathbb{N}}$  and  $(V_k)_{k \in \mathbb{N}}$  be two mutually independent sequences of independent, identically exponentially distributed random variables with parameters  $\mu$  and  $\lambda$ , respectively. Let  $Z_k = U_k - V_k$  for all  $k$ , and let  $P_{\lambda, \mu}$  be the law of the random element of  $\mathcal{H}$  with the following properties. Starting with a first local minimum at  $(0, 0)$ , the heights of the first local maximum and the second local minimum are  $U_1$  and  $(Z_1)^+$ , respectively. The process is stopped whenever it returns to zero. If  $(Z_1)^+ > 0$ , the heights of the second local maximum and the third local minimum are  $U_2 + Z_1$  and  $(Z_1 + Z_2)^+$ , respectively. If  $(Z_1 + Z_2)^+ > 0$ , we continue in the same way (see Figure 4.3).

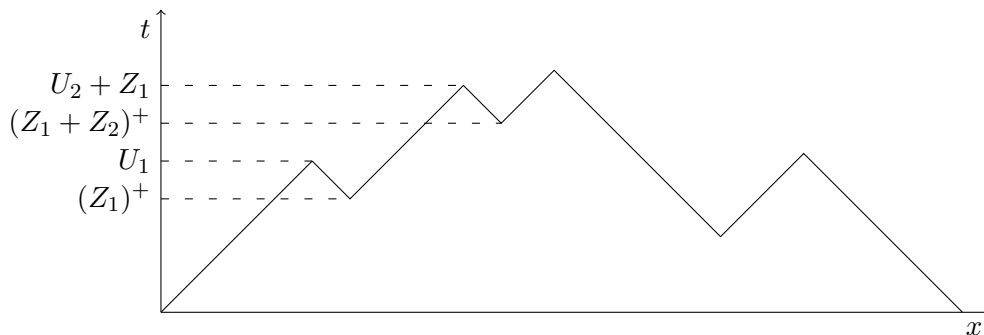


Figure 4.3: The contour process

It can be shown that this process returns to zero almost surely if  $0 < \lambda \leq \mu$  (see Section 2 of [BPS12]).

In Theorem 4.3.8 below, we relate  $P_{\lambda, \mu}$  to the law of a linear birth death process with the same parameters. Since we are also interested in the supercritical case, we also define the law of a random element of  $\mathcal{H}$  that can be related to the law of a supercritical linear birth and death process with birth rate  $\lambda$  and death rate  $\mu < \lambda$ . Since the contour process defined above does not return to zero almost surely if  $\lambda > \mu$ , we need to modify the construction. Therefore, we take the process from above and add a *deflection* at  $T$ : The  $k$ th time the process reaches  $T$ , it stops and goes down with slope  $-1$  for a time  $\min(T, \hat{U}_k)$ , where  $\hat{U}_k \sim \text{Exp}(\lambda)$  independent (also from all other random variables). It can be shown that this deflected process returns to zero almost surely (see Section 2 of [BPS12]). Let  $P_{\lambda, \mu, T}$  denote the *law of the deflected process*. Note that  $P_{\lambda, \mu, \infty} = P_{\lambda, \mu}$ .

Let  $Q_{\lambda, \mu, T}$  be the *law of the random tree corresponding to a linear birth and death process with parameters  $\lambda$  and  $\mu$  killed at time  $T$* , i.e. all individuals alive at time  $T-$  are killed at time  $T$ .<sup>2</sup> From a realization of such a birth and death process, including information about the lifetimes of the individuals, draw the corresponding random tree as follows. Draw the lifetime of the individual that is alive at time 0 as the leftmost vertical line. For each of its direct offspring attach a horizontal line, where the height of this line corresponds to the birth time of the offspring. Draw their lifetimes as further vertical lines. Starting with the latest offspring and keeping the lower horizontal lines long

<sup>2</sup> Here we say an individual is alive at time  $T-$  if and only if it is born but does not die before time  $T$ .

enough, will keep the tree planar. Continue in the same way for the direct offspring of each offspring and so on. Note that the lifetimes of the individuals correspond to the lengths of the vertical lines. Obviously, the ages of the individuals at time  $T$  in a linear birth and death process are just obtained by the total lifetimes of the individuals alive at time  $T-$  in the corresponding killed tree (see Figure 4.4).

We allow  $T = \infty$ , in which case the birth and death process is not killed.

The following theorem reveals the simple relationship between  $Q_{\lambda,\mu,T}$  and  $P_{\lambda,\mu,T}$ .

#### 4.3.8 Theorem

For any  $\lambda, \mu > 0$  and  $T \in (0, \infty)$ , and also for  $T = \infty$  if  $\lambda \leq \mu$ , we have

$$Q_{\lambda,\mu,T} = P_{\lambda,\mu,T} \phi^{-1}.$$

*Proof:* [BPS12], Theorem 3.1. See the proof of that theorem to verify the precise construction of the left-aligned tree based on the birth and death process.  $\square$

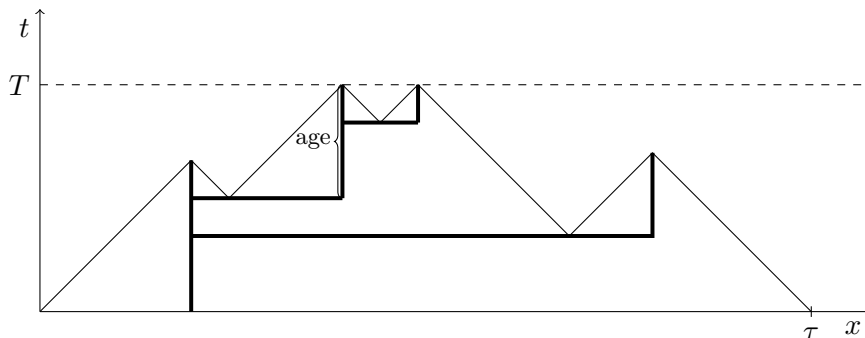


Figure 4.4: The deflected contour process and the corresponding tree representing a linear birth and death process up to time  $T$

#### 4.3.4 Proof of Theorem 4.3.1

We first prove that we obtain the same distribution if ages are read off from the tree under the vertically *mirrored* contour process  $(H_{\tau-t})_{0 \leq t \leq \tau}$ , where  $\tau = \inf\{x > 0 : H_x = 0\}$  (see Figure 4.5). To see this, denote by  $\Psi: \mathcal{T} \rightarrow \mathcal{T}$  the corresponding transformation on the set of trees. Let  $Z \sim Q_{\lambda,\mu,T}$  be the random tree obtained from the above birth and death process, and denote by  $\Theta(z)$  the finite set of trees that have the same vertex positions in the vertical direction as a tree  $z \in \mathcal{T}$ , but not necessarily the same phylogeny. Note that  $\Theta(\Psi(z)) = \Theta(z)$  for every  $z \in \mathcal{T}$ . Given  $\Theta(Z)$ , the random tree  $Z$  is uniformly distributed on  $\Theta(Z)$ , and so is  $\Psi(Z)$  because, restricted to the finite set  $\Theta(Z)$ , the map  $\Psi$  is just a permutation. Hence the distributions of  $\Psi(Z)$  and  $Z$  are equal by integration of the conditional distributions given  $\Theta(Z)$ .

Since it is more customary to draw stochastic processes from left to right, we mirror the situation in Figure 4.5, which gives us the original contour process  $(H_x)_{x \geq 0}$  again, enveloping a right-aligned

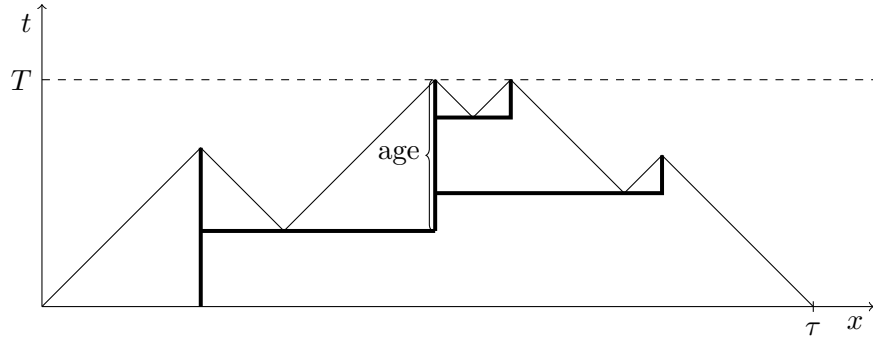


Figure 4.5: The mirrored deflected contour process (the realization corresponds to Figure 4.4) and its (left-aligned) tree

tree (constructed in the analogous way as the left-aligned tree, but from right to left; see Figure 4.6). Of course the ages still correspond to vertical line lengths.

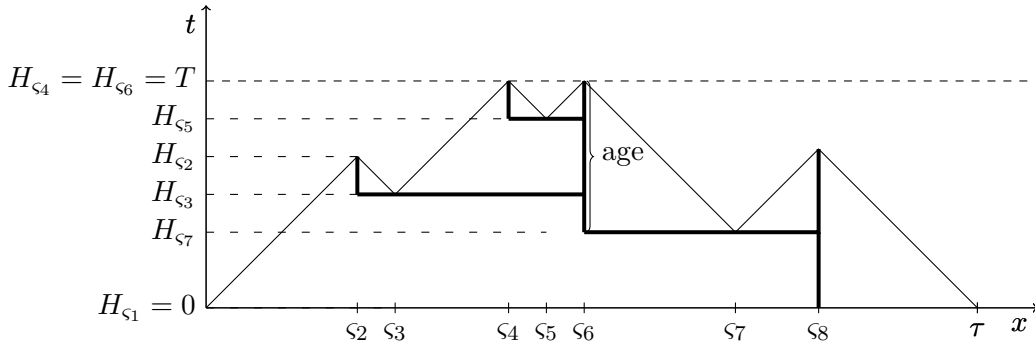


Figure 4.6: The deflected contour process and its right-aligned tree

Obviously, the local maxima of the contour process at height  $T$  correspond to the living individuals in the linear birth and death process at time  $T$ , which we assume to be numbered from left to right in the right-aligned tree here. In total, there are  $\mathfrak{B}_T$  local maxima (at any height) corresponding to the  $\mathfrak{B}_T$  individuals that have been born up to time  $T$ . Let these maxima be numbered from left to right, and let  $\mathfrak{J}_1 < \dots < \mathfrak{J}_{Y_T}$  be the indices of the maxima at height  $T$  and  $\mathfrak{J}_k = \infty$  for  $k > Y_T$ . Then  $V_{\mathfrak{J}_1}, \dots, V_{\mathfrak{J}_{Y_T-1}}, V_{\mathfrak{J}_{Y_T}} \wedge T$  are the ages we are interested in, where  $(V_k)_{k \in \mathbb{N}}$  is the sequence of random variables from the definition of the contour process, i.e.  $(V_k)_{k \in \mathbb{N}}$  are independent and identically  $\text{Exp}(\lambda)$  distributed, and  $\wedge$  denotes the minimum operator. Let  $0 = s_1 < \dots < s_{2\mathfrak{B}_T+1} = \tau$  be the positions of all local extrema of  $(H_x)_{0 \leq x \leq \tau}$  (cf. Figure 4.6). Note that  $s_{2\mathfrak{J}_1}, \dots, s_{2\mathfrak{J}_{Y_T}}$  are the positions of the maxima at height  $T$ . For convenience, we define  $s_k = s_{2\mathfrak{B}_T+1} = \tau$  for all  $k > 2\mathfrak{B}_T$  and  $s_\infty = \lim_{k \rightarrow \infty} s_k = \tau$ .

By the above construction of the contour process, the process  $(H_{s_k}, (-1)^k)_{k \in \mathbb{N}}$  is a Markov chain on

$[0, T] \times \{-1, 1\}$  with initial value  $(0, -1)$  at time 1 and transition probabilities

$$\begin{aligned} \mathbb{P}((H_{\varsigma_{k+1}}, (-1)^{k+1}) \in B_1 \times B_2 | (H_{\varsigma_k}, (-1)^k) = (z_1, z_2)) \\ = \begin{cases} \left( \delta_T(B_1) e^{-\mu(T-z_1)} + \mu \int_{B_1 \cap [z_1, T]} e^{-\mu(u-z_1)} du \right) \delta_1(B_2) & \text{if } z_2 = -1, \\ \left( \delta_0(B_1) e^{-\lambda z_1} + \lambda \int_{B_1 \cap (0, z_1]} e^{-\lambda(z_1-u)} du \right) \delta_{-1}(B_2) & \text{if } z_2 = 1, \end{cases} \end{aligned}$$

where  $B_1$  is a Borel subset of  $[0, T]$ ,  $B_2 \subset \{0, 1\}$  and  $(z_1, z_2) \in [0, T] \times \{-1, 1\}$ . Note that the second component of  $(H_{\varsigma_k}, (-1)^k)_{k \in \mathbb{N}}$  tells us if  $(H_x)_{0 \leq x \leq \tau}$  has a minimum (if it is  $-1$ ) or a maximum (if it is  $1$ ) at  $\varsigma_k$ . Note further that this Markov chain determines  $(H_x)_{x \geq 0}$  completely.

We define the sequence  $(\xi_k)_{k \in \mathbb{N}}$  of hitting times in  $(T, 1)$  recursively by

$$\xi_1 = \inf\{k \in \mathbb{N} : (H_{\varsigma_k}, (-1)^k) = (T, 1)\} \text{ and } \xi_{l+1} = \inf\{k > \xi_l : (H_{\varsigma_k}, (-1)^k) = (T, 1)\}$$

for all  $l \in \mathbb{N}$ , where  $\inf \emptyset = \infty$ . Note that each  $\xi_l$  is a stopping time and that  $\max\{l : \xi_l < \infty\} = Y_T$ .

Moreover, we define the ‘‘horizontally mirrored’’ excursions

$$E^{(l)} = (E_k^{(l)})_{0 \leq k \leq \xi_{l+1} - \xi_l} = (T - H_{\varsigma_k})_{\xi_l \leq k \leq \xi_{l+1}}$$

for  $l \in \mathbb{N}$ , where we set  $\infty - \infty = 0$ .

Since  $\mathbb{P}((H_{\varsigma_{\xi_l}}, (-1)^{\xi_l}) = (T, 1) \mid \xi_l < \infty) = 1$ , the strong Markov property and the fact that the second component of the process is deterministic imply that  $(H_{\varsigma_k})_{0 \leq k \leq \xi_l}$  and  $(H_{\varsigma_k})_{\xi_l \leq k}$  are independent given  $\xi_l < \infty$  and that  $(H_{\varsigma_k})_{\xi_l \leq k}$  has the same distribution given  $\xi_l < \infty$  for any  $l \in \mathbb{N}$ . As a consequence, we have for any  $l \in \mathbb{N}$  that  $(H_{\varsigma_k})_{0 \leq k \leq \xi_1}, E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $(E^{(l)}, (\xi_m)_{m \geq l+1})$  given  $\xi_l < \infty$  and hence that  $(H_{\varsigma_k})_{0 \leq k \leq \xi_1}, E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $E^{(l)}$  given  $\xi_l < \infty$  and given any sub- $\sigma$ -algebra of  $\sigma(\xi_m : m \geq l+1)$ . This implies that  $(H_{\varsigma_k})_{0 \leq k \leq \xi_1}, E^{(1)}, E^{(2)}, \dots, E^{(l-1)}, E^{(l)}$  are independent given  $\xi_l < \infty$  and given any sub- $\sigma$ -algebra of  $\sigma(\xi_m : m \geq l+1)$  for any  $l \in \mathbb{N}$ . Note that for any  $l \in \mathbb{N}$ , the first  $l-1$  ‘‘horizontally mirrored’’ excursions  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  all have the same distribution under this conditioning because  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  all have the same distribution given  $\xi_l < \infty$  and  $E^{(1)}, E^{(2)}, \dots, E^{(l-1)}$  are independent of  $(\xi_m)_{m \geq l+1}$  given  $\xi_l < \infty$ .

Since  $\{Y_T = y_T\} = \{\xi_{y_T} < \infty, \xi_{y_T+1} = \infty\}$ , we obtain that, given  $Y_T = y_T$ , the processes  $(H_{\varsigma_k})_{0 \leq k \leq \xi_1}, E^{(1)}, E^{(2)}, \dots, E^{(y_T-1)}, E^{(y_T)}$  are independent and  $E^{(1)}, E^{(2)}, \dots, E^{(y_T-1)}$  are identically distributed. By having a closer look at the strong Markov property used above, it is seen that neither the distribution of  $E^{(1)}$  given  $Y_T = y_T \geq 2$  nor the distribution of  $E^{(y_T)}$  given  $Y_T = y_T \geq 1$  depend on the concrete value  $y_T$ . Note that on  $\{Y_T = y_T\}$ , we have  $E_1^{(1)} = V_{\mathcal{J}_1}, \dots, E_1^{(y_T-1)} = V_{\mathcal{J}_{y_T-1}}$ , and  $E_1^{(y_T)} = V_{\mathcal{J}_{y_T}} \wedge T$ , so these are the ages we are interested in. The conditional distributions of  $E_1^{(1)}$  and  $E_1^{(y_T)}$  given  $Y_T = y_T$  can be easily rediscovered as conditional distributions in another contour process, which is what we look at next.

Let  $(\tilde{H}_x)_{x \geq 0} \sim P_{\mu, \lambda, T}$ . By Theorem 4.3.8,  $(\tilde{H}_x)_{x \geq 0}$  is the contour process corresponding to a linear birth and death process  $(Z_t)_{t \geq 0}$  with per-capita *birth rate*  $\mu$  and per-capita *death rate*  $\lambda$  that is killed at  $T$ . Let the positions  $(\tilde{\varsigma}_k)_{k \in \mathbb{N}}$  of the extrema of  $(\tilde{H}_x)_{x \geq 0}$  be defined analogously to  $(\varsigma_k)_{k \in \mathbb{N}}$ .

Both  $(\tilde{H}_{\tilde{\zeta}_{k+1}})_{k \in \mathbb{N}_0}$  and  $(E_k^{(l)})_{k \in \mathbb{N}_0}$  for arbitrary  $l \in \{1, \dots, Y_T\}$  start in 0 and alternate between independently adding  $\text{Exp}(\lambda)$  and subtracting  $\text{Exp}(\mu)$  random variables at least until 0 or  $T$  is crossed. Thus until this happens, they have the same distribution.

We have  $\{Y_T \geq 2\} = \{E_{\xi_2 - \xi_1}^{(1)} = 0\} = \{\max_{k \in \mathbb{N}} E_k^{(1)} < T\}$ , and therefore for any  $y_T \geq 2$  that, since the distribution of  $E_1^{(1)}$  does not depend on the concrete value of  $y_T$ ,

$$\begin{aligned} \mathcal{L}(E_1^{(1)} | Y_T = y_T) &= \mathcal{L}(E_1^{(1)} | E^{(1)} \text{ returns to 0 before reaching } T) \\ &= \mathcal{L}\left(\tilde{H}_{\tilde{\zeta}_2} \left| \max_{k \in \mathbb{N}} \tilde{H}_{\tilde{\zeta}_k} < T \right.\right) = \mathcal{L}\left(\tilde{H}_{\tilde{\zeta}_2} \left| \max_{x \geq 0} \tilde{H}_x < T \right.\right). \end{aligned} \quad (4.1)$$

On the other hand we have  $\{Y_T \geq 1\} = \{Y_T \geq 1, \max_{k \in \mathbb{N}} E_k^{(Y_T)} = T\}$  (in fact  $\xi_{Y_T+1} = \infty$  and  $E^{(Y_T)}$  is eventually absorbed in  $T$ ). Therefore, for any  $y_T \geq 1$ , since the distribution of  $E_1^{(y_T)}$  does not depend on the concrete value of  $y_T$ ,

$$\begin{aligned} \mathcal{L}(E_1^{(y_T)} | Y_T = y_T) &= \mathcal{L}(E_1^{(y_T)} | E^{(y_T)} \text{ reaches } T \text{ before returning to 0}) \\ &= \mathcal{L}\left(\tilde{H}_{\tilde{\zeta}_2} \left| \max_{k \in \mathbb{N}} \tilde{H}_{\tilde{\zeta}_k} = T \right.\right) = \mathcal{L}\left(\tilde{H}_{\tilde{\zeta}_2} \left| \max_{x \geq 0} \tilde{H}_x = T \right.\right). \end{aligned} \quad (4.2)$$

In other words, we have identified the desired age distributions at time  $T$  given  $Y_T = y_T$  as the distribution of the lifetime of the first individual in a linear birth and death process  $(Z_t)_{t \geq 0}$  with per-capita *birth rate*  $\mu$  and per-capita *death rate*  $\lambda$  conditioned on extinction of the process by time  $T$  (Equation (4.1), first  $y_T - 1$  individuals), i.e.  $Z_T = 0$ , or conditioned on survival of the process up to time  $T$  (Equation (4.2), last individual, lifetime measured up to time  $T$ ), i.e.  $Z_T > 0$ . Denote by  $L$  the lifetime of the starting individual, and let  $F_*$  and  $F^*$  be the cumulative distribution functions of  $L$  given  $Z_T = 0$  and of  $\min(L, T)$  given  $Z_T > 0$ , respectively. We then obtain from the above that the age distribution given  $Y_T = y_T > 0$  has cumulative distribution function

$$F_{y_T}(t) = \frac{y_T - 1}{y_T} F_*(t) + \frac{1}{y_T} F^*(t) \quad (4.3)$$

for all  $t \geq 0$ .

By Bayes' Theorem,  $F_*$  has density

$$f_*(t) \propto \lambda e^{-\lambda t} \mathbb{P}(Z_T = 0 | L = t), \quad (4.4)$$

for  $t \in [0, T)$ . Given  $L = t$ , the birth times of the offspring of the starting individual form a Poisson process of rate  $\mu$  (up to time  $t$ ). By conditioning on the number of offspring and their birth times and



plugging in their extinction probabilities from Section 2.2, we obtain

$$\begin{aligned}
\mathbb{P}(Z_T = 0 | L = t) &= \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} e^{-\mu t} \frac{1}{t^k} \int_0^t \cdots \int_0^t \frac{\lambda}{\mu} \frac{1 - e^{(\lambda-\mu)(T-t_1)}}{1 - \frac{\lambda}{\mu} e^{(\lambda-\mu)(T-t_1)}} \cdots \frac{\lambda}{\mu} \frac{1 - e^{(\lambda-\mu)(T-t_k)}}{1 - \frac{\lambda}{\mu} e^{(\lambda-\mu)(T-t_k)}} dt_1 \cdots dt_k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} \left( \int_0^t \lambda \frac{e^{-(\lambda-\mu)(T-u)} - 1}{e^{-(\lambda-\mu)(T-u)} - \frac{\lambda}{\mu}} du \right)^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\mu t} \left( \log(\lambda e^{(\lambda-\mu)(T-t)} - \mu) + \lambda t - \log(\lambda e^{(\lambda-\mu)T} - \mu) \right)^k \\
&= \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t}
\end{aligned} \tag{4.5}$$

for  $t < T$  if  $\lambda \neq \mu$ . For  $\lambda = \mu > 0$ , we see analogously that this probability is equal to

$$\begin{aligned}
\mathbb{P}(Z_T = 0 | L = t) &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t} \frac{1}{t^k} \int_0^t \cdots \int_0^t \frac{\lambda(T-t_1)}{1 + \lambda(T-t_1)} \cdots \frac{\lambda(T-t_k)}{1 + \lambda(T-t_k)} dt_1 \cdots dt_k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} \left( \int_0^t \lambda \frac{\lambda(T-u)}{1 + \lambda(T-u)} du \right)^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} e^{-\lambda t} (\log(1 + \lambda(T-t)) - \log(1 + \lambda T) + \lambda t)^k \\
&= \frac{1 + \lambda(T-t)}{1 + \lambda T}.
\end{aligned} \tag{4.6}$$

We may now compute the normalizing constant of the density  $f_*$  in (4.4). For  $\lambda \neq \mu$ , we obtain

$$\begin{aligned}
\int_0^T \lambda e^{-\lambda t} (\lambda e^{(\lambda-\mu)(T-t)} - \mu) e^{(\lambda-\mu)t} dt &= \int_0^T (\lambda e^{(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-\mu t}) dt \\
&= e^{(\lambda-\mu)T} (1 - e^{-\lambda T}) - (1 - e^{-\mu T}) \\
&= e^{(\lambda-\mu)T} - 1,
\end{aligned}$$

and for  $\lambda = \mu > 0$ , we have

$$\begin{aligned}
\int_0^T \lambda e^{-\lambda t} (1 + \lambda(T-t)) dt &= (1 - e^{-\lambda T})(1 + \lambda T) - \int_0^T \lambda^2 e^{-\lambda t} t dt \\
&= (1 - e^{-\lambda T})(1 + \lambda T) - (1 - e^{-\lambda T})(\lambda T + 1) \\
&= \lambda T.
\end{aligned}$$

Thus the density  $f_*$  is given by

$$f_*(t) = \frac{\lambda e^{(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-\mu t}}{e^{(\lambda-\mu)T} - 1} = \frac{\lambda e^{-\lambda t} - \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \tag{4.7}$$

for  $t \in [0, T]$  if  $\lambda \neq \mu$  and by

$$f_*(t) = \frac{\lambda e^{-\lambda t} (1 + \lambda(T-t))}{\lambda T} = \frac{e^{-\lambda t} (1 + \lambda(T-t))}{T} \tag{4.8}$$

for  $t \in [0, T]$  if  $\lambda = \mu > 0$ . By integration, we obtain that the cumulative distribution function  $F_*$  takes the form

$$F_*(t) = 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \quad (4.9)$$

and

$$F_*(t) = 1 - \frac{e^{-\lambda t}(T-t)}{T}$$

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

#### 4.3.9 Remark

In particular, we see that

$$1 - F_*(t) = \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \leq \frac{e^{-\lambda t}(1 - e^{-(\lambda-\mu)T})}{1 - e^{-(\lambda-\mu)T}} = e^{-\lambda t}$$

for  $t \in [0, T]$  if  $\lambda \neq \mu$  and

$$1 - F_*(t) = \frac{e^{-\lambda t}(T-t)}{T} \leq e^{-\lambda t}$$

for  $t \in [0, T]$  if  $\lambda = \mu$ . Thus, given  $Y_T = y_T$ , the age distribution of the first  $y_T - 1$  individuals is stochastically dominated by the  $\text{Exp}(\lambda)$  distribution.

In order to prove Theorem 4.3.1, it remains to derive  $F^*$ , which we do in a similar way. Recall that we have to compute  $\mathcal{L}(\min(L, T) | Z_T > 0)$ . A slight notational complication arises from the fact that  $F^*$  has a discontinuity at  $T$ . We note that  $\min(L, T)$  has a density  $\tilde{f}$  with respect to the measure  $\text{Leb}_{[0, T]} + \delta_T$  given by

$$\tilde{f}(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } 0 \leq t < T, \\ e^{-\lambda T} & \text{if } t = T. \end{cases}$$

Thus by Bayes' Theorem, the age distribution of the last individual given  $Y_T = y_T$  has a density  $f^*$  with respect to  $\text{Leb}_{[0, T]} + \delta_T$  satisfying

$$f^*(t) \propto \tilde{f}(t) \mathbb{P}(Z_T > 0 | \min(L, T) = t) = \begin{cases} \lambda e^{-\lambda t} \mathbb{P}(Z_T > 0 | L = t) & \text{if } 0 \leq t < T, \\ e^{-\lambda T} & \text{if } t = T. \end{cases}$$

Firstly, we consider the case where  $\lambda \neq \mu$ . By (4.5), we have that

$$\mathbb{P}(Z_T > 0 | L = t) = 1 - \mathbb{P}(Z_T = 0 | L = t) = 1 - \frac{\lambda e^{(\lambda-\mu)(T-t)} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} e^{(\lambda-\mu)t}$$

for  $t \in [0, T)$ , which implies

$$f^*(t) \propto \begin{cases} \lambda e^{-\lambda t} \frac{\mu e^{(\lambda-\mu)t} - \mu}{\lambda e^{(\lambda-\mu)T} - \mu} & \text{if } 0 \leq t < T, \\ e^{-\lambda T} & \text{if } t = T. \end{cases}$$

We can compute the normalizing constant of  $f^*$  as

$$\begin{aligned} e^{-\lambda T} + \frac{\lambda\mu}{\lambda e^{(\lambda-\mu)T} - \mu} \int_0^T e^{-\lambda t} (e^{(\lambda-\mu)t} - 1) dt &= e^{-\lambda T} + \frac{\lambda\mu}{\lambda e^{(\lambda-\mu)T} - \mu} \int_0^T (e^{-\mu t} - e^{-\lambda t}) dt \\ &= e^{-\lambda T} + \frac{\lambda(1 - e^{-\mu T}) - \mu(1 - e^{-\lambda T})}{\lambda e^{(\lambda-\mu)T} - \mu} \\ &= \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \mu}. \end{aligned} \quad (4.10)$$

#### 4.3.10 Remark

By Bayes' Theorem, the normalizing constant of the density  $f^*$  is just the probability that a linear birth and death process with birth rate  $\mu$  and death rate  $\lambda$  survives up to time  $T$ . Thus we could also have used the extinction probability from Section 2.2 in order to obtain the right-hand side of (4.10).

We may conclude that  $f^*$  is given by

$$f^*(t) = \begin{cases} \lambda\mu \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu} & \text{if } 0 \leq t < T, \\ \frac{\lambda e^{-\mu T} - \mu e^{-\lambda T}}{\lambda - \mu} & \text{if } t = T. \end{cases} \quad (4.11)$$

By integration, we obtain that the cumulative distribution function  $F^*$  for the age of the last individual given  $Y_T = y_T$  takes the form

$$F^*(t) = \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0, T)}(t) + \mathbb{1}_{\{T\}}(t) \quad (4.12)$$

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

For  $\lambda = \mu > 0$ , we proceed analogously: By (4.6), we have that

$$\mathbb{P}(Z_T > 0 | L = t) = 1 - \mathbb{P}(Z_T = 0 | L = t) = 1 - \frac{1 + \lambda(T - t)}{1 + \lambda T} = \frac{\lambda t}{1 + \lambda T}$$

for  $t \in [0, T)$ , which leads to

$$f^*(t) \propto \begin{cases} \lambda e^{-\lambda t} \frac{\lambda t}{1 + \lambda T} & \text{if } 0 \leq t < T, \\ e^{-\lambda T} & \text{if } t = T. \end{cases}$$

We compute the normalizing constant of  $f^*$  as

$$e^{-\lambda T} + \int_0^T e^{-\lambda t} \frac{\lambda t}{1 + \lambda T} dt = e^{-\lambda T} + \frac{1}{1 + \lambda T} - e^{-\lambda T} = \frac{1}{1 + \lambda T}.$$

This yields that  $f^*$  is given by

$$f^*(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & \text{if } 0 \leq t < T, \\ (1 + \lambda T) e^{-\lambda T} & \text{if } t = T. \end{cases} \quad (4.13)$$

By integration, we obtain that the cumulative distribution function  $F^*$  takes the form

$$F^*(t) = (1 - e^{-\lambda t} (1 + \lambda t)) \mathbb{1}_{[0, T)}(t) + \mathbb{1}_{\{T\}}(t)$$

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

Plugging the above expressions for  $F_*$  and  $F^*$  into Equation (4.3) yields the statement of Theorem 4.3.1.  $\square$

### 4.3.5 Proof of Corollary 4.3.3

It is enough to construct  $A_*$  and  $Z_*$  on  $\{Y_T > 0\}$  that have the right distributions given  $Y_T > 0$  and are such that  $\mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T > 0)$  satisfies the required bound. We construct them explicitly as a quantile coupling using notation from the proof of Theorem 4.3.1. Let  $J_T$  be a random variable that is uniformly distributed on  $\{1, \dots, y_T\}$  given  $Y_T = y_T > 0$ , independent from everything else, and let  $U$  be uniformly distributed on  $[0, 1]$  and also independent from everything else. Defining the generalized inverse of a cumulative distribution function  $F$  by  $F^{-1}(u) = \inf\{t \in \mathbb{R} : F(t) \geq u\}$ , set

$$\begin{aligned} A_* &:= \mathbb{1}_{\{J_T < Y_T\}} F_*^{-1}(U) + \mathbb{1}_{\{J_T = Y_T\}} (F^*)^{-1}(U), \\ Z_* &:= F_\infty^{-1}(U), \end{aligned} \tag{4.14}$$

where  $F_\infty$  denotes the cumulative distribution function of  $\text{Exp}(\lambda)$ . By Equation (4.3) and the independence of  $U$  from  $(Y_T, J_T)$ , we obtain  $\mathcal{L}(A_* \mid Y_T > 0) = \mathcal{L}(A_{J_T} \mid Y_T > 0)$  and  $\mathcal{L}(Z_* \mid Y_T > 0) = \text{Exp}(\lambda)$  as required.

By Remark 4.3.9, we have

$$F_*(t) \geq 1 - e^{-\lambda t} = F_\infty(t) \quad \text{for all } t > 0.$$

Since  $F_\infty$  is continuous and strictly increasing on  $[0, \infty)$ , this implies

$$F_*(F_\infty^{-1}(U)) \geq F_\infty(F_\infty^{-1}(U)) = U \tag{4.15}$$

almost surely. By well-known properties of the generalized inverse, we have that  $F_*^{-1}$  is non-decreasing and  $F_*^{-1}(F_*(t)) \leq t$  for all  $t > 0$ . Thus (4.15) implies

$$F_\infty^{-1}(U) \geq F_*^{-1}(F_*(F_\infty^{-1}(U))) \geq F_*^{-1}(U) \tag{4.16}$$

almost surely.

We may compute

$$\begin{aligned} &\mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T > 0) \\ &= \mathbb{E}\left(\frac{Y_T - 1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T < Y_T) + \frac{1}{Y_T} \mathbb{E}(|e^{-cA_*} - e^{-cZ_*}| \mid Y_T, J_T = Y_T) \mid Y_T > 0\right) \\ &\leq \mathbb{E}(e^{-cF_*^{-1}(U)} - e^{-cF_\infty^{-1}(U)}) + \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right), \end{aligned} \tag{4.17}$$

where we use (4.16) in the last line.

#### 4.3.11 Remark

Note that we could employ a more sophisticated argument, taking care of the sign of the expression  $e^{-c(F^*)^{-1}(U)} - e^{-cF_\infty^{-1}(U)}$  for the second inner conditional expectation in the second line of (4.17). But since  $F^*(t) \not\rightarrow F_\infty(t)$  as  $T \rightarrow \infty$  for any  $t > 0$ , we would not gain anything in terms of the convergence rates; in particular, the factors  $T$  and  $\log(T)$  in the orders of the upper bound if  $\mu < \lambda$  and  $\mu = \lambda$ , respectively, cannot be removed.

It is now a matter of computing Laplace transforms. Since  $F_\infty^{-1}(U) \sim \text{Exp}(\lambda)$ , we have

$$\mathbb{E}(e^{-cF_\infty^{-1}(U)}) = \frac{\lambda}{\lambda + c}. \quad (4.18)$$

For  $F_*^{-1}(U)$  we use the density from (4.11) and (4.13), respectively, and obtain for  $\lambda > \mu$

$$\begin{aligned} \mathbb{E}(e^{-cF_*^{-1}(U)}) &= \frac{1}{1 - e^{-(\lambda-\mu)T}} \int_0^T (\lambda e^{-(\lambda+c)t} - \mu e^{-(\lambda-\mu)T} e^{-(\mu+c)t}) dt \\ &= \frac{1}{1 - e^{-(\lambda-\mu)T}} \left( \frac{\lambda}{\lambda + c} (1 - e^{-(\lambda+c)T}) - \frac{\mu}{\mu + c} e^{-(\lambda-\mu)T} (1 - e^{-(\mu+c)T}) \right) \\ &\leq \frac{\lambda}{\lambda + c} \frac{e^{(\lambda-\mu)T}}{e^{(\lambda-\mu)T} - 1}, \end{aligned} \quad (4.19)$$

and for  $\lambda = \mu > 0$

$$\mathbb{E}(e^{-cF_*^{-1}(U)}) = \frac{1}{T} \int_0^T e^{-(\lambda+c)t} (1 + \lambda(T-t)) dt = \frac{\lambda}{\lambda + c} + \frac{c(1 - e^{-(\lambda+c)T})}{(\lambda + c)^2 T} \leq \frac{\lambda}{\lambda + c} + \frac{c}{(\lambda + c)^2 T}. \quad (4.20)$$

Combining (4.18)–(4.20) to bound the first summand on the right hand side of (4.17) and employing Lemma 2.2.4 for the second summand, we obtain the required bounds.  $\square$

### 4.3.6 Proof of Corollary 4.3.5

Let  $\tilde{A}_*$  and  $\tilde{Z}_*$  be random variables on a common probability space such that

$$\mathcal{L}((\tilde{A}_*, \tilde{Z}_*)) = \mathcal{L}((A_*, Z_*) | Y_T > 0),$$

where the random variables  $A_*$  and  $Z_*$  are defined by (4.14). Then we know from the previous subsection that  $\mathcal{L}(\tilde{A}_*) = \mathcal{L}(A_* | Y_T > 0) = \mathcal{L}(A_{J_T} | Y_T > 0)$  and  $\mathcal{L}(\tilde{Z}_*) = \mathcal{L}(Z_* | Y_T > 0) = \text{Exp}(\lambda)$  as desired. Thus we have

$$d_W(\mathcal{L}(A_{J_T} | Y_T > 0), \text{Exp}(\lambda)) \leq \mathbb{E}(|\tilde{A}_* - \tilde{Z}_*|) = \mathbb{E}(|A_* - Z_*| | Y_T > 0).$$

Analogously to the procedure in (4.17), we compute

$$\begin{aligned} \mathbb{E}(|A_* - Z_*| | Y_T > 0) &= \mathbb{E}\left(\frac{Y_T - 1}{Y_T} \mathbb{E}(|A_* - Z_*| | Y_T, J_T < Y_T) + \frac{1}{Y_T} \mathbb{E}(|A_* - Z_*| | Y_T, J_T = Y_T) \mid Y_T > 0\right) \\ &\leq \mathbb{E}(F_\infty^{-1}(U) - F_*^{-1}(U)) + \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) (\mathbb{E}(F_\infty^{-1}(U)) + \mathbb{E}((F^*)^{-1}(U))), \end{aligned} \quad (4.21)$$

where  $U$  is uniformly distributed on  $[0, 1]$ .

For  $\lambda > \mu$ , we have

$$\begin{aligned} \mathbb{E}(F_*^{-1}(U)) &= \int_0^T t f_*(t) dt = \int_0^T \frac{t \lambda e^{-\lambda t}}{1 - e^{-(\lambda-\mu)T}} dt - \int_0^T \frac{t \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} dt \\ &= \frac{\frac{1}{\lambda} - e^{-\lambda T} T - \frac{e^{-\lambda T}}{\lambda}}{1 - e^{-(\lambda-\mu)T}} - \frac{\frac{1}{\mu} - e^{-\mu T} - \frac{e^{-\mu T}}{\mu}}{1 - e^{-(\lambda-\mu)T}} e^{-(\lambda-\mu)T} \\ &= \frac{1}{\lambda} + O(T e^{-(\lambda-\mu)T}). \end{aligned}$$

Thus

$$\mathbb{E}(F_\infty^{-1}(U) - F_*^{-1}(U)) = \mathbb{E}(F_\infty^{-1}(U)) - \mathbb{E}(F_*^{-1}(U)) = O(Te^{-(\lambda-\mu)T}).$$

For  $\lambda = \mu$ , we obtain

$$\begin{aligned} \mathbb{E}(F_*^{-1}(U)) &= \int_0^T t f_*(t) dt = \int_0^T \frac{te^{-\lambda t}(1 + \lambda(T-t))}{T} dt \\ &= \frac{(\lambda T + 1) - e^{-\lambda T}(\lambda T + 1)^2}{\lambda^2 T} - \frac{e^{-\lambda T}(-\lambda T(\lambda T + 2) - 2) + 2}{\lambda^2 T} \\ &= \frac{1}{\lambda} + O\left(\frac{1}{T}\right). \end{aligned}$$

Thus

$$\mathbb{E}(F_\infty^{-1}(U) - F_*^{-1}(U)) = \mathbb{E}(F_\infty^{-1}(U)) - \mathbb{E}(F_*^{-1}(U)) = O\left(\frac{1}{T}\right).$$

Furthermore, for  $\lambda > \mu$ , we have

$$\begin{aligned} \mathbb{E}((F^*)^{-1}(U)) &= \int_0^T t f^*(t) dt + T f^*(T) = \lambda \int_0^T t^2 \lambda e^{-\lambda t} dt + (1 + \lambda T) e^{-\lambda T} \\ &\leq \lambda \int_0^\infty t^2 \lambda e^{-\lambda t} dt + O(1) = \frac{2}{\lambda} + O(1) = O(1). \end{aligned}$$

For  $\lambda = \mu$ , we obtain similarly

$$\mathbb{E}((F^*)^{-1}(U)) = \int_0^T t f^*(t) dt + T f^*(T) = \int_0^T t \lambda \mu \frac{e^{-\mu t} - e^{-\lambda t}}{\lambda - \mu} dt + T \frac{\lambda e^{-\mu T} - \mu e^{-\lambda T}}{\lambda - \mu} = O(1).$$

Thus the right hand side of (4.21) is of the desired order in both cases by Proposition 2.2.4.  $\square$

### 4.3.7 Proof of Corollary 4.3.6

Let  $f$  denote the density of  $\mathcal{L}(A_{J_T} | Y_T > 0)$  with respect to  $\zeta_T = \text{Leb}_{[0,\infty)} \setminus \{T\} + \delta_T$ . Equation (4.3) implies that

$$f(t) = (1 - c_T) f_*(t) \mathbb{1}_{[0,T)}(t) + c_T f^*(t),$$

where  $f_*$  and  $f^*$  are given by (4.7), (4.8) and (4.13) and

$$c_T = \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right) = \begin{cases} O(Te^{-(\lambda-\mu)T}) & \text{if } \lambda > \mu, \\ O(\log(T)/T) & \text{if } \lambda = \mu. \end{cases} \quad (4.22)$$

by Lemma 2.2.4.

Noting that  $f_*(t) = f^*(t) = 0$  for  $t > T$ , we have

$$\begin{aligned} d_{TV}(\mathcal{L}(A | Y_T > 0), \text{Exp}(\lambda)) &= \frac{1}{2} \int_0^\infty |f(t) - \lambda e^{-\lambda t} \mathbb{1}_{[0, \infty) \setminus \{T\}}| \zeta_T(dt) \\ &\leq \frac{1}{2} \left( (1 - c_T) \int_0^\infty |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^\infty |f^*(t) - \lambda e^{-\lambda t}| dt + c_T f^*(T) \right) \\ &\leq \frac{1}{2} \left( \int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt + c_T \int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt + O(Te^{-\lambda T}) \right). \end{aligned} \quad (4.23)$$

If  $\lambda > \mu$ , we obtain for the first integral on the right-hand side

$$\begin{aligned} \int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt &= \int_0^T \left| \frac{\lambda e^{-(\lambda-\mu)T} e^{-\lambda t} - \mu e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}} \right| dt \\ &= \frac{e^{-(\lambda-\mu)T}}{1 - e^{-(\lambda-\mu)T}} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt = O(e^{-(\lambda-\mu)T}) \end{aligned} \quad (4.24)$$

and for the second integral on the right-hand side

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \frac{\lambda}{\lambda - \mu} \int_0^T |\lambda e^{-\lambda t} - \mu e^{-\mu t}| dt = O(1). \quad (4.25)$$

In conclusion, plugging (4.24), (4.25) and (4.22) into Inequality (4.23), we have that

$$d_{TV}(\mathcal{L}(A | Y_T > 0), \text{Exp}(\lambda)) = O(Te^{-(\lambda-\mu)T}).$$

If  $\lambda = \mu > 0$ , we obtain for the first integral on the right-hand side of Inequality (4.23)

$$\int_0^T |f_*(t) - \lambda e^{-\lambda t}| dt = \frac{1}{T} \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O(1/T). \quad (4.26)$$

and for the second integral

$$\int_0^T |f^*(t) - \lambda e^{-\lambda t}| dt = \lambda \int_0^T e^{-\lambda t} |\lambda t - 1| dt = O(1). \quad (4.27)$$

Thus, plugging (4.26), (4.27) and (4.22) into Inequality (4.23) yields

$$d_{TV}(\mathcal{L}(A | Y_T > 0), \text{Exp}(\lambda)) = O(\log(T)/T).$$

□

### 4.3.8 Simulations

In the following, we present some simulations that confirm the main results of this section.

Firstly, we consider the expression that we derived for the age of the first  $y_T - 1$  individuals<sup>3</sup> given  $Y_T = y_T$  for  $\lambda > \mu$ , i.e. Equation (4.9). We know that, given  $Y_T = y_T$ , these ages correspond to the first  $y_T - 1$  downsteps of the deflected contour process that start in  $T$ , which we denoted by  $V_{J_1}, \dots, V_{J_{y_T-1}}$ . We simulate the contour process and plot the empirical distribution function for  $\lambda = 2$ ,  $\mu = 1$  and

<sup>3</sup> Note that the individuals are still numbered from left to right in the right-aligned tree and that this ordering is rather artificial.

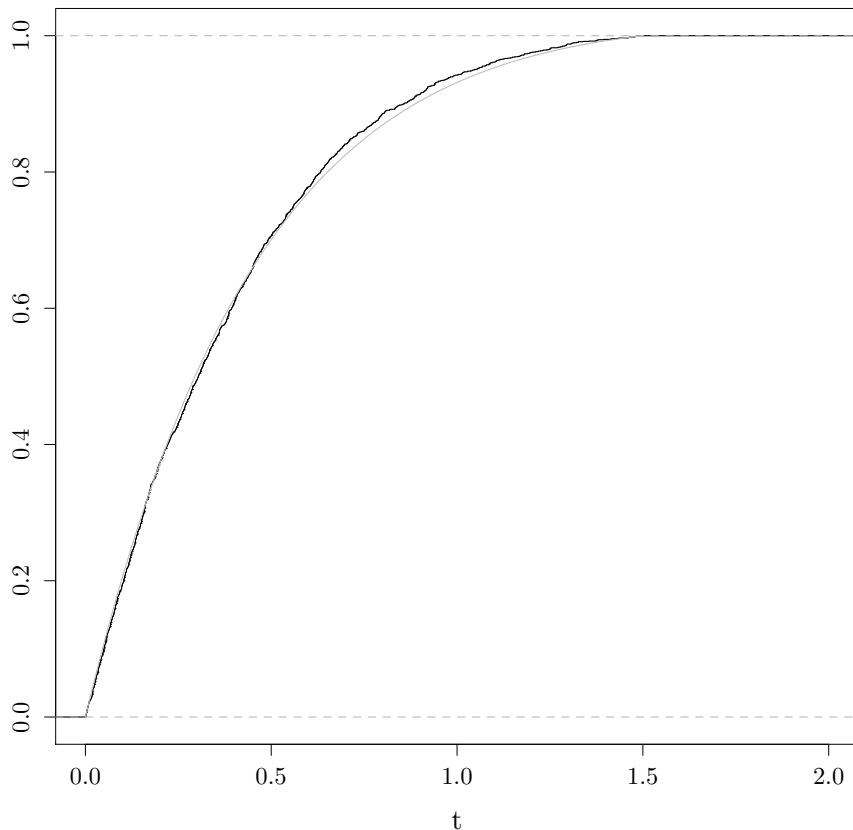


Figure 4.7: True  $F_*$  (grey) and simulation-based empirical distribution function for the downsteps from  $T$  in the contour process corresponding to  $F_*$  (black)

$T = 1.5$  together with the corresponding theoretical cumulative distribution function  $F_*$ . The result is displayed in Figure 4.7, where we used 500 iterations.

For the cumulative distribution function  $F^*$  of the age  $V_{\mathcal{J}_T} \wedge T$  of the last individual given  $Y_T = y_T$ , we derived Equation (4.12). In Figure 4.8, this cumulative distribution function is compared with the empirical distribution function of the last downstep of the deflected contour process given  $Y_T = y_T$ . The parameters and time are chosen as above. Here we used 5000 iterations.

Figures 4.7 and Figure 4.8 illustrate also that we have  $F_*(t) \geq F^*(t)$  for all  $t \geq 0$ , i.e. that the ages of the first  $Y_T - 1$  individuals are stochastically dominated by the age of the last individual. Finally, we consider the age distribution of  $(Y_t)_{t \geq 0}$  at time  $T$ , which is given by Corollary 4.3.2. In order to obtain Figure 4.9 and Figure 4.10, we simulated 5000 realizations of a linear birth and death process up to time  $T$  and plotted the empirical distribution function of the age of an individual picked uniformly at random at time  $T$  together with the theoretical cumulative distribution function from Corollary 4.3.2 and the cumulative distribution function of the asymptotic age distribution, i.e. of the  $\text{Exp}(\lambda)$  distribution. We chose the rates  $\lambda$  and  $\mu$  as above and  $T = 1.5$  and  $T = 2$ , respectively, such that we also can see that the age distribution approaches the asymptotic one.



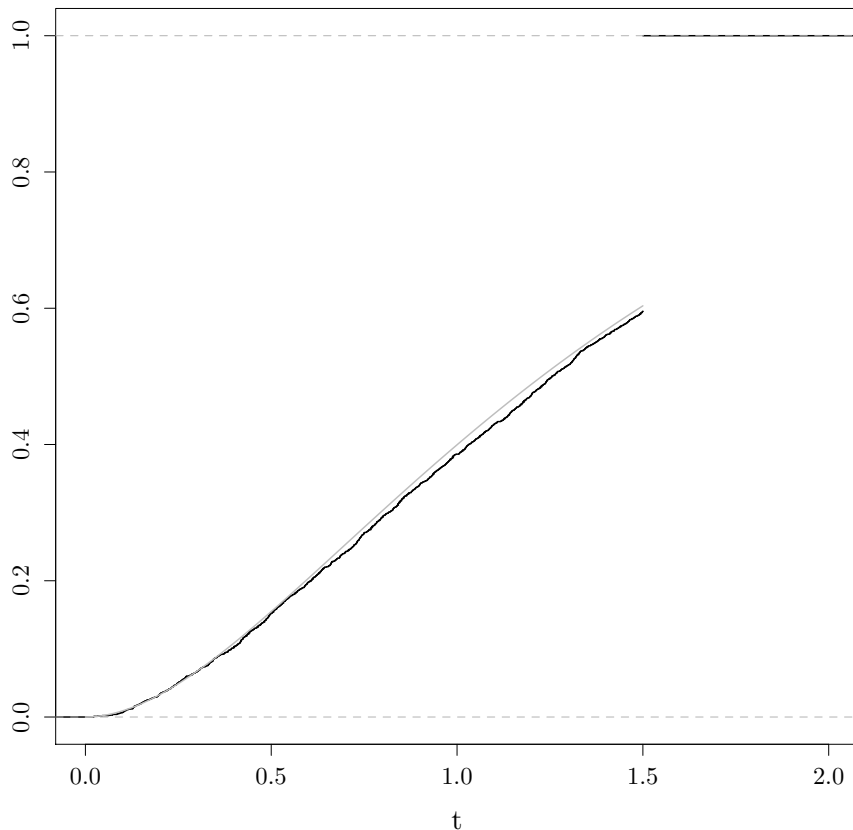


Figure 4.8: True  $F^*$  (grey) and simulation-based empirical distribution function for the downsteps from  $T$  in the contour process corresponding to  $F^*$  (black)

## 4.4 Upper bound for the time since the last event

### 4.4.1 The main result

Let  $0 = T_1 < T_2 < \dots$  be the event times of our linear birth and death process. Since  $\mathfrak{B}_T + \mathfrak{D}_T$  is the number of events up to time  $T$ , the random variable  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  describes the time passed since the last event as seen from time  $T$ . Note that, given the population size  $Y_T = y_T$ , the distribution of the difference  $T_{\mathfrak{B}_T + \mathfrak{D}_T + 1} - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  between the previous and the next event time at time  $T$  is typically larger than the  $\text{Exp}(y_T(\lambda + \mu))$  distribution due to the inspection paradox. Therefore, we do not bound  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  from above by  $T_{\mathfrak{B}_T + \mathfrak{D}_T + 1} - T_{\mathfrak{B}_T + \mathfrak{D}_T}$ . Instead we use that, given  $Y_T = y_T \geq 2$ , the time since the last event is bounded from above by the minimum of the ages of the first  $y_T - 1$  individuals, where the individuals are ordered as in the previous section. We know that these  $y_T - 1$  ages are independent and identically distributed from the proof of Theorem 4.3.1, and their distribution is stochastically dominated by the  $\text{Exp}(\lambda)$  distribution by Remark 4.3.9. Since the minimum of  $y_T - 1$  independent  $\text{Exp}(\lambda)$  distributed random variables is  $\text{Exp}((y_T - 1)\lambda)$  distributed, we obtain the following theorem.

#### 4.4.1 Theorem

Given  $Y_T = y_T$ , the distribution of  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  is stochastically dominated by the distribution with

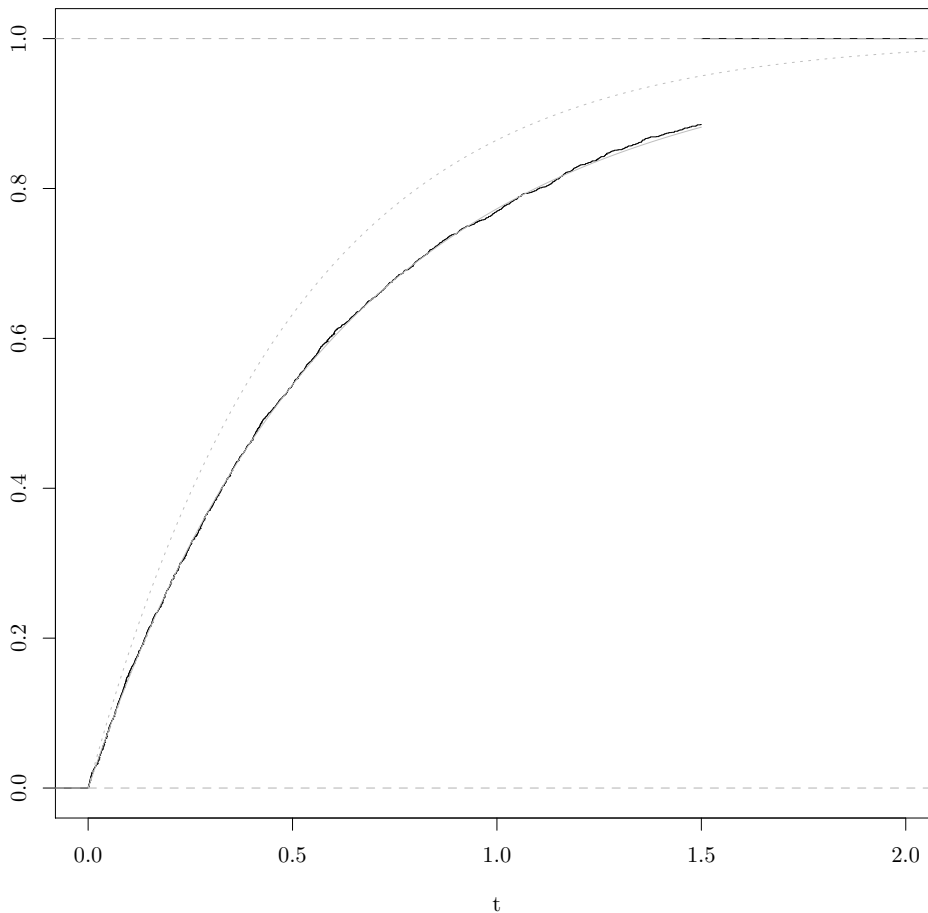


Figure 4.9: Simulation-based empirical distribution function of the distribution of the age of an individual picked uniformly at random for  $T = 1.5$  (black line), the corresponding cumulative distribution function (grey solid line) and the cumulative distribution function of the asymptotic age distribution (grey dotted line)

cumulative distribution function

$$G(t) = \begin{cases} 1 - e^{-(y_T-1)\lambda t} & \text{if } y_T > 1 \\ \mathbb{1}_{[T, \infty)}(t) & \text{if } y_T \leq 1. \end{cases}$$

#### 4.4.2 Comparison with results about reconstructed trees

We compare our result to recent results obtained in the reconstructed tree; see [Ger08].

Note that  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  corresponds to the time from the maximum of all branching times in the tree representing the linear birth and death process up to time  $T$ . In order to obtain an upper bound for the difference between  $T$  and this maximal branching time, we modify the tree in an appropriate way. In order to be consistent with [Ger08], we draw the trees in the common centered way in the following.

**4.4.2 Definition** (see e.g. the introduction in [Ger08])

Consider a binary tree representing a linear birth and death process at time  $T$  and remove the extinct

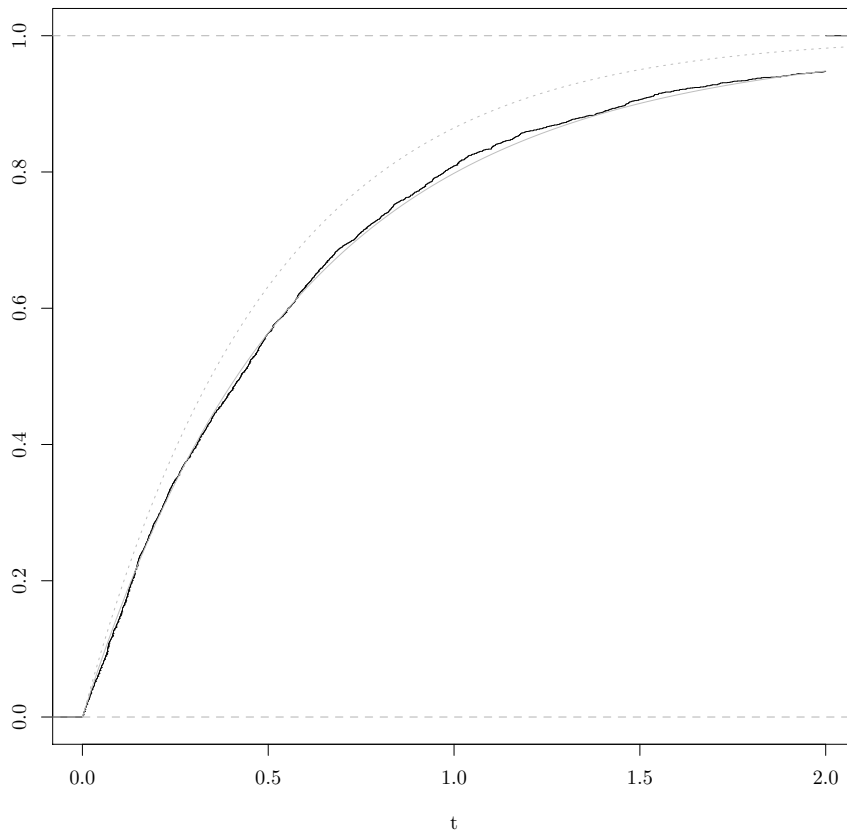


Figure 4.10: Simulation-based empirical distribution function of the distribution of the age of an individual picked uniformly at random for  $T = 2$  (black line), the corresponding cumulative distribution function (grey solid line) and the cumulative distribution function of the asymptotic age distribution (grey dotted line)

lineages, i.e. all edges that do not lead to an edge living at time  $T$  (see Figure 4.11). The resulting tree is called *reconstructed tree*. The original tree is referred to as the *complete tree*.

The tree defined above is called *reconstructed tree* since it is the one that is often considered if only the individuals alive at time  $T$  are observed.

#### 4.4.3 Remark

Since the set of branching times in the reconstructed tree is a subset of all branching times in the complete tree, the time  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  since the last event is smaller than or equal to the minimum of the differences between  $T$  and the branching times in the reconstructed tree. Consequently, we are interested in the distribution of the branching times and these differences.

#### 4.4.4 Definition

Given the population size  $Y_T = y_T$  at time  $T$ , let  $X_1^+ := 0$  and let  $X_2^+, \dots, X_{y_T}^+$  denote the *unordered* (i.e. randomly permuted) *non-zero branching times* in the reconstructed tree.

The following result from [Ger08] gives us the (conditional) distribution of these times.

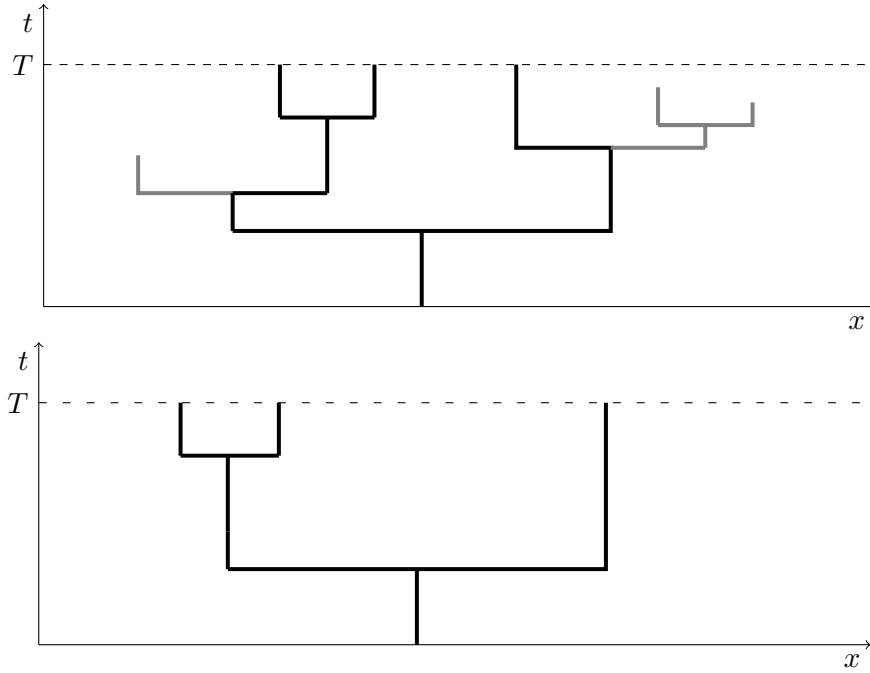


Figure 4.11: A complete tree, where the gray edges belong to the extinct lineages, and the corresponding reconstructed tree

#### 4.4.5 Theorem ([Ger08], Theorem 2.5)

Given  $Y_T = y_T$ , the random times  $T - X_2^+, \dots, T - X_{y_T}^+$  are independent and identically distributed with cumulative distribution function

$$F(t) := \begin{cases} \frac{1 - e^{-(\lambda - \mu)t}}{\lambda - \mu e^{-(\lambda - \mu)t}} \frac{\lambda - \mu e^{-(\lambda - \mu)T}}{1 - e^{-(\lambda - \mu)T}} & \text{if } 0 \leq t \leq T, \\ 1 & \text{if } t > T. \end{cases}$$

#### 4.4.6 Remark

Note that Theorem 4.4.5 allows us to obtain a result for a distribution that is in some sense similar to the age distribution we considered in Section 4.2 and Section 4.3. However, it seems to be unlikely that the results from Section 4.3 can be deduced from known results about reconstructed trees.

The distribution from Theorem 4.4.5 is dominated by a truncated exponential distribution:

#### 4.4.7 Corollary

Given  $Y_T = y_T$ , for each  $j \in \{2, \dots, y_T\}$ , the conditional distribution of  $T - X_j^+$  is stochastically dominated by the  $\text{Exp}(\lambda - \mu)$  distribution truncated at  $T$ .

*Proof:* Given  $Y_T = y_T$ , for each  $j \in \{2, \dots, y_T\}$ , the cumulative distribution function  $F$  of  $T - X_j^+$  is given by Theorem 4.4.5. For  $t > T$ , we have  $F(t) = 1 \geq (1 - e^{-(\lambda - \mu)t}) / (1 - e^{-(\lambda - \mu)T})$ . For  $t \leq T$ , the expression  $\lambda - \mu e^{-(\lambda - \mu)T}$  is larger than or equal to  $\lambda - \mu e^{-(\lambda - \mu)t}$ . Thus  $F(t) \geq (1 - e^{-(\lambda - \mu)t}) / (1 - e^{-(\lambda - \mu)T})$  also holds for  $t \leq T$ .  $\square$

#### 4.4.8 Remark

By Remark 4.4.3, we have that, given  $Y_T$ , the difference  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  is stochastically dominated by

$\min\{T - X_j^+ | j \in \{1, \dots, Y_T\}\}$ . Given  $Y_T = 1$ , we have that  $\min\{T - X_j^+ | j \in \{1, \dots, Y_T\}\}$  is equal to  $T - X_1^+ = T$ . Given  $Y_T = y_T > 1$ , we know that there is an individual that is younger than the first individual and thus  $\min\{T - X_j^+ | j \in \{1, \dots, y_T\}\} = \min\{T - X_j^+ | j \in \{2, \dots, y_T\}\}$ . Since the minimum of  $y_T - 1$  independent  $\text{Exp}(\lambda - \mu)$  distributed random variables is  $\text{Exp}((y_T - 1)(\lambda - \mu))$  distributed for  $y_T > 1$ , Theorem 4.4.5 and Corollary 4.4.7 imply that, given  $Y_T = y_T \in \mathbb{N}$ , the distribution of  $\min\{T - X_j^+ | j \in \{1, \dots, y_T\}\}$  is stochastically dominated by the distribution with cumulative distribution function

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

It follows that, given  $Y_T = y_T$ , the distribution of  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  is stochastically dominated by the distribution with cumulative distribution function  $\tilde{G}$ . Note  $\tilde{G}(t) \leq G(t)$  for all  $t \geq 0$ . Thus the distribution corresponding to  $G$  is stochastically dominated by the distribution with cumulative distribution function  $\tilde{G}$ . We may conclude that we do not obtain a better result, i.e. a better upper bound, using the reconstructed tree than from Theorem 4.4.1. This is plausible because by adding the “missing” branches to the reconstructed tree, we see that the ages in the complete tree cannot be larger than than the ages in the corresponding reconstructed one.



# 5 Random networks

In this chapter, we give a short introduction to the large field of random networks. We refer to [vdH16] and [New10] for extensive surveys. A well-written introduction to the most important random network models can also be found in [DM10].

## 5.1 Introduction to graphs

In the first part of this section, which is based on [New10], [vdH16] and [Die06], we give basic definitions and interpretations. We begin with the definition of a (deterministic) graph, which is the main mathematical concept for modeling networks.

### 5.1.1 Definition

The tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is called a *graph*, where  $\mathcal{V}$  is a countable set and the definition of  $\mathcal{E}$  depends on the type of graph as follows.

For an *undirected* graph without multiple edges and loops,  $\mathcal{E}$  is a subset of  $[\mathcal{V}]^2$ , where  $[\mathcal{V}]^2$  denotes the set of all two-element subsets of  $\mathcal{V}$ .

For a *directed* graph without multiple edges,  $\mathcal{E}$  is a subset of  $\mathcal{V} \times \mathcal{V}$ .

The elements of  $\mathcal{V}$  and  $\mathcal{E}$  are called *nodes* and *edges*, respectively. An edge is called *directed* or *undirected* depending on the type of graph it belongs to. For simplicity, we denote each edge  $\{i, j\}$  of an undirected graph also by  $(i, j)$  (or  $(j, i)$ ). An edge of the form  $(i, i)$ ,  $i \in \mathcal{V}$ , is called *loop*.

For nodes  $i, j \in \mathcal{V}$ , we say there exists an edge between  $i$  and  $j$  or an edge connects  $i$  and  $j$  if and only if  $(i, j) \in \mathcal{E}$ . Furthermore, we say an edge  $(i_1, i_2) \in \mathcal{E}$  is *incident* to a node  $i_3 \in \mathcal{V}$  if and only if  $i_3 = i_1$  or  $i_3 = i_2$ .

An *undirected* multigraph is a tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  of disjoint sets together with a map  $\iota : \mathcal{E} \rightarrow \mathcal{V} \cup [\mathcal{V}]^2$  assigning to every edge either one or two nodes, which are the nodes it is incident to.

Analogously, a *directed* multigraph is a tuple  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  of disjoint sets together with a map  $\iota : \mathcal{E} \rightarrow \mathcal{V} \times \mathcal{V}$ .

For both undirected and directed multigraphs, we identify each edge  $\epsilon \in \mathcal{E}$  with  $\iota(\epsilon)$  and interpret  $\mathcal{E}$  as a *multiset*, which is a generalization of a set that may contain each element multiple times (see e.g. Chapter 1 of [Rig16]). Each element that appears multiple times in this multiset is called *multiple edge*. Furthermore, for an undirected multigraph, each edge  $\epsilon$  with  $\iota(\epsilon) = i$  for some node  $i \in \mathcal{V}$  is called *loop* and denoted by  $(i, i)$ .

Note that we consider multigraphs as special graphs although that is not always the case in the literature (see e.g. Chapter 1 of [Die06]).

Nodes are also referred to as vertices, sites and actors in the literature. Alternative expressions for edges are links, bonds and ties. Note that we consider undirected graphs in this thesis if not stated otherwise.

The nodes are often interpreted as individuals, for example persons, and the edges represent relations between the individuals then, for example friendships. Thus graphs are a very useful tool for analyzing phenomena with an underlying network structure. A lot of interpretations and applications of graphs can, for example, be found in the surveys mentioned above.

Note that multiple edges can be interpreted as weighted edges, where the weight is given by the number of times the edge is contained in  $\mathcal{E}$ . In many application, it is difficult to give useful interpretations for such weights. However, several empirical networks where integer-valued weights for edges are desired can for example be found in Subsection 2.4.2 of [BLM<sup>+</sup>06].

Now we define the degree of a node, which plays a crucial role in this thesis.

### 5.1.2 Definition

The number of edges incident to a node  $i \in \mathcal{V}$  is called *degree* of node  $i$ . Furthermore, for a fixed node  $i$  of a directed graph, the number of edges  $(j_1, j_2)$  such that  $j_2 = i$  is called *in-degree* of node  $i$ . The *out-degree* is defined analogously.

Now we give a definition concerning the connectivity of a graph.

### 5.1.3 Definition (cf. e.g. Section 1.2 and Definition 1.6 of [vdH16])

- (i) A sequence of edges  $(i_1, i_2), (i_2, i_3), \dots, (i_{m-1}, i_m)$ ,  $m \in \mathbb{N}$ , is called *path* between the nodes  $i_1$  and  $i_m$ . We say nodes are *connected* if and only if there exists a path between them.
- (ii) The *graph distance*  $d_{\mathcal{G}}(i, j)$  between nodes  $i$  and  $j$  is the minimal number of edges of paths between  $i$  and  $j$ , where we set  $d_{\mathcal{G}}(i, j) = \infty$  if  $i$  and  $j$  are not connected.
- (iii) Let  $(\mathcal{G}^{(n)})_{n \in \mathbb{N}}$  be a sequence of graphs such that the node set  $\mathcal{V}^{(n)}$  of  $\mathcal{G}^{(n)}$  has exactly  $n$  nodes for all  $n \in \mathbb{N}$ . Let  $\mathfrak{C}^{(n)}(i) := \{j \in \mathcal{V}^{(n)} \mid d_{\mathcal{G}^{(n)}}(i, j) < \infty\}$  for all  $n \in \mathbb{N}$  be the *connected component* of node  $i$  in  $\mathcal{G}^{(n)}$ , and let  $\mathfrak{C}_{\max}^{(n)}$  denote the *maximal connected component*, i.e.  $|\mathfrak{C}_{\max}^{(n)}| = \max_{i \in \mathcal{V}^{(n)}} |\mathfrak{C}^{(n)}(i)|$  for all  $n \in \mathbb{N}$ , where  $|\cdot|$  denotes the cardinality. Then the sequence of graphs  $(\mathcal{G}^{(n)})_{n \in \mathbb{N}}$  is called *highly-connected* if

$$\liminf_{n \rightarrow \infty} \frac{|\mathfrak{C}_{\max}^{(n)}|}{n} > 0,$$

and then  $\mathfrak{C}_{\max}^{(n)}$  is called *giant component*.

### 5.1.4 Remark

The existence of a giant component has important consequences for applications. For example, assume that the graph  $\mathcal{G}^{(n)}$  models a population with  $n$  individuals and the spread of an epidemic such that each individual infects all other individuals it is connected to in the graph. Then connectivity corresponds to the impact of the epidemic. See e.g. [DM10] for a precise modeling of epidemics using (random) graphs.



The small world property describes the phenomenon that individuals are surprisingly close to each other in social networks. It is formalized in the definition below. It was discovered in the population of the USA by the sociologist Stanley Milgram in a famous experiment in 1967, where people were supposed to forward letters addressed to someone they did not know personally to persons in their circle of acquaintances they thought were likely to know the addressee (see [Mil67]).

**5.1.5 Definition** (see (1.2.11) and Definition 1.7 in [vdH16])

- (i) Consider a graph  $\mathcal{G}^{(n)}$  with nodes  $1, \dots, n$ . Let  $J_n^{(1)}$  and  $J_n^{(2)}$  be independent and uniformly distributed on  $\{1, \dots, n\}$ . Then  $d_{typ}(\mathcal{G}^{(n)}) := d_{\mathcal{G}^{(n)}}(J_n^{(1)}, J_n^{(2)})$  is called *typical distance* of  $\mathcal{G}^{(n)}$ .
- (ii) A graph sequence  $(\mathcal{G}^{(n)})_{n \in \mathbb{N}}$  has the *small world property* if and only if there exists a constant  $C > 0$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P}(d_{typ}(\mathcal{G}^{(n)}) \leq C \log(n)) = 1.$$

Note that the choice of the term  $C \log(n)$  in the definition of the small world property can be motivated by the fact that the typical distance in the well-known classical preferential attachment model (defined in Definition 5.4.1 below) is of logarithmic order; see Theorem 7.1 in [vdH15] for details.

**5.1.6 Remark** (see e.g. (1.2.13) in [vdH16] and Section 6.1 of [DM10])

Sometimes the small world property is also defined using the *diameter* of a graph  $\mathcal{G}$ , i.e. the quantity

$$\max_{i, j \in \mathcal{G}} d_{\mathcal{G}}(i, j),$$

instead of the typical distance.

**5.1.7 Remark**

There are many more interesting properties of graphs. For example, one might be interested in the homogeneity of a graph. In this context, the *clustering coefficient*, which is for example defined in Section 1.5 of [vdH16], is an important measure for the clustering in a graph. Intuitively, the clustering coefficient is high if and only if two nodes with a common neighbour are more likely to be connected than those without common neighbours.

Furthermore, each node is often associated with quantities that measure its importance in some sense. For example, a node can be considered important if its degree is high. Further *importance/centrality measures* can for example be found in Section 1.5 of [vdH16] and Chapter 7 of [New10].

We model *random networks* by using random graphs, i.e. graphs where the set of nodes  $\mathcal{V}$  and/or the vertices  $\mathcal{E}$  can be random. We assume that all random variables are defined on an underlying common probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

## 5.2 Heavy tailed and power law distributions

Before we present various random network models, we introduce the concept of power law distributions since, as already mentioned in the introduction, many random network models have such distributions as asymptotic degree distributions (see e.g. Chapter 3 of [DM03]).

Firstly, we define heavy tail and power-law distributions. Note that there are several different definitions of those in the literature.

**5.2.1 Definition** (cf. Section 2.1 of [FKZ11])

A distribution on  $\mathbb{R}$  is *heavy-tailed* if and only if its cumulative distribution function  $F$  satisfies

$$\lim_{x \rightarrow \infty} e^{\vartheta x} (1 - F(x)) = \infty \quad \forall \vartheta > 0.$$

Theorem 2.6 in [FKZ11] gives further characterizations of heavy-tailed distributions.

In particular, all distributions that obey power laws in the weak or strong sense that we define now are heavy-tailed. As in [vdH16], we concentrate on distributions on  $\mathbb{N}_0$  here.

**5.2.2 Definition** (cf. the introduction of [vdH16])

We say a distribution on  $\mathbb{N}_0$  obeys a *power law* in the *strong sense* if and only if the probability mass function  $(f(k))_{k \in \mathbb{N}_0}$  satisfies

$$f(k) \propto k^{-\theta}$$

for all  $k \in \mathbb{N}$  and some  $\theta > 1$ , and in the *weak sense* with power law exponent  $\theta$  if and only if its cumulative distribution function  $F$  satisfies

$$1 - F(m) = \varphi(m)m^{1-\theta} \tag{5.1}$$

for all  $m \in \mathbb{N}$  and some  $\theta > 1$ , where the function  $m \mapsto \varphi(m)$  is *slowly varying* (see Definition 3.3.5).

**5.2.3 Remark**

We know from Chapter 2 of [FKZ11] that distributions that obey a power law in the weak sense are heavy-tailed. Note that these distributions are sometimes also referred to as *regularly varying* distributions (see e.g. Section 2.8 of [FKZ11]).

**5.2.4 Remark** (see e.g. Section 2.1 of [FKZ11])

The Pareto distribution is often referred to as *continuous power law distribution*.

For a survey on power law distributions, including a lot of real-world examples, we refer to [New13].

## 5.3 Static random network models

### 5.3.1 The Erdős-Rényi model and related models

The best-known random network model is the following.

**5.3.1 Definition** (cf. [Gil59])

Let  $\mathcal{V} = \{1, \dots, n\}$  for some  $n \in \mathbb{N}$ . Assume that for each pair of nodes  $(i, j)$  with  $1 \leq i < j \leq n$  the corresponding edge exists with probability  $p$  (referred to as edge probability in the following) independently from the existence of edges corresponding to other pairs of nodes. Then the resulting random graph is called *Erdős-Rényi graph*.

Note that Erdős-Rényi graphs are often simply referred to as random graphs.

Before we consider the degree distribution, we state a very classical result about the connected components of Erdős-Rényi graphs.

**5.3.2 Proposition** (see e.g. Section 2.2 of [DM10])

A Erdős-Rényi graph has almost surely a giant component if and only if  $p \geq 1/n$ .

For more detailed results about giant components, we refer to [DM10].

Before we draw our attention to generalizations of Erdős-Rényi graphs, we briefly consider the degree and a corresponding convergence rate.

**5.3.3 Remark** (cf. e.g. Section 12.3 of [New10])

Obviously, the distribution of the degree of a fixed node  $i \in \mathcal{V}$  is the  $\text{Bin}(n-1, p)$  distribution. From Section 3.2, we know that this distribution can be approximated by a Poisson distribution in a reasonable way if  $p$  is small. Namely, we have  $d_{TV}(\text{Bin}(n-1, p), \text{Po}(\tilde{\lambda})) \leq p$ , where  $\tilde{\lambda} = (n-1)p$ . In particular, for a sequence of Erdős-Rényi graphs with  $n$  nodes and edge probability  $p(n) = \frac{1}{n-1}$ ,  $n \in \mathbb{N} \setminus \{1\}$ , we have that the  $\text{Bin}(n-1, p(n))$  distribution of the degree of a fixed node  $i \in \mathcal{V}$  converges to the  $\text{Po}(1)$  distribution, and we obtain an explicit convergence rate from Section 3.2:

$$d_{TV}(\text{Bin}(n-1, p(n)), \text{Po}(\tilde{\lambda})) \leq \frac{1}{n}.$$

Due to the homogeneity of the Erdős-Rényi graph, we obtain the same asymptotic distribution and convergence rate also for the distribution of the degree of a node picked uniformly at random (cf. (5.4.7) and (5.4.8) in [vdH16]).

The following is a generalization of the Erdős-Rényi model.

**5.3.4 Definition** (see Chapter 6 of [vdH16])

Let  $\mathcal{V} = \{1, \dots, n\}$  for some  $n \in \mathbb{N}$ , and let  $w_1, \dots, w_n$  be positive real numbers, which we refer to as weights. Furthermore, let  $\ell_n := \sum_{i=1}^n w_i$ . Assume that for each pair of nodes  $(i, j)$  with  $1 \leq i < j \leq n$  the corresponding edge exists with probability

$$p_{ij}(n) := \frac{w_i w_j}{\ell_n + w_i w_j} \tag{5.2}$$

independently from the existence of edges corresponding to other pairs of nodes. Then the resulting random graph is called *generalized random graph with deterministic weights*.

Note that we obtain an Erdős-Rényi graph if we choose  $w_i = \frac{np}{1-p}$  for  $i \in \{1, \dots, n\}$  (see Exercise 6.1 in [vdH16]).

Throughout this section, we assume that the following regularity conditions from Chapter 6 of [vdH16] for the weights hold. For the first condition, let  $J_n$  be uniformly distributed on  $\{1, \dots, n\}$  for  $n \in \mathbb{N}$ . We assume that the weight  $w_{J_n}$  of a node picked uniformly at random converges weakly to some random variable  $W$ , i.e. the probability measure  $\frac{1}{n} \sum_{i=1}^n \delta_{w_i}$  converges weakly to a probability measure  $\mathcal{L}(W)$ , where  $\delta_{w_i}$  denotes the Dirac measure in  $w_i$ . Furthermore, we assume that  $\lim_{n \rightarrow \infty} \mathbb{E}(w_{J_n}) = \mathbb{E}(W) > 0$ . These statements are discussed after Condition 6.4 in [vdH16].

**5.3.5 Theorem** (cf. Theorem 6.7 in [vdH16])

Let  $(w_i)_{i \in \mathbb{N}}$  be a sequence of positive real numbers. Let  $D_i^{(n)}$  be the degree of node  $i$ ,  $i \in \{1, \dots, n\}$ , in the generalized random graph  $\mathcal{G}^{(n)}$  with deterministic weights  $w_1, \dots, w_n$  and  $n$  nodes. We have

$$d_{TV}(\mathcal{L}(D_i^{(n)}), \text{Po}(w_i)) \leq \frac{w_i^2}{\sum_{j=1}^n w_j} \left( 1 + 2 \frac{\sum_{k=1}^n w_k^2}{\sum_{l=1}^n w_l} \right)$$

for all  $i \in \{1, \dots, n\}$ .

This theorem has the following consequence for the degree of a node picked uniformly at random.

**5.3.6 Corollary** (cf. Corollary 6.9 in [vdH16])

The degree of a node picked uniformly at random converges weakly to the  $\text{MixPo}(W)$  distribution, where the asymptotic weight  $W$  is defined as above. For any  $m \in \mathbb{N}$ , the degrees of  $m$  nodes picked uniformly at random from  $\{1, \dots, n\}$  are asymptotically independent.

**5.3.7 Remark**

Note that a rate for the convergence stated in Corollary 5.3.6 can be deduced from Theorem 5.3.5 by conditioning on the node picked uniformly at random.

The following example shows that *every* mixed Poisson distribution can be obtained as asymptotic distribution of the degree of a node picked uniformly at random.

**5.3.8 Example** (see page 169 in [vdH16])

Let  $F$  be an arbitrary cumulative distribution function with  $F(0) = 0$ , and let  $(1 - F)^{-1}$  denote the generalized inverse of  $1 - F$  defined by  $(1 - F)^{-1}(y) = \inf\{x \in \mathbb{R} : (1 - F)(x) \leq y\}$ . Furthermore, let  $w_i = (1 - F)^{-1}(i/n)$ . Then it can be easily deduced from (6.1.17) in [vdH16] that the weight  $w_{J_n}$  of a node picked uniformly at random converges weakly to the distribution with cumulative distribution function  $F$ . By Corollary 5.3.6, the asymptotic distribution of the degree of a node picked uniformly at random is thus the  $\text{MixPo}(W)$  distribution, where  $W$  has cumulative distribution function  $F$ . In particular, the  $\text{MixPo}(W)$  distribution can obey a power law in the weak sense (cf. Section 5.2). An example for such a mixed Poisson distribution will be considered in Section 6.5

Since the weights in this example lack a plausible heuristic explanation, we consider the case where the weights are independent and identically distributed random variables now. In this case, we can also obtain every mixed Poisson distribution as asymptotic distribution of the degree of a node picked uniformly at random (see below).

**5.3.9 Definition** (cf. Chapter 6 of [vdH16])

Let  $\mathcal{V} = \{1, \dots, n\}$  for some  $n \in \mathbb{N}$ , and let  $W_1, \dots, W_n$  be positive independent and identically distributed random variables, which we refer to as weights. Furthermore, let  $\mathcal{L}_n := \sum_{i=1}^n W_i$ . Assume that, given  $W_i = w_i$  for  $i \in \{1, \dots, n\}$  and  $\mathcal{L}_n = \ell_n$ , for each pair of nodes  $(i, j)$  with  $1 \leq i < j \leq n$  the corresponding edge exists with probability

$$p_{ij}(n) := \frac{w_i w_j}{\ell_n + w_i w_j}$$

independently from the existence of edges corresponding to other pairs of nodes. Then the resulting random graph is called *generalized random graph with independent and identically distributed weights*.

The asymptotic degree distribution of the generalized random graph with independent and identically distributed weights is given by the following theorem.

**5.3.10 Theorem** (cf. Corollary 6.11 in [vdH16])

Let a sequence  $(W_i)_{i \in \mathbb{N}}$  of positive independent and identically distributed random variables be given. Let  $D_i^{(n)}$  be the degree of node  $i$ ,  $i \in \{1, \dots, n\}$ , in the generalized random graph  $\mathcal{G}^{(n)}$  with independent and identically distributed weights  $W_1, \dots, W_n$  and  $n$  nodes. Then the distribution of  $D_i^{(n)}$  converges weakly to the  $\text{MixPo}(W_i)$  distribution as  $n \rightarrow \infty$  for all  $i \in \{1, \dots, n\}$ .

**5.3.11 Remark**

Note that we can obtain a rate for the convergence of the distribution of the degree of a node picked uniformly at random from Theorem 5.3.5 by conditioning on the weights and on the node picked uniformly at random (cf. the proof of Theorem 5.3.13 and Corollary 5.3.14 below).

As stated in the introduction, the asymptotic degree distribution in the Britton-Lindholm model without loops sketched in the introduction and treated in Chapter 6 is also a mixed Poisson distribution. In order to illustrate the connection between the Britton-Lindholm model and the generalized random graph with independent and identically distributed weights, we modify the latter such that we obtain a static random network model that is in some sense an analogue to the Britton-Lindholm model without loops from the introduction.

**5.3.12 Definition**

Let  $\mathcal{V} = \{1, \dots, n\}$  for some  $n \in \mathbb{N} \setminus \{1\}$ ,  $\alpha > 0$ , and let  $W_1, \dots, W_n$  be positive independent and identically distributed random variables such that  $W_i \leq \alpha/2$  almost surely for all  $i \in \{1, \dots, n\}$ . Assume that, given  $W_i = w_i$  for  $i \in \{1, \dots, n\}$ , for each pair of nodes  $(i, j)$  with  $1 \leq i < j \leq n$  the corresponding edge exists with probability

$$p_{ij}(n) := \frac{\alpha(w_i + w_j)}{n - 1}$$

independently from the existence of edges corresponding to other pairs of nodes. Then the resulting random graph is called *modified generalized random graph* with independent and identically distributed weights.

The following theorem gives us a convergence rate for the distribution of the degree of a fixed node in the modified generalized random graph with independent and identically distributed weights and is proved similarly to the proof of Theorem 6.7 in [vdH16].

**5.3.13 Theorem**

Let  $(W_i)_{i \in \mathbb{N}}$  be a sequence of positive independent and identically distributed random variables with  $\mathbb{E}(W_1^2) < \infty$ . Let  $D_i^{(n)}$  be the degree of node  $i$ ,  $i \in \{1, \dots, n\}$ , in the modified generalized random graph  $\mathcal{G}^{(n)}$  with independent and identically distributed weights  $W_1, \dots, W_n$  and  $n$  nodes. Then we have

$$d_{TV}(\mathcal{L}(D_i^{(n)}), \text{MixPo}(\alpha(W_i + \mathbb{E}(W_i)))) \leq \frac{2\alpha^2(\mathbb{E}(W_1^2) + (\mathbb{E}(W_1))^2)}{n - 1} + \frac{\alpha\sqrt{\text{Var}(W_1)}}{\sqrt{n - 1}}$$

for all  $i \in \{1, \dots, n\}$ , and the distribution of  $D_i^{(n)}$  converges weakly to the  $\text{MixPo}(W_i + \mathbb{E}(W_i))$  distribution as  $n \rightarrow \infty$  for all  $i \in \{1, \dots, n\}$ .

*Proof:* At first, we assume that the weights are given by  $W_i = w_i$ ,  $i \in \mathbb{N}$ . For  $i \neq j$  and  $n \in \mathbb{N}$ , let  $I_{ij}^{(n)}$  be 1 if  $i$  and  $j$  are connected in  $\mathcal{G}^{(n)}$  and 0 otherwise. Then the degree of node  $i$  can be written as  $D_i^{(n)} = \sum_{j=1}^n I_{ij}^{(n)}$ . Note that  $I_{ij}^{(n)} \sim \text{Be}(p_{ij})$  for  $i \neq j$ , where  $p_{ij} = \frac{\alpha(w_i + w_j)}{n-1}$ , and  $I_{ii}^{(n)} = 0$  almost surely for all  $i$ . Note further that  $I_{i1}^{(n)}, \dots, I_{in}^{(n)}$  are independent.

By Inequality (3.6), we thus obtain

$$d_{TV} \left( \mathcal{L}(D_i^{(n)}), \text{Po} \left( \sum_{j \neq i} p_{ij} \right) \right) \leq \sum_{j \neq i} p_{ij}^2 = \sum_{j \neq i} \frac{\alpha^2 (w_i + w_j)^2}{(n-1)^2}. \quad (5.3)$$

Let us now relieve the assumption that the weights  $W_i$ ,  $i \in \mathbb{N}$ , are given, and let

$$X \sim \text{MixPo} \left( \alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1} \right).$$

Then we have

$$\begin{aligned} d_{TV} \left( \mathcal{L}(D_i^{(n)}), \text{MixPo} \left( \alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1} \right) \right) &\leq \sup_{B \subset \mathbb{N}_0} |\mathbb{P}(D_i^{(n)} \in B) - \mathbb{P}(X \in B)| \\ &= \sup_{B \subset \mathbb{N}_0} |\mathbb{E}(\mathbb{P}(D_i^{(n)} \in B | W_1, \dots, W_n)) - \mathbb{E}(\mathbb{P}(X \in B | W_1, \dots, W_n))| \\ &\leq \sup_{B \subset \mathbb{N}_0} \mathbb{E}(|\mathbb{P}(D_i^{(n)} \in B | W_1, \dots, W_n) - \mathbb{P}(X \in B | W_1, \dots, W_n)|) \\ &\leq \mathbb{E} \left( \sup_{B \subset \mathbb{N}_0} |\mathbb{P}(D_i^{(n)} \in B | W_1, \dots, W_n) - \mathbb{P}(X \in B | W_1, \dots, W_n)| \right) \\ &= \mathbb{E}(d_{TV}(\mathcal{L}(D_i^{(n)} | W_1, \dots, W_n), \mathcal{L}(X | W_1, \dots, W_n))). \end{aligned}$$

By Inequality (5.3), the right-hand side is smaller than or equal to

$$\mathbb{E} \left( \alpha^2 \sum_{j \neq i} \frac{W_i^2 + 2W_i W_j + W_j^2}{(n-1)^2} \right) \quad (5.4)$$

because  $\mathcal{L}(X | W_1, \dots, W_n) = \mathcal{L}(X | W_i + \sum_{j \neq i} \frac{W_j}{n-1})$ . Since the sum has  $n-1$  summands, we obtain that (5.4) is equal to

$$\frac{2\alpha^2(\mathbb{E}(W_i^2) + (\mathbb{E}(W_i))^2)}{n-1}.$$

Note that this expression converges to zero for  $n \rightarrow \infty$ . Thus

$$d_{TV} \left( \mathcal{L}(D_i^{(n)}), \text{MixPo} \left( \alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1} \right) \right)$$

converges to zero for  $n \rightarrow \infty$ .

By the triangle inequality, we have

$$\begin{aligned} &d_{TV}(\mathcal{L}(D_i^{(n)}), \text{MixPo}(\alpha(W_i + \mathbb{E}(W_i)))) \\ &\leq d_{TV} \left( \mathcal{L}(D_i^{(n)}), \text{MixPo} \left( \alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1} \right) \right) \\ &\quad + d_{TV} \left( \text{MixPo} \left( \alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1} \right), \text{MixPo}(\alpha(W_i + \mathbb{E}(W_i))) \right) \end{aligned} \quad (5.5)$$

Thus it remains to show that the second summand of the right-hand side of (5.5) converges to zero as  $n \rightarrow \infty$ . In order to do so, we use Theorem 3.4.1 and obtain

$$d_{TV}\left(\text{MixPo}\left(\alpha W_i + \alpha \sum_{j \neq i} \frac{W_j}{n-1}\right), \text{MixPo}(\alpha(W_i + \mathbb{E}(W_i)))\right) \leq \mathbb{E}\left(\left|\frac{\alpha}{n-1} \sum_{j \neq i} W_j - \alpha \mathbb{E}(W_i)\right|\right).$$

Applying  $\mathbb{E}|Z - \mathbb{E}(Z)| \leq \sqrt{\text{Var}(Z)}$  to  $Z = \frac{\alpha}{n-1} \sum_{j \neq i} W_j$  reveals that the right-hand side is smaller than or equal to

$$\sqrt{\text{Var}\left(\frac{\alpha}{n-1} \sum_{j \neq i} W_j\right)} = \frac{\alpha}{n-1} \sqrt{(n-1)\text{Var}(W_i)} = \frac{\alpha}{\sqrt{n-1}} \sqrt{\text{Var}(W_i)}$$

and this expression converges to zero as  $n \rightarrow \infty$ . Since convergence in total variation distance implies weak convergence, the desired result follows.  $\square$

The following corollary is an immediate consequence of Theorem 5.3.13

#### 5.3.14 Corollary

Let  $(W_i)_{i \in \mathbb{N}}$  be a sequence of positive independent and identically distributed random variables with  $\mathbb{E}(W_1^2) < \infty$ . Then in the modified generalized random graph  $\mathcal{G}^{(n)}$  with independent and identically distributed weights  $W_1, \dots, W_n$  and  $n$  nodes, the distribution of the degree of a node picked uniformly at random among all  $n$  nodes converges weakly to the  $\text{MixPo}(\alpha(W_1 + \mathbb{E}(W_1)))$  distribution as  $n \rightarrow \infty$  for all  $i \in \{1, \dots, n\}$ , and we have the same convergence rate as in Theorem 5.3.13.

#### 5.3.15 Remark

Note that we obtain an Erdős-Rényi graph if we set  $W_i = 1$  for all  $i$  and  $\alpha = 1/2$  in the modified generalized random graph with independent and identically distributed weights.

#### 5.3.16 Remark

The weights in the modified generalized random graph model correspond to the social indices  $(S_i)_{i \in \mathbb{N}}$  in the Britton-Lindholm model. Recall that in the latter each node creates edges at rate  $\alpha S_i$  and the second node is picked uniformly among all living nodes. Thus we would expect that the probability that two nodes  $i$  and  $j$  with social indices  $S_i$  and  $S_j$  are connected is in average approximately proportional to  $S_i + S_j$ , which corresponds to the probability  $p_{ij}(n)$  in the modified generalized random graph model.

We obtain a in some sense similar asymptotic degree distribution in both models. The reason for the additional factor  $(1 - e^{-(\beta+\mu)A})/(\beta + \mu)$  in the random variable determining the mixing distribution of the Britton-Lindholm model is that this model is a (time-continuous) dynamic model in contrast to the modified generalized random graph model.

### 5.3.2 The configuration model

In this Section, which is based on Chapter 7 of [vdH16], we consider the case where an arbitrary distribution on  $\mathbb{N}_0$  is given and we introduce a model that produces random graphs with node set

$\{1, \dots, n\}$ , possibly containing multiple edges and loops, whose asymptotic degree distribution, i.e. the distribution of the degree of a node picked uniformly at random, is this given distribution. If we condition on the absence of multiple edges and loops, we obtain a uniform distribution on all graphs with node set  $\{1, \dots, n\}$  that do not have any multiple edges and loops. Thus this model can be used in order to find out if the network is rather “purely” random or if it contains additional structure.

### The configuration model for a given deterministic degree sequence

#### 5.3.17 Definition and Remark (cf. Section 7.2 of [vdH16])

Let  $n \in \mathbb{N}_0$  and  $\mathbf{d} = (d_i)_{i \in \{1, \dots, n\}}$  be a deterministic sequence of positive integers such that the sum  $\sum_{i=1}^n d_i$  is even. We would like to produce a graph with node set  $\{1, \dots, n\}$  such that node  $i$  has degree  $d_i$  for all  $i \in \{1, \dots, n\}$ . If we only consider graphs without multiple edges and loops, this is not always possible (see Section 7.2 of [vdH16]). Therefore, we allow our graphs to have such edges.

Imagine that node  $i$  is equipped with  $d_i$  half-edges for any  $i \in \{1, \dots, n\}$  such that we have  $\sum_{i=1}^n d_i$  in total. Now we pick two of those half-edges uniformly at random among all half-edges and combine them, i.e. if one of the randomly picked half-edge belongs to node  $i$  and the other one to node  $j$ , we add the edge  $(i, j)$ . Then we pick two half-edges uniformly at random among the  $\sum_{i=1}^n d_i - 2$  remaining half-edges, add the corresponding edge and continue this procedure until no half-edge is left. We call the resulting graph *configuration model with degree sequence*  $\mathbf{d}$ .

Note that the first half-edges are not directly picked uniformly at random but in an arbitrary order in the definition of the configuration model in [vdH16]. However, as pointed out in [vdH16], this leads to the same random graph.

Obviously, the configuration model has the desired degree sequence. Moreover, we obtain from the following example that we can obtain any feasible distribution as asymptotic degree distribution.

#### 5.3.18 Example (see Section 7.2 of [vdH16])

Let  $F$  be the cumulative distribution function of an arbitrary distribution on  $\mathbb{N}_0$ , and let  $n \in \mathbb{N}$ . Note that a degree sequence  $\mathbf{d}$  is, apart from the node labels, determined by the sequence  $(n_k)_{k \in \mathbb{N}_0}$  with  $n_k = \sum_{i=1}^n \mathbb{1}_{\{d_i=k\}}$ , i.e.  $n_k$  denotes the number of nodes with degree  $k$ , for all  $k \in \mathbb{N}_0$ . Thus we can always find a corresponding configuration model if the sequence  $(n_k)_{k \in \mathbb{N}_0}$  is given. Let this sequence now be given by  $n_k = \lceil nF(k) \rceil - \lceil nF(k-1) \rceil$  for all  $k \in \mathbb{N}_0$ , where  $\lceil \cdot \rceil$  denotes the ceiling function, and let  $D_{J_n}$  denote the degree of a node picked uniformly at random among  $\{1, \dots, n\}$  in the configuration model with degree sequence  $\mathbf{d}$ , where  $\mathbf{d}$  satisfies  $n_k = \sum_{i=1}^n \mathbb{1}_{\{d_i=k\}}$  for all  $k \in \mathbb{N}_0$ . Then

$$\mathbb{P}(D_{J_n} \leq x) = \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{d_j \leq x\}}$$

converges to  $F(x)$  for all  $x \geq 0$  as  $n \rightarrow \infty$ , i.e. the degree  $D_{J_n}$  converges weakly to the distribution with cumulative distribution function  $F$ .

### The configuration model with independent and identically distributed degrees

Let  $(D_i)_{i \in \{1, \dots, n\}}$  be a sequence of  $\mathbb{N}$ -valued independent and identically distributed random variables. Since the sum of the degrees  $\sum_{i=1}^n d_i$  has to be even in Definition and Remark 5.3.17, the configuration



model with random degrees  $(D_i)_{i \in \{1, \dots, n\}}$  cannot be defined in the most natural way such that, given  $D_i = d_i$  for all  $i \in \{1, \dots, n\}$ , the random graph is just given by the configuration model with degrees  $d = (d_i)_{i \in \{1, \dots, n\}}$  as defined above. Therefore, we add an additional half-edge to node  $n$  if the sum of the degrees would not be even otherwise. Formally, we let

$$\tilde{D}_n = \begin{cases} D_n & \text{if } \sum_{i=1}^n D_i \text{ is even,} \\ D_n + 1 & \text{if } \sum_{i=1}^n D_i \text{ is odd} \end{cases}$$

and define the model as described above but substitute  $(D_i)_{i \in \{1, \dots, n\}}$  by the sequence  $D_1, \dots, D_{n-1}, \tilde{D}_n$ . For convenience, we denote the latter sequence by  $(\tilde{D}_i)_{i \in \{1, \dots, n\}}$ . By this procedure, we obtain a random graph where the degree  $\tilde{D}_{J_n}$  of a node picked uniformly at random among  $\{1, \dots, n\}$  converges weakly to the distribution  $\mathcal{L}(D_1)$  as  $n \rightarrow \infty$  (cf. Section 7.6 of [vdH16]).

### The erased configuration model

As mentioned above, multiple edges and loops are not desired in many applications. Therefore, we erase all such edges in the configuration model now if they exist in the sense that we merge all multiple edges into single edges and ignore loops. This amounts in ignoring the additional edges of a multiple edge, which is the same we did when we considered the Britton-Lindholm model without loops in the introduction. We call the resulting model the *erased configuration model*. The following theorem states that the distribution of the degree of a node picked uniformly at random still converges weakly to  $\mathcal{L}(D_1)$  as  $n \rightarrow \infty$  in this model if the degrees are deterministic and a couple of weak conditions are fulfilled. See Theorem 7.10 in [vdH16] for a similar result. The analogous theorem for independent and identically distributed degrees can be found in [BDML06] and is stated in Theorem 5.3.20 below.

#### 5.3.19 Theorem

Let  $D_{J_n}^{(er)}$  denote the degree of a node picked uniformly at random in the erased configuration model with deterministic degrees. Assume that the degree sequence  $d = (d_i)_{i \in \{1, \dots, n\}}$  satisfies the following regularity conditions (cf. Condition 7.7 in [vdH16]).

- (i) The distribution of the degree  $d_{J_n}$  of a node picked uniformly at random among  $\{1, \dots, n\}$  converges weakly to some asymptotic distribution with cumulative distribution function  $F$ .
- (ii) Let  $D$  have the cumulative distribution function  $F$  from (i). Then we have  $\mathbb{E}(d_{J_n}) \rightarrow \mathbb{E}(D) < \infty$  as  $n \rightarrow \infty$ .

Then we obtain

$$\mathcal{L}(D_{J_n}^{(er)}) \xrightarrow{w} \mathcal{L}(D)$$

as  $n \rightarrow \infty$ .

*Proof:* By Remark 3.1.10, it is sufficient to show that the total variation distance  $d_{TV}(\mathcal{L}(D_{J_n}^{(er)}), \mathcal{L}(D))$  converges to zero as  $n \rightarrow \infty$ , where  $D$  is defined as in (ii).

By the triangle inequality, we have

$$d_{TV}(\mathcal{L}(D_{J_n}^{(er)}), \mathcal{L}(D)) \leq d_{TV}(\mathcal{L}(D_{J_n}^{(er)}), \mathcal{L}(d_{J_n})) + d_{TV}(\mathcal{L}(d_{J_n}), \mathcal{L}(D)).$$

Since we know from above that  $\mathcal{L}(d_{J_n})$  converges weakly to  $\mathcal{L}(D)$  as  $n \rightarrow \infty$ , we obtain that  $d_{TV}(\mathcal{L}(d_{J_n}), \mathcal{L}(D))$  converges to zero as  $n \rightarrow \infty$  by Remark 3.1.10. It remains to show that the total variation distance  $d_{TV}(\mathcal{L}(D_{J_n}^{(er)}), \mathcal{L}(d_{J_n}))$  converges to zero as  $n \rightarrow \infty$  as well. In order to do so, we note that  $d_{TV}(\mathcal{L}(D_{J_n}^{(er)}), \mathcal{L}(d_{J_n}))$  is smaller than or equal to  $\mathbb{P}(D_{J_n}^{(er)} \neq d_{J_n})$  by Theorem 3.1.8. For any node  $i \in \{1, \dots, n\}$ , let  $M_i^{(n)}$  denote the number of edges that are multiple edges or loops and are incident to  $i$ . Then we obtain by conditioning on  $J_n$

$$\mathbb{P}(D_{J_n}^{(er)} \neq d_{J_n}) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(D_i^{(er)} \neq d_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(M_i^{(n)} > 0) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{M_i^{(n)} > 0\}}) \quad (5.6)$$

In order to show that the right-hand side converges to zero, we use the idea of the last part of the proof of Proposition 7.11 in [vdH16]. Note that we know from [vdH16] that  $\max_{i \in \{1, \dots, n\}} d_i = o(n)$  as  $n \rightarrow \infty$  under the conditions (i) and (ii) of the theorem. Thus we can choose a sequence  $(a_n)_{n \in \mathbb{N}}$  such that  $a_n \rightarrow \infty$  and  $a_n \max_{i \in \{1, \dots, n\}} d_i = o(n)$  as  $n \rightarrow \infty$ . For this sequence, we obtain that the right-hand side of (5.6) is smaller than or equal to

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{M_i^{(n)} > 0\}}) \mathbb{1}_{\{d_i \leq a_n\}} + \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\mathbb{1}_{\{M_i^{(n)} > 0\}}) \mathbb{1}_{\{d_i > a_n\}} \leq \frac{1}{n} \sum_{i=1}^n \mathbb{E}(M_i^{(n)}) \mathbb{1}_{\{d_i \leq a_n\}} + \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{d_i > a_n\}}$$

The first summand of the right-hand side converges to zero as  $n \rightarrow \infty$  by (7.3.30) in [vdH16]. The second summand is equal to the percentage of nodes with degree higher than  $a_n$ . Since  $a_n \rightarrow \infty$  as  $n \rightarrow \infty$  and (i) is equivalent to

$$\frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{d_i \leq x\}} \rightarrow F(x)$$

as  $n \rightarrow \infty$  for all  $x \geq 0$  (cf. Exercise 7.5 in [vdH16]), this percentage vanishes also as  $n \rightarrow \infty$ ; see also the proof of Proposition 7.11 in [vdH16].  $\square$

### 5.3.20 Theorem (cf. Theorem 2.1 in [BDML06])

Let  $D_{J_n}^{(er)}$  denote the degree of a node picked uniformly at random in the erased configuration model with independent and identically distributed degrees with finite expectation. Then we have

$$\mathcal{L}(D_{J_n}^{(er)}) \xrightarrow{w} \mathcal{L}(D_1)$$

as  $n \rightarrow \infty$ .

### Further properties of the configuration model

A lot of further results for the configuration model and related models are given in [vdH16]. We only state the ones most interesting to us here.

The asymptotic probability that the random graph from the ordinary configuration model does not have any multiple edges or loops is considered in Theorem 7.11 and Theorem 7.19 in [vdH16]. This asymptotic probability is positive under some weak conditions. Thus instead of applying the erased configuration model, a random graph without multiple edges or loops can also be obtained by repeating the algorithm described in the definition of the configuration model until it produces such

graph. Formally, we condition on the graph having no multiple edges or loops. The corresponding model is called *repeated configuration model* (see also [BDML06]). For any deterministic degree sequence  $d$ , the repeated configuration model with degree sequence  $d$  produces a random graph that is *uniformly* distributed on the set of all graphs without multiple edges or loops and degree sequence  $d$  (see Proposition 7.13 in [vdH16]).

### 5.3.3 Inhomogeneous random graphs

The *inhomogeneous random graph* model is treated extensively in Chapter 2 of [vdH15] and generalizes the generalized random graph model further. In this model, each of the  $n$  nodes is assigned a random type, and the probability of two nodes being connected depends on the types of the nodes in a very general way (see Section 2.2 of [vdH15] for details). There exist a lot of interesting results about this model. In particular, it can be shown that the distribution of the degree of a node picked uniformly at random converges to a mixed Poisson distribution as  $n \rightarrow \infty$  (cf. Theorem 2.4 in [vdH15]).

Theorem 2.7 in [vdH15] relates the asymptotic size of the largest components to the survival probability of a multitype branching process. This result is used in [BLT11] in order to justify results about the largest component in the Britton-Lindholm model sketched in the introduction and treated in Chapter 6. In particular, the authors argue heuristically that the random graphs from the Britton-Lindholm model behave as special inhomogeneous random graphs.

### 5.3.4 Further static random network models

A further very well-known model, which illustrates the small world property, is the random graph model by Strogatz and Watts (see e.g. Section 6.2 of [DM10]).

Several *spatial* random graph models, i.e. models that consider the positions of the nodes, are for example presented in Section IV of [Bar11]. The best-known spatial random graphs are the so-called *geometric random graphs* (see also [Pen03]). Note that those and other static spatial random graphs can be interpreted as special inhomogeneous random graphs (see also the introduction of [Pen15]).

## 5.4 Dynamic random network models

### 5.4.1 Preferential attachment

Probably the most popular random network models are the preferential attachment models, which formalize the "rich-get-richer" mechanism.

#### Time-discrete Preferential attachment

There are several variants of preferential attachment models. Here we first introduce one of them that is inspired by the famous paper [BA99]. It is given by a sequence of *directed* graphs. However, an undirected graph can obviously be obtained from a directed one by interpreting all directed edges as undirected edges. We also present the convergence rate from [PRR13] for this model below.

**5.4.1 Definition** (see e.g. Section 6 of [PRR13])

Let  $\mathcal{G}^{(1)}$  be a graph with one node labeled 1, which has a loop. For  $n \in \mathbb{N}$ , the graph  $\mathcal{G}^{(n+1)}$  is obtained from  $\mathcal{G}^{(n)}$  by adding one node labeled  $n+1$  and one directed edge from the new node  $n+1$  to one of the nodes  $1, \dots, n$  of  $\mathcal{G}^{(n)}$ . The probability that  $(n+1, i)$  is the new edge is proportional to the degree of node  $i$  in  $\mathcal{G}^{(n)}$  for  $i \in \{1, \dots, n\}$ , and the probability that the new edge becomes a loop is  $\frac{1}{2n+1}$ . We call the resulting model *preferential attachment model*.

**5.4.2 Remark**

Note that the model in [BA99] also allows to add more than one edge per time step. However, since this changes the main results about the model only marginally, we focus on the simpler case here.

Let us now define the distribution that turns out to be the asymptotic degree distribution in the preferential attachment model.

**5.4.3 Definition** (cf. Definition 3.1 in [Rei09])

A random variable  $X$  is said to have the *Yule-Simon distribution* with parameter  $\rho > 0$  if and only if

$$\mathbb{P}(X = k) = \rho B(k, \rho + 1) = \frac{\rho \Gamma(k) \Gamma(\rho + 1)}{\Gamma(k + \rho + 1)}$$

for all  $k \in \mathbb{N}$ , where  $B$  and  $\Gamma$  denote the beta and gamma function, respectively.

Now we state a main result from [PRR13], which gives us a convergence rate for the degree distribution of a node picked uniformly at random.

**5.4.4 Theorem** (see Theorem 6.1 in [PRR13])

Let  $D_i^{(n)}$  denote the degree of node  $i$  in  $\mathcal{G}^{(n)}$  for  $i \in \{1, \dots, n\}$  in the preferential attachment model, and let  $J_n$  be a random variable that is uniformly distributed on  $\{1, \dots, n\}$  independent of  $D_i^{(n)}$  for all  $i \in \{1, \dots, n\}$ . If  $X$  has the Yule-Simon distribution with parameter 2, we have

$$d_{TV}(D_{J_n}^{(n)}, X) \leq O\left(\frac{\log(n)}{n}\right)$$

as  $n \rightarrow \infty$ .

**5.4.5 Remark**

Theorem 5.4.4 implies that the asymptotic degree distribution in the preferential model defined above is the Yule-Simon distribution with parameter 2. Since this distribution has probability mass function

$$\frac{4}{k(k+1)(k+2)},$$

it obeys a power law with power law exponent 3 in the weak sense.

As mentioned before, power law degree distributions can be observed in many real-world networks. However, in many empirical networks, exponents have been observed that are significantly different from 3 (see e.g. page 27 in [CSN09]). Therefore, the following generalization is of great interest (see Chapter 8 of [vdH16]). We use the same mechanism as in Definition 5.4.1 except that the probability that  $(n+1, i)$  is the new edge is proportional to the sum of the degree of node  $i$  in  $\mathcal{G}^{(n)}$  and  $\delta - 1$

for  $i \in \{1, \dots, n\}$ , and the probability that the new edge becomes a loop is  $\frac{\delta}{n(\delta+1)+\delta}$ , where  $\delta > 0$ . This modified model is called *affine preferential attachment model*. Note that we again omit the case where more than one edge is added per time step for the reason stated in Remark 5.4.2 above. For the affine preferential attachment model, we have that the distribution of a node picked uniformly at random converges to a distribution that obeys a power law with exponent  $\theta = \delta + 2$  (see (8.4.11), (8.4.12) and Exercise 8.14 in [vdH16]). We give a precise definition for a loop-free version of an affine preferential model that is considered in [Rei09] now, again omitting the addition of more than one edge per time step.

**5.4.6 Definition** (see Section 2.1 of [Rei09])

Let  $\mathcal{G}^{(1)}$  be a graph with  $m_0 \in \mathbb{N}$  nodes labeled  $1, \dots, m_0$  that has no edges. For  $n \in \mathbb{N}$ , the graph  $\mathcal{G}^{(n+1)}$  is obtained from  $\mathcal{G}^{(n)}$  by adding one node labeled  $n+1$  and one directed edge from the new node  $n+1$  to one of the nodes  $1, \dots, n$  of  $\mathcal{G}^{(n)}$ . The probability that  $(n+1, i)$  is the new edge is proportional to the sum of the in-degree (i.e. the total degree minus one) of node  $i$  in  $\mathcal{G}^{(n)}$  and  $\delta$  for a constant  $\delta > 0$  for  $i \in \{1, \dots, n\}$ , and the probability that the new edge becomes a loop is zero. We call the resulting model *affine loop-free preferential attachment model*.

The asymptotic degree distribution is essentially a generalisation of the Yule-Simon distribution that we define now.

**5.4.7 Definition and Remark** (cf. Definition 3.3 in [Rei09])

A random variable  $X$  is said to have the *generalized Yule-Simon distribution* with parameters  $\gamma > 0$  and  $\rho > 0$  if and only if

$$\mathbb{P}(X = k) = \frac{B(k-1+\gamma, \rho+1)}{B(\gamma, \rho)},$$

for all  $k \in \mathbb{N}$ , where  $B$  denotes the beta function. Note that we obtain the Yule-Simon distribution with parameter  $\rho > 0$  for  $\gamma = 1$ .

The following result from [Rei09] gives us a convergence rate for the degree distribution of a node picked uniformly at random.

**5.4.8 Theorem** (see Theorem 8.2 in [Rei09])

Let  $D_i^{(n)}$  denote the degree of node  $i$  in  $\mathcal{G}^{(n)}$  for  $i \in \{1, \dots, n\}$  in the affine loop-free preferential attachment model with parameter  $\delta > 0$ , and let  $J_n$  be a random variable that is uniformly distributed on  $\{1, \dots, n\}$  and independent of  $D_i^{(n)}$  for all  $i \in \{1, \dots, n\}$ . If  $X$  has the generalized Yule-Simon distribution with parameters  $\gamma = \delta$  and  $\rho = \delta + 1$ , we have

$$d_{TV}(D_{J_n}^{(n)} + 1, \mathcal{L}(X)) \leq O\left(\frac{\log(n)}{n}\right)$$

as  $n \rightarrow \infty$ .

**5.4.9 Remark**

Note that the generalized Yule-Simon distribution with parameters  $\gamma > 0$  and  $\rho > 0$  obeys a power law with power law exponent  $\rho + 1$  in the weak sense (see Subsection 3.2.3 of [Rei09]). Consequently, we know from Theorem 5.4.8 that the asymptotic degree distribution in the affine loop-free preferential model with parameter  $\delta > 0$  obeys a power law with power law exponent  $\theta = \delta + 2$  in the weak sense.

#### 5.4.10 Remark

Note that the preferential attachment model defined in Definition 5.4.1 and the affine loop-free preferential attachment model with parameter  $\gamma = 1$  are very similar. In particular, the asymptotic degree distribution is the Yule-Simon distribution with parameter 2 for both models. Furthermore, the rates for the convergence towards this distribution that are given by Theorem 5.4.4 and Theorem 5.4.8, respectively, are the same. Note that the former theorem was proven in [PRR13] in 2013 by applying Stein's method whereas the result stated in the latter was already proven in 2009 [Rei09] by using recurrence equations. It seems to be very likely that the rate from Theorem 5.4.4 could also be obtained by slightly adapting the proof from the analogous result in [Rei09]. It is worth mentioning that the procedure in [PRR13] is more similar to ours in Chapter 6 and Chapter 7 in the following sense: In [PRR13], the authors use that the asymptotic degree distribution is a mixture of geometric distributions and apply general approximation results for geometric distributions. In Chapter 6 and Chapter 7, the asymptotic degree distribution is a mixture of Poisson distributions and we thus apply general approximation results for (mixed) Poisson distributions. An approach that is even more similar to ours in those chapters can be found in Chapter 7 of [Rei09], where a rate that is slightly worse than the one from Theorem 5.4.8 above is obtained for the affine loop-free preferential attachment model in the following way. Reinert uses that the asymptotic degree distribution as well as the degree distribution of a node picked uniformly at random is a mixture of negative binomial distributions and consequently applies general approximation results for negative binomial distributions.

#### 5.4.11 Remark

Further publications considering other variations of discrete-time preferential attachment are for example summarized in Section 2.2 of [Rei09], Section VIII.A of [AB02] and Subsection 2.3.5 of [BLM<sup>+</sup>06]. *Spatial* variants of preferential attachment models are for example considered in Subsection IV.D. of [Bar11].

#### 5.4.12 Remark

In [DHH10], it is shown for three different preferential attachment models that they have the small world property.

### Continuous-time preferential attachment

Now we introduce the continuous-time preferential attachment model introduced in Subsection 3.3.5 of [Rei09]. As pointed out in Section 3.3 of [Rei09], this model is very similar to the well-known population model that was considered by Yule in 1925 (see [Yul25]). The connection between continuous-time preferential attachment and Yule's population model is also considered in [PPS15].

#### 5.4.13 Definition (see Subsection 3.3.5 of [Rei09])

Let the node population develop according to a linear birth process with per-capita birth rate  $\lambda$ . Each living node gives birth to half-edges independently from everything else at a rate that is proportional to the sum of  $\delta$  for  $\delta > 0$  and the current number of its half-edges. The half-edges are joined together in the order of their births. The resulting model is called *continuous-time preferential attachment model* with parameters  $\delta$  and  $\lambda$ .

**5.4.14 Remark**

There are alternative ways of joining together half-edges that are also considered in Subsection 3.3.5 of [Rei09].

The asymptotic degree distribution is essentially the generalized Yule-Simon distribution from Definition 5.4.7. More precisely, the asymptotic distribution of the sum of one and the degree of a randomly picked node in the continuous-time preferential attachment model with parameters  $\delta$  and  $\lambda$  is the generalized Yule-Simon distribution with parameters  $\delta$  and  $\lambda$  (see Section 3.6 of [Rei09]). The following theorem gives us rates for the convergence to this asymptotic degree distribution in terms of the number of nodes.

**5.4.15 Theorem** (see Theorem 3.22 in [Rei09])

Let  $T_n$  and  $T_{n+1}$  be the birth time of node  $n$  and  $n + 1$  for  $n \in \mathbb{N}$  in the continuous-time preferential attachment model with parameters  $\delta \geq 1$  and  $\lambda > 1$ , respectively, and let  $\vartheta \in (0, 1)$ . Furthermore, let  $Z$  be a random variable that is distributed according to the generalized Yule-Simon distribution with parameters  $\delta$  and  $\lambda$ , and let  $D_{J_{T_n + \vartheta T_{n+1}}}^{(T_n + \vartheta T_{n+1})}$  denote the degree of a randomly picked node at time  $T_n + \vartheta T_{n+1}$ . Then we have

$$d_{TV}(\mathcal{L}(D_{J_{T_n + \vartheta T_{n+1}}}^{(T_n + \vartheta T_{n+1})} + 1), \mathcal{L}(Z)) \leq \frac{1}{n + 1} + \frac{\delta}{(n + 1)(\lambda - 1)}$$

for all  $n \in \mathbb{N}$ .

**5.4.2 Uniform attachment**

Apart from preferential attachment, we also consider uniform attachment.

**5.4.16 Definition** (see e.g. Section 5 of [PRR13])

Let  $\mathcal{G}^{(1)}$  be a graph with one node labeled 1, which has a loop. For  $n \in \mathbb{N}$ , the graph  $\mathcal{G}^{(n+1)}$  is obtained from  $\mathcal{G}^{(n)}$  by adding one node labeled  $n + 1$  and one directed edge from the new node  $n + 1$  to one of the nodes  $1, \dots, n + 1$  of  $\mathcal{G}^{(n+1)}$ . The probability that  $(n + 1, i)$  is the new edge is  $\frac{1}{n+1}$  for  $i \in \{1, \dots, n + 1\}$ . We call the resulting model *uniform attachment model*.

It is well-known that the asymptotic distribution of the degree, which is the in-degree plus one, of a node picked uniformly at random obeys the  $\text{Geo}(1/2)$  distribution, which is given by the probability mass function  $(1/2)^k$ . A rather heuristic proof for this result can be found in Section 4 of [BRS<sup>+</sup>01]. A different and more rigorous proof (for a different initial condition) is presented in Section 2.5 of [VR07].

The following theorem from [PRR13] gives us a rate for the convergence towards this degree distribution. It can be proved by using Stein's method (see Section 5 of [PRR13]).

**5.4.17 Theorem** (see Theorem 5.1 in [PRR13])

Let  $D_i^{(n)}$  denote the degree of node  $i$  in  $\mathcal{G}^{(n)}$  for  $i \in \{1, \dots, n\}$  in the uniform attachment model, and let  $J_n$  be a random variable that is uniformly distributed on  $\{1, \dots, n\}$  independent of  $D_i^{(n)}$  for all  $i \in \{1, \dots, n\}$ . If  $X$  has the  $\text{Geo}(1/2)$  distribution, we have

$$d_{TV}(D_{J_n}^{(n)}, X) \leq \frac{1}{n}.$$

#### 5.4.18 Remark

Note that the case where  $m > 1$  edges are added per time step is also considered in [BRS<sup>+</sup>01]. In this case, the asymptotic degree distribution is the  $\text{Geo}(\frac{1}{m+1})$  distribution (see Section 4 of [BRS<sup>+</sup>01]). In a further variant of uniform attachment, which is treated in [CHK<sup>+</sup>01], the new edge is only added with some fixed positive probability  $\delta$  and is not necessarily incident to the new node. Instead, *both* nodes of a newly added edge are picked uniformly at random. The asymptotic distribution of the degree of a node picked uniformly at random turns out to be a geometric as well, namely the  $\text{Geo}(\frac{1}{2\delta+1})$  distribution (see Section III of [CHK<sup>+</sup>01]).

#### 5.4.19 Remark

A *continuous-time* uniform attachment model is briefly considered in Section 12.2 of [Rei09].

### 5.4.3 Fitness models

Now we consider another class of models that can lead to power law degree distributions with a high flexibility in the power law exponent. The so-called fitness models depict the "fit-get-richer" in addition to or rather than the "rich-get-richer" mechanism. That means that each node is assigned a fitness at its birth that influences its probability to receive edges.

We first consider models that combine fitness and preferential attachment. Then we present models that do not have any preferential attachment component, which we call pure fitness models. Note that also the Britton-Lindholm model is a pure fitness model.

For some real-world networks, fitness is the crucial factor for their evolution. For example, it is concluded in Section 4 of [HW13] that "efficient supply chain systems demonstrate a 'fit-gets-richer' mechanism of growth". Note that pure fitness models still cover a large range of power law exponents, which make them even more attractive for applications.

An introduction to fitness models can also be found in Chapter 2 of [TP11].

#### Preferential attachment with fitness

The most popular fitness model was introduced in the well-known paper [BB01] and is often referred to as *Bianconi–Barabási model* (see e.g. Section 6.2 of [Bar13]). In this model, we equip every node with a positive fitness (or social index)  $S_i$ , where the  $S_i$  are independent and identically distributed. The attachment probability is proportional to the product of  $S_i$  and the degree of node  $i$ . If, for example, the  $S_i$  are uniformly distributed on  $[0, 1]$ , the corresponding asymptotic distribution of the degree of a node picked uniformly at random obeys a power law with exponent 2.255 in the weak sense (see page 439 in [BB01]).

In [ER02], the authors consider a fitness model where the attachment probability is proportional to the *sum* of the fitness and the degree. They show that the corresponding asymptotic distribution of the degree of a node picked uniformly at random obeys a power law with exponent  $2 + \mathbb{E}(S)$  in the weak sense for a fitness distribution with finite mean  $\mathbb{E}(S)$  (see Section 2 of [ER02]).



### Pure fitness models

For some applications, we expect that preferential attachment does not have any significant impact on the network evolution as mentioned above. Thus we consider a pure fitness model now. Note that the generalized random graph model can be seen as a static fitness model. A more general static fitness model was introduced in [CCRM02]. Here we focus on the dynamic fitness model from [SHR13] by Smolyarenko, Hoppe and Rodgers. We do not consider the optional rewiring component. Like this, the model is more similar to both the uniform and preferential attachment models and the Britton-Lindholm model, and more results are known.

#### 5.4.20 Definition (see Section II of [SHR13])

Let  $(S_i)_{i \in \mathbb{N}}$  be independent identically distributed positive random variables with finite mean  $\mathbb{E}(S)$  and strictly monotonically increasing, continuously differentiable cumulative distribution function  $G$  and density  $g$ , and let  $f : (0, \infty)^2 \rightarrow (0, \infty)$  be an arbitrary function. We begin with a graph  $\mathcal{G}^{(1)}$  with one node labeled 1 with fitness  $S_1$ . For  $n \in \mathbb{N}$ , the graph  $\mathcal{G}^{(n+1)}$  is obtained from  $\mathcal{G}^{(n)}$  by adding one node labeled  $n+1$  with fitness  $S_{n+1}$  and one (undirected) edge from the new node  $n+1$  to one of the nodes  $1, \dots, n+1$  of  $\mathcal{G}^{(n+1)}$ . Given  $S_1 = s_1, \dots, S_{n+1} = s_{n+1}$ , the probability that  $(n+1, i)$  is the new edge is

$$\frac{f(s_{n+1}, s_i)}{\sum_{j=1}^{n+1} f(s_{n+1}, s_j)}$$

for  $i \in \{1, \dots, n+1\}$ . We call the resulting model *SHR model*.

Let

$$\tilde{\varrho}(u, n) = \frac{1}{n} \int_0^\infty \frac{f(x_1, G^{-1}(u))}{\int_0^\infty f(x_1, x_2)g(x_2)dx_2} g(x_1)dx_1 = \frac{1}{n} \int_0^1 \frac{f(G^{-1}(v), G^{-1}(u))}{\int_0^1 f(G^{-1}(v), G^{-1}(w))dw} dv,$$

which is approximately the probability for a node with fitness  $G^{-1}(u)$  to receive a new edge at time  $n$  for large  $n$  (see Section II of [SHR13]). Furthermore, let  $\varrho(u) = n\tilde{\varrho}(u, n)$ . The asymptotic probability mass function of the degree of a node picked uniformly at random in the SHR model is then given by

$$p_k = \int_0^1 \frac{1}{1 + \varrho(u)} \left( \frac{\varrho(u)}{1 + \varrho(u)} \right)^{k-1} du \quad (5.7)$$

by Equation (7) in [SHR13], i.e. the asymptotic degree distribution is mixed geometric (namely the  $\text{Geo}(\frac{1}{1+\varrho(U)})$  distribution, where  $U$  is uniformly distributed on  $[0, 1]$ ).

#### 5.4.21 Example

For a constant linking function  $f$ , the SHR model is just the uniform attachment model. In this case, the function  $\varrho$  is also constant and equal to 1. Thus Equation (5.7) implies that the asymptotic distribution of the degree of a node picked uniformly at random is the  $\text{Geo}(\frac{1}{2})$  distribution for the uniform attachment model. This corresponds to the result stated above.

We assume now that  $\varrho$  is continuously differentiable and strictly monotone, i.e.  $\varrho$  has a continuously differentiable inverse  $\varrho^{-1}$ . By substitution, we then can derive

$$p_k = \int_{\varrho(0)}^{\varrho(1)} (\varrho^{-1})'(y) \frac{1}{1+y} \left( \frac{y}{1+y} \right)^{k-1} dy$$

from (5.7). Since  $(\varrho^{-1})'(y) = \frac{1}{\varrho'(\varrho^{-1}(y))}$ , we finally obtain

$$p_k = \int_{\varrho(0)}^{\varrho(1)} \frac{1}{\varrho'(\varrho^{-1}(y))} \frac{1}{1+y} \left( \frac{y}{1+y} \right)^{k-1} dy \quad (5.8)$$

for the asymptotic probability mass function of the degree of a node picked uniformly at random (cf. Equation (10) in [SHR13]).

Now we consider a special case that leads to a model that is in some sense very similar to the Britton-Lindholm model. Let the linking function be given by  $f(x_1, x_2) = x_2$ , and let  $\mathbb{E}(S) = 1$ . Then we obtain

$$\varrho(u) = \frac{\int_0^\infty G^{-1}(u)g(x_1)dx_1}{\int_0^\infty x_2g(x_2)dx_2} = \frac{G^{-1}(u)}{\mathbb{E}(S)} = G^{-1}(u).$$

This implies  $\varrho^{-1}(y) = G(y)$  and  $\varrho'(u) = (G^{-1})'(u) = \frac{1}{G'(G^{-1}(u))} = \frac{1}{g(G^{-1}(u))}$ . Consequently, Equation (5.8) implies

$$p_k = \int_0^{G^{-1}(u)} \frac{1}{1+y} \left( \frac{y}{1+y} \right)^{k-1} g(y) dy.$$

Note that we rediscover this mixture of geometric distributions in Remark 6.7.2, where we treat a version of the model by Britton and Lindholm without loops in which neither nodes nor edges can die.

#### 5.4.22 Remark

If we choose  $f(x_1, x_2) = x_2$  as above and, additionally, the lognormal distribution as fitness distribution, we are in the setting of the lognormal fitness attachment model that is analyzed numerically in [GKST10]. The choice of the lognormal distribution as fitness distribution is motivated in Section 2 of [GKST10].

#### 5.4.4 Node and edge deletion

Obviously, in many real-world networks nodes can die. This is taken into account in some models. For example, in [SR04], a preferential attachment model with node deletion is considered, and it is shown that the asymptotic degree distribution is still a power law distribution in the weak sense, but with an exponent depending on the rate at which nodes are deleted (see Section II of [SR04]). In [CFV04], it is proved that the same holds also for edge deletion.

In [KSR08], node deletion is added to the popular fitness model from [BB01] mentioned above. It is shown that the power law exponent of the asymptotic degree distribution does not depend on the rate at which nodes are deleted if the fitness distribution is chosen to be a truncated exponential distribution. The truncated exponential distribution is concluded to be the suitable fitness distribution for the World Wide Web in [KSR08].

Further results for models with node deletion can also be found in [MGN06].

#### 5.4.5 A further dynamic network model

The *actor-based network model* introduced in [Sni01] and [SvdBS10] is a continuous-time directed random graph model that is very popular in social science. In this model, each node chooses an optimal new neighbour based on an individual objective function each time it creates a new edge. This function may depend on the current network structure as well as on quantities associated to the nodes. See [Sni01] and [SvdBS10] for details.



# 6 The network model by Britton and Lindholm

## 6.1 The model

In this chapter, we consider the *loop-free* version of the original *dynamic network model* by Britton and Lindholm (see [BL10] and [BLT11]) sketched in the introduction. We first give a precise definition of the model.

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. 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 that  $(Y_t)_{t \geq 0}$  has right-continuous trajectories.

We assume  $\lambda > \mu$ , so that the node process  $(Y_t)_{t \geq 0}$  is a supercritical continuous-time Markov branching process. Thus we can apply the results from Chapter 2 to the node process. In particular, the random variable  $\tilde{W} := \lim_{t \rightarrow \infty} Y_t e^{-(\lambda - \mu)t}$  exists almost surely and satisfies  $\mathbb{P}(\tilde{W} = 0) = \mu/\lambda$  and  $\mathcal{L}(\tilde{W} | \tilde{W} > 0) = \text{Exp}(\frac{\lambda - \mu}{\lambda})$ .

We equip every node  $i$  with a positive random *social index*  $S_i$ , where the  $(S_i)_{i \in \mathbb{N}}$  are independent and identically distributed with finite expectation and independent of all other random variables.

This allows us to define the development of the edge process. 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  $\alpha S_i$  and per-edge death rate  $\beta$ , where  $\alpha, \beta > 0$ .<sup>4</sup> 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.

### 6.1.1 Remark

Note that the proofs become slightly simpler if we consider the original model by Britton and Lindholm, which allows loops - essentially because times where  $Y_t = 1$  need not obtain special treatment. The upper bounds remain largely the same; see also Remark 6.3.3 for the pure birth case.

### 6.1.2 Remark

Britton and Lindholm also introduced a modified model where the “second” node of each newly born

---

<sup>4</sup> We also examine the case where  $\beta = 0$  in Section 6.7.

edge is not picked uniformly, but from a probability distribution that is proportional to the social indices of the living nodes. 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.

We briefly mention some results about the model by Britton and Lindholm now, which are stated in [BL10], [BLT11] and [Küc12]. Note that the authors give rather heuristic arguments for their results instead of rigorous proofs in [BL10], [BLT11].

In Section 4 of [BL10], it is argued that the clustering coefficient (cf. Remark 5.1.7) vanishes as time  $T$  goes to infinity. Furthermore, results about a phase transition and the type distribution of neighbouring nodes, where type refers to the age and the social index, are considered in [BLT11]. These results from [BL10] and [BLT11] are also presented in [Küc12] in a more detailed way. The degree correlation, i.e. the correlation of the degrees of two connected nodes, is considered in [BLT11] as well. However, the corresponding proof is wrong as it was pointed out in Subsection 1.10.2 of [Küc12]. Moreover, in Section 1.11 of [Küc12], it is examined how the asymptotic degree distribution depends on the distribution of the social indices.

In the appendix of [BL10], Britton and Lindholm give (rather heuristic) arguments for the negligibility of multiple edges, which are not sufficient to find convergence rates for the case where we ignore multiple edges though. Thus we treat the negligibility of multiple edges rigorously in Section 6.4.

Numerical evaluations of the asymptotic degree distribution can be found in Section 5 of [BL10] and Section 1.12 of [Küc12].

One of the most important characteristics of a random network is its degree distribution (see e.g. [vdH16]), which we focus on here. As mentioned in the introduction, we derive rates for the convergence of the degree distribution towards an asymptotic degree distribution. As a by-product, we obtain a rigorous proof of the convergence result itself.

At the end of this chapter, we furthermore reveal a couple of additional interesting properties of our network model. In particular, in Section 6.5, we prove that the asymptotic degree distribution can exhibit heavy tails and power laws, which often occur in real random networks as mentioned in the introduction.

## 6.2 Degree distribution at finite time

In this section, we deduce the degree distribution at finite time  $T$ .

Firstly, we consider the case where the nodes do not die, i.e.  $\mu = 0$ . 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  $\{1, \dots, y_T\}$  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 chosen at time  $T$  among all living nodes.

### 6.2.1 The number of outgoing edges in the pure birth case

Firstly, we condition on the population size  $y_T$ , the randomly picked node  $j_T$ , its social index  $s_{j_T}$  and its age  $a_{j_T}$  at time  $T$ . Furthermore, we condition on the age  $a_2$  of the second node if  $y_T > 1$  and set  $a_2 = 0$  otherwise. If  $y_T > 1$ , the number of living edges that  $j_T$  creates during its lifetime can be described by a birth and death process with constant birth and linear death rate started in zero at time  $T - a_{\max(j_T, 2)}$  since no edges are created as long as there is only one node. The birth rate is  $\alpha s_{j_T}$  and the death rate is the product of  $\beta$  and the number of (living) edges that  $j_T$  has created. Such a birth and death process is examined in Chapter XVII of [Fel67] and can be interpreted as a queueing system with constant arrival rate and infinitely many servers. We denote the state probabilities at time  $t$  of such a process started at a general state  $i$  with constant birth rate  $\tilde{\lambda}$  and linear death rate  $n\tilde{\mu}$  by  $P_n(t)$ , where  $n$  is the current state of the process. This process is described by the following Kolmogorov forward differential equations:

$$\frac{dP_n(t)}{dt} = \tilde{\lambda}P_{n-1}(t) - (\tilde{\lambda} + n\tilde{\mu})P_n(t) + \tilde{\mu}(n+1)P_{n+1}(t) \text{ for } n \geq 1, \quad (6.1)$$

$$\frac{dP_0(t)}{dt} = \tilde{\mu}P_1(t) - \tilde{\lambda}P_0(t). \quad (6.2)$$

Thus we have for the generating function  $P(s, t) = \sum_{n=0}^{\infty} P_n(t)s^n$  for all  $s$  with  $|s| < 1$ :

$$\begin{aligned} \frac{\partial P(s, t)}{\partial t} &= \frac{\partial \sum_{n=0}^{\infty} P_n(t)s^n}{\partial t} = \sum_{n=0}^{\infty} \frac{dP_n(t)}{dt} s^n \\ &= \sum_{n=1}^{\infty} s^n \left( \tilde{\lambda}P_{n-1}(t) - (\tilde{\lambda} + n\tilde{\mu})P_n(t) + \tilde{\mu}(n+1)P_{n+1}(t) \right) + \tilde{\mu}P_1(t) - \tilde{\lambda}P_0(t) \\ &= (1-s) \left( -\tilde{\lambda}P(s, t) + \tilde{\mu} \frac{\partial P(s, t)}{\partial s} \right), \end{aligned} \quad (6.3)$$

where the second line follows from the Kolmogorov forward differential equations (6.1) and (6.2). The solution of the partial differential equation (6.3) is given by (cf. page 481 in [Fel67])

$$P(s, t) = \exp\left(-\tilde{\lambda}(1-s) \frac{(1-e^{-\tilde{\mu}t})}{\tilde{\mu}}\right) (1 - (1-s)e^{-\tilde{\mu}t})^i. \quad (6.4)$$

The function  $P(\cdot, t)$  is the product of the generating function of the  $\text{Po}\left(\frac{\tilde{\lambda}}{\tilde{\mu}}(1-e^{-\tilde{\mu}t})\right)$  and the  $\text{Bin}(i, e^{-\tilde{\mu}t})$  distribution and hence the generating function of their convolution. In particular, we obtain for  $i = 0$  the generating function of the Poisson distribution with parameter

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

This result can be applied to the edge process we are interested in: Conditioned on the population size  $y_T$ , the chosen node  $j_T$ , its social index  $s_{j_T}$  and age  $a_{j_T}$ , the distribution of the number of edges created by  $j_T$  is the following Poisson distribution:

$$\text{Po}\left(\frac{\alpha s_{j_T}}{\beta} (1 - e^{-\beta a_{\max(j_T, 2)}})\right),$$

where the  $\text{Po}(0)$  distribution is the Dirac measure in zero. Note that we obtain the Dirac measure if only one individual is alive at time  $T$ , i.e.  $y_T = 1$ , since we set  $a_2 = 0$  in this case.

Consequently, the corresponding unconditional distribution is the following mixed Poisson distribution:

$$\text{MixPo}\left(\frac{\alpha S_{J_T}}{\beta} (1 - e^{-\beta A_{\max(J_T, 2)}(T)})\right).$$

## 6.2.2 The number of incoming edges in the pure birth case

Recall that the nodes are ordered by their birth times. Firstly, 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  $T = a_1 > \dots > a_{y_T}$  denote the corresponding ages of the individuals.

Consider a fixed node  $i \in \{1, \dots, y_T\} \setminus \{j_T\}$  and some time interval of the form  $[T - a_l, T - a_{l+1})$  for  $l \geq i \vee 2$ , where  $\vee$  denotes the maximum operator. We only consider edges that are created by the fixed node  $i$ . Then 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 again. The population size is constant and equal to  $l$  in the interval  $[T - a_l, T - a_{l+1})$ . Thus the birth rate is also constant and equal to  $\alpha s_i \frac{1}{l-1}$  because node  $i$  creates edges at rate  $\alpha s_i$  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  $\beta$ . As in the previous subsection, we obtain a Poisson distribution, i.e. the number of edges that are created by  $i$  in  $[T - a_l, T - a_{l+1})$ , connect  $i$  to  $j_T$  and survive until  $T - a_{l+1}$  follows the distribution

$$\text{Po}\left(\frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l - a_{l+1})})\right).$$

Now the distribution of the number of edges that  $i$  creates in  $[T - a_l, T - a_{l+2})$ , connect  $i$  to  $j_T$  and survive until  $T - a_{l+2}$  can be deduced from (6.4). In order to do this, let  $Z$  be the number of edges that  $i$  creates in  $[T - a_l, T - a_{l+1})$ , connect  $i$  to  $j_T$  and survive until  $T - a_{l+1}$ . We have already shown that  $Z$  is  $\text{Po}\left(\frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l - a_{l+1})})\right)$  distributed. First, we condition on  $Z = z$ . Then the number of edges that  $i$  creates in  $[T - a_l, T - a_{l+2})$ , connect  $i$  to  $j_T$  and survive until  $T - a_{l+2}$  is described by a birth and death process with constant birth rate  $\alpha s_i / l$ , linear death rate with factor  $\beta$  and initial state  $z$ . For such a process, the probability generating function (6.4) was stated in the previous subsection. Since (6.4) is the product of the probability generating function of a Poisson distribution and of the probability generating function of a binomial distribution, we obtain the sum of a  $\text{Po}\left(\frac{\alpha s_i}{l\beta} (1 - e^{-\beta(a_{l+1} - a_{l+2})})\right)$  distributed and a  $\text{Bin}(z, e^{-\beta(a_{l+1} - a_{l+2})})$  distributed random variable for time  $T - a_{l+2}$ , where these two random variables are independent. The second one is a Poisson mixture of binomial distributions if we do not condition on  $Z = z$ . This mixture is the Poisson distribution with parameter  $e^{-\beta(a_{l+1} - a_{l+2})} \frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l - a_{l+1})})$ . Consequently, it follows that

$$\text{Po}\left(e^{-\beta(a_{l+1} - a_{l+2})} \frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l - a_{l+1})}) + \frac{\alpha s_i}{l\beta} (1 - e^{-\beta(a_{l+1} - a_{l+2})})\right).$$

is the desired distribution.



We obtain inductively by mixing and independence that

$$\text{Po}\left(\sum_{l=i \vee j_T}^{y_T-1} \left(\prod_{k=1}^{y_T-l-1} e^{-\beta(a_{l+k}-a_{l+k+1})}\right) \frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l-a_{l+1})})\right)$$

is the distribution of the number of edges that connect  $i$  to  $j_T$  surviving until  $T - a_{y_T}$ .

Repeating the same argument as above, it follows that

$$\text{Po}\left(\frac{\alpha s_i}{(y_T-1)\beta} (1 - e^{-\beta a_{y_T}}) + e^{-\beta a_{y_T}} \sum_{l=i \vee j_T}^{y_T-1} \left(\prod_{k=1}^{y_T-l-1} e^{-\beta(a_{l+k}-a_{l+k+1})}\right) \frac{\alpha s_i}{(l-1)\beta} (1 - e^{-\beta(a_l-a_{l+1})})\right)$$

is the distribution of the number of edges between  $i$  and  $j_T$  being alive at time  $T$ .

By independence of the edge processes of edges generated by the various nodes  $i$ , we obtain the following distribution for the total number of incoming edges in node  $j_T$ :

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

Thus without conditioning, we obtain a mixed Poisson distribution:

$$\text{MixPo}\left(\frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{Y_T} \frac{S_i}{Y_T-1} (1 - e^{-\beta A_{Y_T}(T)}) + \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{Y_T} \sum_{l=i \vee j_T}^{Y_T-1} \frac{S_i}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\right).$$

### 6.2.3 Degree distribution in the pure birth case

By conditional independence of the numbers of incoming and outgoing edges<sup>5</sup>, the previous considerations imply that the degree distribution of a randomly chosen node is a mixed Poisson distribution. The random parameter of this mixed Poisson distribution is the sum of the random parameters of the two mixed Poisson distributions that were deduced in the previous two subsections. Thus we obtain the  $\text{MixPo}(\Lambda_T)$  distribution, where

$$\begin{aligned} \Lambda_T &= \frac{\alpha S_{j_T}}{\beta} (1 - e^{-\beta A_{\max(j_T, 2)}(T)}) + \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{Y_T} \frac{S_i}{Y_T-1} (1 - e^{-\beta A_{Y_T}(T)}) \\ &+ \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{Y_T} \sum_{l=i \vee j_T}^{Y_T-1} \frac{S_i}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}). \end{aligned} \quad (6.5)$$

<sup>5</sup> Note that those numbers are *not* independent unconditionally. For example, an old node is expected to have a high number of incoming as well as outgoing edges, which leads to a positive correlation.

### 6.2.4 Degree distribution in the general case

If the nodes can die, the degree distribution can be determined similarly to the pure birth case.

Let  $0 = T_1^+ < \dots < T_{\mathfrak{B}_T}^+$  be the birth times. Recall that  $0 = T_1 < \dots < T_{\mathfrak{B}_T + \mathfrak{D}_T}$  are the event times up to  $T$ . Moreover, let  $T_i^-$  denote the death time of node  $i$  for each  $i$ , where the nodes are still numbered according to their birth times. Finally, given the subset of  $\{1, \dots, \mathfrak{B}_T\}$  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 events as before. Note that  $S_{J_T}$  is (stochastically) independent of  $((Y_t)_{0 \leq t \leq T}, J_T)$ .

The crucial differences compared to the procedure for the pure birth case are:

- We have to condition on  $Y_T > 0$  since there has to exist a node at time  $T$  that can be picked randomly.
- We have to consider  $[T_l, T_{l+1})$  instead of  $[T - A_l(T), T - A_{l+1}(T))$  as time intervals in which the parameters of the process of incoming edges are constant.
- We do not sum over all nodes  $i$ , but over all that are still alive at time  $T$ .
- In general, the population size does not increase almost surely by one per time interval. Thus the population size is not necessarily equal to  $l$  in the time interval  $[T_l, T_{l+1})$ .

Since deaths of other nodes reduce the number of outgoing edges, 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 proceed in a similar way as for the incoming nodes in the pure birth case. Therefore, we do not derive the distribution of the numbers of outgoing and incoming edges separately here.

We condition on  $(Y_t)_{0 \leq t \leq T} = (y_t)_{0 \leq t \leq T}$ , the social indices  $(S_i)_{i \in \mathbb{N}} = (s_i)_{i \in \mathbb{N}}$  and  $J_T = j_T$ . Let  $b_T$  and  $d_T$  denote the corresponding number of births and deaths up to time  $T$ , respectively. Furthermore, let  $0 = t_1 < t_2 < \dots < t_{b_T + d_T}$  and  $0 = t_1^+ < t_2^+ < \dots < t_{b_T}^+$  be the corresponding event times and birth times up to time  $T$ , respectively. We 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 at rate  $\alpha s_i$  and  $\alpha s_{j_T}$ , 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 that connect  $i$  to  $j_T$  and 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  $\beta$  like before. Analogously to Subsection 6.2.2, we obtain that the number of edges that are created in  $[t_l, t_{l+1})$ , connect  $i$  to  $j_T$  and survive until  $t_{l+1}$  has the following Poisson distribution:

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

Let the function  $r = r_{(y_t)_{0 \leq t \leq T}} : \{1, \dots, b_T\} \rightarrow \{1, \dots, b_T + d_T\}$  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. Note that the sum  $b_T + d_T$  of the number of births and the number of deaths up to time  $T$  is the number of events up to time  $T$ .

Applying the same iterative procedure as in Subsection 6.2.2, we can see that

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

is the distribution of the number of edges between  $i$  and  $j_T$  being alive at time  $T$ .

Let the death times be denoted by  $t_1^-, t_2^-, \dots, t_{d_T}^-$ . Then by independence of the processes of edges generated by the various nodes  $i$ , we obtain the following conditional degree distribution:

$$\begin{aligned} & \text{Po}\left(\sum_{\substack{i=1 \\ i \neq j_T}}^{b_T} \frac{\alpha(s_i + s_{j_T})}{(y_T - 1)\beta} \mathbb{1}_{\{t_i^- > T\}}(1 - e^{-\beta(T-t_{b_T+d_T})})\mathbb{1}_{\{y_T>1\}} + \sum_{\substack{i=1 \\ i \neq j_T}}^{b_T} \mathbb{1}_{\{t_i^- > T\}} e^{-\beta(T-t_{b_T+d_T})}\right. \\ & \quad \cdot \sum_{l=r(i)\vee r(j_T)}^{b_T+d_T-1} \left(\left(\prod_{k=1}^{b_T+d_T-l-1} e^{-\beta(t_{l+k+1}-t_{l+k})}\right) \frac{\alpha(s_i + s_{j_T})}{(y_{t_l} - 1)\beta}(1 - e^{-\beta(t_{l+1}-t_l)})\right)\mathbb{1}_{\{y_{t_l}>1\}}\Big) \\ & = \text{Po}\left(\frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{b_T} \frac{s_i + s_{j_T}}{(y_T - 1)} \mathbb{1}_{\{t_i^- > T\}}(1 - e^{-\beta(T-t_{b_T+d_T})})\mathbb{1}_{\{y_T>1\}}\right. \\ & \quad \left.+ \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{b_T} \mathbb{1}_{\{t_i^- > T\}} \sum_{l=r(i)\vee r(j_T)}^{b_T+d_T-1} \frac{s_i + s_{j_T}}{(y_{t_l} - 1)} (e^{-\beta(T-t_{l+1})} - e^{-\beta(T-t_l)})\mathbb{1}_{\{y_{t_l}>1\}}\right). \end{aligned}$$

Thus the degree distribution is  $\text{MixPo}(\Lambda_T^*)$ , where  $\mathcal{L}(\Lambda_T^*) = \mathcal{L}(\Lambda_T | Y_T > 0)$  and

$$\begin{aligned} \Lambda_T & := \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{\mathfrak{B}_T} \frac{S_i + S_{j_T}}{(Y_T - 1)} \mathbb{1}_{\{T_i^- > T\}} (1 - e^{-\beta(T-T_{\mathfrak{B}_T+d_T})}) \mathbb{1}_{\{Y_T>1\}} \\ & \quad + \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq j_T}}^{\mathfrak{B}_T} \mathbb{1}_{\{T_i^- > T\}} \sum_{l=r(i)\vee r(j_T)}^{\mathfrak{B}_T+d_T-1} \frac{S_i + S_{j_T}}{(Y_{T_l} - 1)} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \mathbb{1}_{\{Y_{T_l}>1\}}. \end{aligned} \quad (6.6)$$

## 6.3 Bounds on the total variation distance between the finite time and asymptotic degree distribution

Since in the pure birth case we obtain a much better bound with considerably less work, we treat the cases  $\mu = 0$  and the general case separately.

### 6.3.1 The pure birth case

Firstly, we consider the pure birth case, i.e.  $\mu = 0$ , again. Fix  $T > 0$ , and let the random variable  $\Lambda_T$  be defined by (6.5). Furthermore, let  $S_{J_\infty}$  and  $A_{J_\infty}$  be independent random variables such that  $S_{J_\infty}(\omega) = S_{J_T}(\omega)$  and  $A_{J_\infty}(\omega) := F_\infty^{-1}(F_T(A_{\max(J_T, 2)}(T, \omega)))$  for all  $\omega \in \Omega$ , where  $F_T$  and  $F_\infty$  are the cumulative distribution functions of  $A_{\max(J_T, 2)}(T)$ , i.e. of the  $\text{Exp}(\lambda)$  distribution truncated at  $T$  by Proposition 4.2.1, and the  $\text{Exp}(\lambda)$  distribution, respectively. Note that we know that

$\mathcal{L}(A_{J_T}) \rightarrow \text{Exp}(\lambda)$  weakly as  $T \rightarrow \infty$  from Chapter 4. The random variable  $A_{J_\infty}$  is  $\text{Exp}(\lambda)$  distributed since  $F_T$  is continuous and hence  $F_T(A_{\max(J_T, 2)}(T))$  is uniformly distributed on  $[0, 1]$ . Moreover, since the exponential distribution dominates the truncated exponential distribution, we have  $F_\infty(A_{\max(J_T, 2)}(T, \omega)) \leq F_T(A_{\max(J_T, 2)}(T, \omega))$  for all  $\omega \in \Omega$ . Since  $F_\infty^{-1}$  is increasing, it follows that  $A_{\max(J_T, 2)}(T, \omega) \leq A_{J_\infty}(\omega)$  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

$$M := \frac{\alpha S_{J_\infty}}{\beta} (1 - e^{-\beta A_{J_\infty}}) + \frac{\alpha \mathbb{E}(S)}{\beta} (1 - e^{-\beta A_{J_\infty}}). \quad (6.7)$$

We know from Subsection 6.2.3 that  $\text{MixPo}(\Lambda_T)$  is the degree distribution at time  $T$  in the pure birth case. Theorems 6.3.1 and 6.3.2 below imply that  $\text{MixPo}(\Lambda_T)$  converges at rate of just a bit slower than  $e^{-\frac{1}{2}\lambda T}$  to  $\text{MixPo}(M)$  as  $T \rightarrow \infty$ , which is the asymptotic degree distribution already stated in Section 3.2 of [BL10]. Note that these theorems are much more powerful as they give an exact distance bound for finite  $T$ .

### 6.3.1 Theorem

We have

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

where  $\sigma_S$  is the standard deviation of  $S$ .

The main idea of the proof is to make use of Theorem 3.4.1 to obtain  $\mathbb{E}(|\Lambda_T - M|)$  as an upper bound for the total variation distance on the left-hand side. In order to bound  $\mathbb{E}(|\Lambda_T - M|)$  further, we use that the expected value of the second summand of the right-hand side of (6.5) becomes small as  $T \rightarrow \infty$  since the age  $A_{Y_T}(T)$  of the youngest individual at time  $T$  converges quickly to zero, and compare the other summands of the right-hand sides of (6.5) and (6.7).

For the comparison of the last summands in (6.5) and (6.7), 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 zero and that  $A_{J_T}(T)$  converges quickly to  $A_{J_\infty}$  by Proposition 4.2.1.

Finally, the expected absolute value of the difference between the first summand of (6.5) and the first summand of (6.7) again becomes small since  $A_{J_T}(T)$  converges quickly to  $A_{J_\infty}$ .

We prove Theorem 6.3.1 in detail after the next theorem, which gives us the exponential rate for the convergence of the degree distribution. In order to reveal this rate, we further estimate  $\mathbb{E}(\frac{1}{Y_T})$  and  $\mathbb{E}(\frac{1}{\sqrt{Y_T}})$  on the right-hand side in Theorem 6.3.1 by using Proposition 2.2.4 and Corollary 2.2.5.

### 6.3.2 Theorem

For  $T \geq \log(2)/\lambda$ , we have

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

where  $\sigma_S$  is the standard deviation of  $S$ .

*Proof:* The inequality follows directly from Theorem 6.3.1, Proposition 2.2.4 and Corollary 2.2.5.  $\square$

### 6.3.3 Remark

For the original model of Britton and Lindholm, which allows loops, we can adapt the proof of Theorem 6.3.1 in such a way that the upper bounds in Theorem 6.3.1 and Theorem 6.3.2 remain exactly the same.

#### Proof of Theorem 6.3.1

From Theorem 3.4.1 follows

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

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

$$\begin{aligned} & \mathbb{E} \left| \frac{\alpha}{\beta} \left( S_{J_\infty} (1 - e^{-\beta A_{J_\infty}}) + \mathbb{E}(S) (1 - e^{-\beta A_{J_\infty}}) - S_{J_T} (1 - e^{-\beta A_{\max(J_T, 2)}(T)}) \right. \right. \\ & \quad \left. \left. - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{S_i}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{S_i}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right) \right| \\ & \leq \frac{\alpha}{\beta} \mathbb{E} \left| \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{S_i}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \right| \\ & \quad + \frac{\alpha}{\beta} \mathbb{E} \left| \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i}^{Y_T-1} \frac{(S_i - \mathbb{E}(S))}{l-1} \mathbb{1}_{\{J_T \leq l\}} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \\ & \quad + \frac{\alpha}{\beta} \mathbb{E} \left| \mathbb{E}(S) (1 - e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{\mathbb{E}(S)}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \\ & \quad + \frac{\alpha}{\beta} \mathbb{E} \left| S_{J_\infty} (1 - e^{-\beta A_{J_\infty}}) - S_{J_T} (1 - e^{-\beta A_{\max(J_T, 2)}(T)}) \right|, \end{aligned} \tag{6.8}$$

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

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

$$\begin{aligned} \mathbb{E} \left| \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{S_i}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \right| &= \mathbb{E} \left( \mathbb{E} \left( \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{S_i}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \middle| Y_T, J_T \right) \right) \\ &= \mathbb{E} \left( \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{\mathbb{E}(S_i)}{Y_T - 1} (1 - \mathbb{E}(e^{-\beta A_{Y_T}(T)} | Y_T)) \right), \end{aligned} \tag{6.9}$$

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  $y_T - 1$  independent and identically distributed truncated exponentially distributed random variables by Proposition 4.2.1. Since the minimum of  $y_T - 1$  independent  $\text{Exp}(\lambda)$  distributed random variables is  $\text{Exp}((y_T - 1)\lambda)$  distributed, the distribution

of  $A_{y_T}(T)$  is stochastically dominated by the  $\text{Exp}((y_T - 1)\lambda)$  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  $\mathbb{E}(e^{-\beta Z}) = \frac{(y_T-1)\lambda}{(y_T-1)\lambda-\beta}$  for  $Z \sim \text{Exp}((y_T - 1)\lambda)$  implies that the right-hand side of (6.9) is smaller than or equal to

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

Note that

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

where  $p_n(T)$  is the probability mass function from Section 2.2. Thus the right-hand side of (6.9) is smaller than or equal to

$$\frac{\beta}{\lambda}\mathbb{E}(S)\mathbb{E}\left(\frac{1}{Y_T}\right). \quad (6.10)$$

For the second line of the right-hand side of (6.8), we obtain

$$\begin{aligned} &\mathbb{E}\left|\sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i}^{Y_T-1} \frac{(S_i - \mathbb{E}(S))}{l-1} \mathbb{1}_{\{J_T \leq l\}} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\right| \\ &= \mathbb{E}\left|\sum_{l=2}^{Y_T-1} \frac{1}{l-1} \sum_{\substack{i=1 \\ i \neq J_T}}^l (S_i - \mathbb{E}(S)) \mathbb{1}_{\{J_T \leq l\}} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\right| \\ &= \mathbb{E}\left(\mathbb{E}\left(\left|\sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{\substack{i=1 \\ i \neq J_T}}^l (S_i - \mathbb{E}(S)) (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)})\right|\right) \middle| Y_T, J_T\right) \\ &\leq \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}\left|\frac{1}{l-1} \sum_{i=2}^l (S_i - \mathbb{E}(S))\right| \mathbb{E}\left(e^{-\beta A_{l+1}(T)} (1 - e^{-\beta(A_l(T) - A_{l+1}(T))}) \middle| Y_T\right)\right) \\ &\leq \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sqrt{\sum_{i=2}^l \mathbb{E}\left((S_i - \mathbb{E}(S))^2\right)} \mathbb{E}(1 - e^{-\beta(A_l(T) - A_{l+1}(T))} \middle| Y_T)\right) \\ &\leq \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{\sqrt{l-1}} \sigma_S \mathbb{E}(1 - e^{-\beta(A_l(T) - A_{l+1}(T))} \middle| Y_T)\right), \end{aligned} \quad (6.11)$$

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  $\mathbb{E}|Z - \mathbb{E}(Z)| \leq \sqrt{\text{Var}(Z)}$  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  $l-1$  independent and identically distributed truncated exponentially distributed random variables by Proposition 4.2.1,

Corollary 4.2.3 and Remark 4.2.4 (for  $l \geq 2$ ). Thus it is stochastically dominated by an  $\text{Exp}((l-1)\lambda)$  distributed random variable  $Z_l$ , where  $Z_l$  can be assumed to be 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 (6.9), it follows that (6.11) is smaller than or equal to

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

Since we arranged  $A_{\max(J_T, 2)}(T) \leq A_{J_\infty}$ , for the third line of the right-hand side of (6.8) we obtain

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

$$\begin{aligned}
&= \mathbb{E}\left(\mathbb{E}(S)(1 - e^{-\beta A_{Y_T}(T)}) + \mathbb{E}(S)(e^{-\beta A_{\max(J_T,2)}(T)} - e^{-\beta A_{J_\infty}})\right) \\
&= \mathbb{E}(S)\mathbb{E}\left(1 - \mathbb{E}(e^{-\beta A_{Y_T}(T)}|Y_T)\right) + \mathbb{E}(S)(\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) - \mathbb{E}(e^{-\beta A_{J_\infty}})). \quad (6.12)
\end{aligned}$$

By Proposition 4.2.1

$$\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) = \int_0^T e^{-\beta x} \frac{\lambda e^{-\lambda x}}{1 - e^{-\lambda T}} dx = \frac{\lambda}{1 - e^{-\lambda T}} \int_0^T e^{-(\beta+\lambda)x} dx = \frac{\lambda}{\beta + \lambda} \frac{1 - e^{-(\beta+\lambda)T}}{1 - e^{-\lambda T}}, \quad (6.13)$$

whence follows that the right-hand side of (6.12) is smaller than or equal to

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

Since we arranged  $S_{J_T} = S_{J_\infty}$  and  $A_{\max(J_T,2)}(T) \leq A_{J_\infty}$ , for the fourth line of the right-hand side of (6.8) we have

$$\mathbb{E}\left|S_{J_\infty}(1 - e^{-\beta A_{J_\infty}}) - S_{J_T}(1 - e^{-\beta A_{\max(J_T,2)}(T)})\right| = \mathbb{E}(S)(\mathbb{E}(e^{-\beta A_{\max(J_T,2)}(T)}) - \mathbb{E}(e^{-\beta A_{J_\infty}})). \quad (6.14)$$

Note that the right-hand side of (6.14) is the same as the second summand of the right-hand side of (6.12), which is smaller than or equal to  $\mathbb{E}(S)\frac{\lambda}{\beta+\mu} \frac{1}{e^{\lambda T} - 1}$  (see above).

Altogether, we obtain the following upper bound for the right-hand side of (6.8):

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

□

### 6.3.2 The general case

#### The main theorem

Let  $S_{J_\infty}(\omega) = S_{J_T}(\omega)$  as in the pure birth case. Furthermore, let, for the time being, the individuals numbered as in the proof of Theorem 4.3.1, and let

$$U(\omega) = \mathbb{1}_{\{J_T(\omega) < Y_T(\omega)\}} F_*^*(A_{J_T}(T, \omega)) + \mathbb{1}_{\{J_T(\omega) = Y_T(\omega)\}} F^*(A_{J_T}(T, \omega)),$$

where  $F_*(t) = 1 - \frac{e^{-\lambda t} - e^{-(\lambda-\mu)T} e^{-\mu t}}{1 - e^{-(\lambda-\mu)T}}$  and  $F^*(t) = \frac{\lambda(1 - e^{-\mu t}) - \mu(1 - e^{-\lambda t})}{\lambda - \mu} \mathbb{1}_{[0, T)}(t) + \mathbb{1}_{\{T\}}(t)$ . Then  $A_{J_T} = A_*$ , where  $A_*$  is the random variable from Corollary 4.3.3, i.e.  $A_* = \mathbb{1}_{\{J_T < Y_T\}} F_*^{-1}(U) + \mathbb{1}_{\{J_T = Y_T\}} (F^*)^{-1}(U)$  (cf. the proof of Corollary 4.3.3). Moreover, let  $A_{J_\infty} = Z_*$ , where  $Z_*$  is the random variable from Corollary 4.3.3, i.e.  $A_{J_\infty}(\omega) = F^{-\infty}(U(\omega))$ . Note, given  $Y_T > 0$ , we have  $A_{J_\infty} \sim \text{Exp}(\lambda)$ .

Recall the definition of the parameter random variable  $\Lambda_T$  from (6.6), and let  $\Lambda_T^*$  be a random variable with  $\mathcal{L}(\Lambda_T^*) = \mathcal{L}(\Lambda_T | Y_T > 0)$  as before. Moreover, set

$$M := \frac{\alpha}{\beta + \mu} (S_{J_\infty} + \mathbb{E}(S))(1 - e^{-(\beta+\mu)A_{J_\infty}}), \quad (6.15)$$

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



We already know that  $\text{MixPo}(\Lambda_T^*)$  is the degree distribution at time  $T$  in the general case. Theorem 6.3.4 below implies that  $\text{MixPo}(\Lambda_T^*)$  converges at a rate of just a bit over  $e^{-\frac{1}{6}(\lambda-\mu)T}$  to the  $\text{MixPo}(\mathbf{M}^*)$  distribution, which is the asymptotic degree distribution stated in Section 3.2 of [BL10], as  $T \rightarrow \infty$ . Note again that this theorem is much more powerful since it gives an exact bound for the total variation distance for finite  $T$ .

### 6.3.4 Theorem

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

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

Note that the right-hand side is of order

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

as  $T \rightarrow \infty$ .

The main idea of the proof of this theorem is the same as in the pure birth case: We make use of Theorem 3.4.1 to obtain  $\mathbb{E}(|\Lambda_T - \mathbf{M}| | Y_T > 0)$  as an upper bound for  $d_{TV}(\text{MixPo}(\Lambda_T^*), \text{MixPo}(\mathbf{M}^*))$  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 (6.6) converges quickly to zero since the time  $T - T_{\mathfrak{B}_T + \mathfrak{D}_T}$  since the last event before  $T$  converges quickly to zero, and compare the remaining summand of the right-hand side of (6.6) with the right-hand side of (6.15). 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 6.3.19).

A further important ingredient is that the reciprocal of the node process  $(Y_t)_{t \geq 0}$  conditioned on survival is a supermartingale (see Corollary 6.3.9), 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  $A_{J_\infty}$  by Corollary 4.3.3.

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, we introduce a deterministic number  $\kappa(T)$  that 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 Definition 6.3.11 and Lemma 6.3.12(ii) below). Essentially, we substitute the number of events  $\mathfrak{B}_T + \mathfrak{D}_T$  up to time  $T$  by  $\kappa(T)$  in the proof of Lemma 6.3.21 below since it is difficult to treat the dependencies between the number of events up to time  $T$  and the event times.

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 at time  $T$  was born up to time  $T/2$  decreases exponentially in  $T$  (see Lemma 6.3.12(i) below). Thus we essentially only have to consider the time interval  $[T/2, \infty)$  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  $T/2$  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.

### Some lemmas of general interest

Before we prove Theorem 6.3.4 in detail, we formulate several important results that are used in the proof and could also be useful in other situations. Together with the core results from Chapter 2 and the more specialized results in the next paragraph, the reader obtains a comprehensive body of knowledge on linear birth and death processes.

Our first result gives an expression for the extinction probability given that the process survives up to time  $T$ . In order to simplify the notation, we define  $Y_\infty := \lim_{t \rightarrow \infty} Y_t$ . Note that  $Y_\infty \in \{0, \infty\}$  almost surely.

#### 6.3.5 Lemma

For the conditioned extinction probability given  $Y_T > 0$ , we have

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

*Proof:* By conditioning on the population size, we obtain

$$\mathbb{P}(Y_\infty = 0 | Y_T > 0) = \mathbb{E}(\mathbb{P}(Y_\infty = 0 | Y_T) | Y_T > 0) = \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T} \mid Y_T > 0\right) \quad (6.16)$$

since  $(\mu/\lambda)^m$  is the extinction probability of a linear birth and death process with birth rate  $\lambda$ , death rate  $\mu$  and initial value  $m$  (see Remark 2.2.2).

On the one hand, we have

$$\begin{aligned} \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T}\right) &= \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T} \mid Y_T > 0\right)\mathbb{P}(Y_T > 0) + \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T} \mid Y_T = 0\right)\mathbb{P}(Y_T = 0) \\ &= \mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T} \mid Y_T > 0\right)(1 - p_0(T)) + p_0(T). \end{aligned}$$

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

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

This yields

$$\mathbb{E}\left(\left(\frac{\mu}{\lambda}\right)^{Y_T} \mid Y_T > 0\right) = \left(\frac{\mu}{\lambda} - p_0(T)\right)(1 - p_0(T))^{-1} = \frac{\mu(\lambda - \mu)}{\lambda(\lambda e^{(\lambda - \mu)T} - \mu)} \frac{\lambda e^{(\lambda - \mu)T} - \mu}{(\lambda - \mu)e^{(\lambda - \mu)T}} \leq \frac{\mu}{\lambda} e^{-(\lambda - \mu)T}.$$

□

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

### 6.3.6 Lemma

We have for  $T \geq \frac{1}{\lambda - \mu} \log(2)$

$$\mathbb{P}(Y_T = 1 | Y_T > 0) \leq \frac{2(\lambda - \mu)}{\lambda} e^{-(\lambda - \mu)T}.$$

*Proof:* Using the probability mass functions  $p_n$  and the function  $\tilde{p}$  from Section 2.2, for  $T \geq \frac{1}{\lambda - \mu} \log(2)$ , we obtain

$$\mathbb{P}(Y_T = 1 | Y_T > 0) = \frac{p_1(T)}{1 - p_0(T)} = 1 - \lambda \tilde{p}(T) = \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \mu} \leq \frac{2(\lambda - \mu)}{\lambda e^{(\lambda - \mu)T}}.$$

□

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

### 6.3.7 Lemma

- (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*  $\mu$ , *death rate*  $\lambda$  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]), which can be proved by using the embedded generation process (see e.g. Section IV.3 of [AN72] or Subsection 6(a) of [Wau58]) and the corresponding result for discrete-time branching processes (see e.g. Theorem 3 in Section I.12 of [AN72]) or the general result from [Wau58] for conditional Markov processes (cf. Subsection 6(a) of [Wau58]). We know from Corollary 2.2.8 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, we know from Corollary 2.2.8 that  $(Y_t)_{t \geq 0}$  is a submartingale. Thus we obtain for  $t > s \geq 0$  and  $y_s \in \mathbb{N}$

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

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. □

Since a more explicit proof of the analogous results for the embedded jump chain is easier, we also state and prove these here.

### 6.3.8 Lemma

- (i)  $(Y_{T_k})_{k \in \mathbb{N}}$  conditioned on  $Y_\infty = 0$  is a supermartingale.
- (ii)  $(Y_{T_k})_{k \in \mathbb{N}}$  conditioned on  $Y_\infty > 0$  is a submartingale.

*Proof:*

- (i) Consider a linear birth and death process  $(\tilde{Y}_t)_{t \geq 0}$  with *birth rate*  $\mu$ , *death rate*  $\lambda$  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 the proof of Lemma 6.3.7). Thus it remains to show that  $(\tilde{Y}_{\tilde{T}_k})_{k \in \mathbb{N}}$  is a supermartingale in order to show (i), where  $(\tilde{T}_k)_{k \in \mathbb{N}}$  are the event times of  $(\tilde{Y}_t)_{t \geq 0}$ .

Since  $(\tilde{Y}_t)_{t \geq 0}$  is a birth and death process, we have

$$\begin{aligned} \mathbb{E}(\tilde{Y}_{\tilde{T}_{k+1}} | \tilde{Y}_{\tilde{T}_1}, \dots, \tilde{Y}_{\tilde{T}_k}) &= \tilde{Y}_{\tilde{T}_k} + \mathbb{P}(\tilde{Y}_{\tilde{T}_{k+1}} = \tilde{Y}_{\tilde{T}_k} + 1 | \tilde{Y}_{\tilde{T}_1}, \dots, \tilde{Y}_{\tilde{T}_k}) \\ &\quad - \mathbb{P}(\tilde{Y}_{\tilde{T}_{k+1}} = \tilde{Y}_{\tilde{T}_k} - 1 | \tilde{Y}_{\tilde{T}_1}, \dots, \tilde{Y}_{\tilde{T}_k}) \\ &= \tilde{Y}_{\tilde{T}_k} + \frac{\mu}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} \\ &\leq \tilde{Y}_{\tilde{T}_k}. \end{aligned}$$

Thus  $(Y_{T_k})_{k \in \mathbb{N}}$  conditioned on  $Y_\infty = 0$  is a supermartingale.

- (ii) Consider a process  $(\hat{Y}_t)_{t \geq 0}$  that has the law of  $(Y_t)_{t \geq 0}$  conditioned on  $Y_\infty > 0$  and let the event times of  $(\hat{Y}_t)_{t \geq 0}$  be denoted by  $(\hat{T}_k)_{k \in \mathbb{N}}$ . Then we have

$$\begin{aligned} \mathbb{E}(\hat{Y}_{\hat{T}_{k+1}} | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}) &= \hat{Y}_{\hat{T}_k} + \mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} + 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}) \\ &\quad - \mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} - 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}). \end{aligned}$$

Consequently, it remains to show that

$$\mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} + 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}) - \mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} - 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}) \geq 0.$$

In order to do so, we compute

$$\begin{aligned} \mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} + 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}) &= \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty > 0) \\ &= \frac{\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1, Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \\ &= \frac{\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}) - \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1, Y_\infty = 0 | Y_{T_1}, \dots, Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \\ &= \mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})^{-1} \left( \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}) \right. \\ &\quad \left. - \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty = 0) \mathbb{P}(Y_\infty = 0 | Y_{T_k}) \right). \quad (6.17) \end{aligned}$$

Since

$$\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty = 0) = \mathbb{P}(\tilde{Y}_{\hat{T}_{k+1}} = \tilde{Y}_{\hat{T}_k} + 1, | \tilde{Y}_{\hat{T}_1}, \dots, \tilde{Y}_{\hat{T}_k}) = \frac{\mu}{\lambda + \mu}$$

and

$$\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} + 1 | Y_{T_1}, \dots, Y_{T_k}) = \frac{\lambda}{\lambda + \mu},$$

the right-hand side of (6.17) is equal to

$$\frac{\frac{\lambda}{\lambda + \mu} - \frac{\mu}{\lambda + \mu} \mathbb{P}(Y_\infty = 0 | Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \geq \frac{\frac{\mu}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} \mathbb{P}(Y_\infty = 0 | Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})}. \quad (6.18)$$

Since

$$\frac{\lambda}{\lambda + \mu} = \mathbb{P}(\tilde{Y}_{\hat{T}_{k+1}} = \tilde{Y}_{\hat{T}_k} - 1 | \tilde{Y}_{\hat{T}_1}, \dots, \tilde{Y}_{\hat{T}_k}) = \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty = 0)$$

and

$$\frac{\mu}{\lambda + \mu} = \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}),$$

the right-hand side of (6.18) is equal to

$$\begin{aligned} & \frac{\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}) - \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty = 0) \mathbb{P}(Y_\infty = 0 | Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \\ &= \frac{\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}) - \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1, Y_\infty = 0 | Y_{T_1}, \dots, Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \\ &= \frac{\mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1, Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})}{\mathbb{P}(Y_\infty > 0 | Y_{T_1}, \dots, Y_{T_k})} \\ &= \mathbb{P}(Y_{T_{k+1}} = Y_{T_k} - 1 | Y_{T_1}, \dots, Y_{T_k}, Y_\infty > 0) \\ &= \mathbb{P}(\hat{Y}_{\hat{T}_{k+1}} = \hat{Y}_{\hat{T}_k} - 1 | \hat{Y}_{\hat{T}_1}, \dots, \hat{Y}_{\hat{T}_k}). \end{aligned}$$

Thus  $(Y_{T_k})_{k \in \mathbb{N}}$  conditioned on ultimate survival is a submartingale.  $\square$

Lemma 6.3.7 yields a useful result about  $(Y_t^{-1})_{t \geq 0}$  conditioned on ultimate survival:

### 6.3.9 Corollary

$(Y_t^{-1})_{t \geq 0}$  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  $t > s \geq 0$

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

Thus  $(Z_t^{-1})_{t \geq 0}$  is a supermartingale. Consequently, Lemma 6.3.7 implies that  $(Y_t^{-1})_{t \geq 0}$  conditioned on ultimate survival is a supermartingale.  $\square$

### 6.3.10 Remark

It can be proved analogously to Corollary 6.3.9 that  $(Y_{T_k}^{-1})_{k \in \mathbb{N}}$  conditioned on  $Y_\infty > 0$  is a supermartingale.

Now we introduce the deterministic number  $\kappa(T)$  as mentioned before, which helps us to cope with dependencies. Furthermore, we define the random number  $\mathcal{K}(T)$  as index of the last event time before  $T/2$  such that  $r(J_T) > \mathcal{K}(T)$  is equivalent to  $T_{r(J_T)} \geq T/2$  and  $\mathcal{M}_T$  as the sum of the number of births and the number of deaths to simplify notation.

### 6.3.11 Definition

Let  $\kappa(T) := \lfloor e^{\frac{3}{2}(\lambda+\mu)T} \rfloor$  and  $\mathcal{K}(T) := \max\{k : T_k < T/2\}$ , such that  $Y_{T_{\mathcal{K}(T)}} = Y_{T/2}$  almost surely. Moreover, let  $\mathcal{M}_T := \mathfrak{B}_T + \mathfrak{D}_T$  be the *number of events* up to time  $T$ .

The following lemma implies the properties of  $\kappa(T)$  and  $\mathcal{K}(T)$  that were stated before and that we need to prove the convergence rate for the degree distribution in the general case.

### 6.3.12 Lemma

- (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

$$\begin{aligned} \mathbb{P}(r(J_T) \leq \mathcal{K}(T) | Y_T > 0) &= \mathbb{P}\left(T_{r(J_T)} < \frac{T}{2} \mid Y_T > 0\right) \\ &\leq e^{-\frac{1}{2}\lambda T} + \frac{\lambda - \mu}{\lambda e^{(\lambda-\mu)T} - \lambda} \left( \log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T \right). \end{aligned}$$

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

$$\mathbb{P}(\kappa(T) \leq \mathcal{M}_T | Y_T > 0) \leq \frac{60\lambda^3(\lambda + \mu)}{(\lambda - \mu)^4} e^{-(\lambda + \mu)T}$$

*Proof:*

- (i) By Remark 4.3.9, we have

$$\begin{aligned} \mathbb{P}(r(J_T) \leq \mathcal{K}(T) | Y_T > 0) &= \mathbb{P}(T_{r(J_T)} \leq T_{\mathcal{K}(T)} | Y_T > 0) = \mathbb{P}\left(T - T_{r(J_T)} \geq \frac{T}{2} \mid Y_T > 0\right) \\ &\leq \mathbb{E}\left(\frac{Y_T - 1}{Y_T} e^{-\frac{1}{2}\lambda T} + \frac{1}{Y_T} \mid Y_T > 0\right) \leq e^{-\frac{1}{2}\lambda T} + \mathbb{E}\left(\frac{1}{Y_T} \mid Y_T > 0\right). \end{aligned} \tag{6.19}$$

By Proposition 2.2.4, the right-hand side of (6.19) is smaller than or equal to

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

- (ii) Using Proposition 2.3.1, we compute

$$\begin{aligned} \kappa(T) - 2\mathbb{E}(\mathfrak{B}_T) &\geq e^{\frac{3}{2}(\lambda+\mu)T} - 1 - \frac{2\lambda}{\lambda - \mu} e^{(\lambda-\mu)T} + \frac{2\mu}{\lambda - \mu} \\ &= e^{(\lambda-\mu)T} \left( e^{2\mu T} e^{\frac{1}{2}(\lambda+\mu)T} - \frac{2\lambda}{\lambda - \mu} \right) + \frac{3\mu - \lambda}{\lambda - \mu}. \end{aligned} \tag{6.20}$$

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

$$\frac{1}{2}e^{2\mu T}e^{\frac{1}{2}(\lambda + \mu)T} \geq \frac{2\lambda}{\lambda - \mu}. \quad (6.21)$$

Thus for  $T \geq \frac{2(\log(4\lambda) - \log(\lambda - \mu))}{\lambda - \mu}$ , the right-hand side of (6.20) is larger than or equal to

$$e^{(\lambda - \mu)T} \frac{1}{2}e^{2\mu T}e^{\frac{1}{2}(\lambda + \mu)T} + \frac{3\mu - \lambda}{\lambda - \mu}. \quad (6.22)$$

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

$$e^{(\lambda - \mu)T} \geq 1 + (\lambda - \mu)T \geq 1 + 2\log\left(\frac{4\lambda}{\lambda - \mu}\right) \geq 1 + 2\log(4) \geq 2, \quad (6.23)$$

Inequality (6.21) implies that for  $T \geq \frac{2(\log(4\lambda) - \log(\lambda - \mu))}{\lambda - \mu}$ , the expression (6.22) 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

$$\kappa(T) - 2\mathbb{E}(\mathfrak{B}_T) \geq \frac{3}{8}e^{\frac{3}{2}(\lambda + \mu)T} > 0. \quad (6.24)$$

Consequently, we can apply Chebyshev's inequality and obtain

$$\begin{aligned} \mathbb{P}(\kappa(T) \leq \mathcal{M}_T | Y_T > 0) &\leq \frac{\mathbb{P}(\kappa(T) \leq \mathcal{M}_T)}{\mathbb{P}(Y_T > 0)} \leq \frac{\mathbb{P}(\kappa(T) \leq 2\mathfrak{B}_T)}{\mathbb{P}(Y_T > 0)} \\ &= \frac{1}{\mathbb{P}(Y_T > 0)} \mathbb{P}(\kappa(T) - 2\mathbb{E}(\mathfrak{B}_T) \leq 2\mathfrak{B}_T - 2\mathbb{E}(\mathfrak{B}_T)) \\ &\leq \frac{1}{\mathbb{P}(Y_T > 0)} \frac{4\text{Var}(\mathfrak{B}_T)}{(\kappa(T) - 2\mathbb{E}(\mathfrak{B}_T))^2}, \end{aligned} \quad (6.25)$$

where  $T \geq \frac{2(\log(4\lambda) - \log(\lambda - \mu))}{\lambda - \mu}$ . By  $\mathbb{P}(Y_T > 0) \geq \mathbb{P}(Y_\infty > 0) = \frac{\lambda - \mu}{\lambda}$  and Inequality (6.24), for  $T \geq \frac{2(\log(4\lambda) - \log(\lambda - \mu))}{\lambda - \mu}$ , the right-hand side of (6.25) is smaller than or equal to

$$\frac{256}{9} \frac{\lambda}{\lambda - \mu} \frac{\text{Var}(\mathfrak{B}_T)}{e^{3(\lambda + \mu)T}} \leq \frac{30\lambda}{\lambda - \mu} \left( \frac{\lambda^2(\lambda + \mu)}{(\lambda - \mu)^3} e^{-(\lambda + \mu)T} + \frac{2\lambda^2\mu}{(\lambda - \mu)^3} e^{-2(\lambda + \mu)T} \right), \quad (6.26)$$

where we used Proposition 2.3.1 in order to obtain the upper bound on the right-hand side. Since  $e^{-(\lambda + \mu)T} \leq 1/2$  by (6.23), the right-hand side of (6.26) 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 6.3.4.

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

#### 6.3.13 Lemma

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

$$\mathbb{E}(1 - e^{-c(T - T_{\mathcal{M}_T})} | Y_T > 0) \leq \frac{2(\lambda - \mu)}{\lambda e^{(\lambda - \mu)T}} + \frac{c}{\lambda} \mathbb{E}\left(\frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \mid Y_T > 0\right). \quad (6.27)$$

*Proof:* Let  $X$  have the cumulative distribution function

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

Conditionally on  $Y_T$ , we then have  $T - T_{\mathcal{M}_T} \leq_{st} X$  by Theorem 4.4.1, which implies

$$\begin{aligned} \mathbb{E}(1 - e^{-c(T - T_{\mathcal{M}_T})} | Y_T > 0) &\leq \mathbb{E}(1 - e^{-cX} | Y_T > 0) \\ &= \mathbb{E}\left((1 - e^{-cT})\mathbb{P}(X = T | Y_T) + \mathbb{E}(1 - e^{-cX} | X < T, Y_T)\mathbb{P}(X < T | Y_T) \mid Y_T > 0\right). \end{aligned} \quad (6.28)$$

Since  $\mathbb{P}(X = T | Y_T = 1) = 1$  and  $\mathcal{L}(X | Y_T) = \text{Exp}((Y_T - 1)\lambda)$  on  $\{Y_T \geq 2\}$ , we obtain that the right-hand side of (6.28) is smaller than or equal to

$$\begin{aligned} \mathbb{E}\left(\mathbb{1}_{\{Y_T = 1\}} + \frac{c}{c + (Y_T - 1)\lambda} \mathbb{1}_{\{Y_T > 1\}} \mid Y_T > 0\right) &\leq \mathbb{P}(Y_T = 1 | Y_T > 0) \\ &\quad + \frac{c}{\lambda} \mathbb{E}\left(\frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \mid Y_T > 0\right). \end{aligned}$$

For the conditional probability of  $Y_T = 1$ , by Lemma 6.3.6, we have for  $T \geq \frac{1}{\lambda - \mu} \log(2)$

$$\mathbb{P}(Y_T = 1 | Y_T > 0) \leq \frac{2(\lambda - \mu)}{\lambda e^{(\lambda - \mu)T}}.$$

□

The following lemma gives us an upper bound for the conditional expectation on the right-hand side of (6.27) and is proved similarly to Proposition 2.2.4.

#### 6.3.14 Lemma

We have

$$\mathbb{E}\left(\frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \mid Y_T > 0\right) \leq \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \mu} \left(\log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right).$$

*Proof:* We have

$$\mathbb{E}\left(\frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \mid Y_T > 0\right) = \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \mu} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{\lambda e^{(\lambda - \mu)T} - \lambda}{\lambda e^{(\lambda - \mu)T} - \mu}\right)^n, \quad (6.29)$$

and in the proof of Proposition 2.2.4, we computed an upper bound for the series in (6.29) that yields that the right-hand side of (6.29) is smaller than or equal to

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

□

In order to prove our next lemma, we use the supermartingale from Corollary 6.3.9.



**6.3.15 Lemma**

For all  $\delta, \gamma > 0$ , we have

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

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

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

where the last line follows from Corollary 6.3.9 and the submartingale inequality (2) in Theorem 6.14 on page 99 in [Yeh95] applied to  $(-Y_t^{-1})_{t \geq 0}$  (i.e. we use the continuous-time analogue to the supermartingale inequality (1) in Corollary 6.8 on page 94 in [Yeh95]).  $\square$

We use the following lemma to bound the conditional expectation in the second summand of the right-hand side of Lemma 6.3.15 from above.

**6.3.16 Lemma**

We have for  $T \geq \frac{2}{\lambda-\mu} \log(2)$

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

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

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

where the second line follows from Proposition 2.2.4 and the last inequality holds for  $T \geq \frac{2}{\lambda-\mu} \log(2)$ .  $\square$

Before we continue to state and prove further useful results, we introduce another random variable.

**6.3.17 Definition**

Let  $R_{T_l, T}$  be 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  $R_{T_l, T}$ .

### 6.3.18 Lemma

We have

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

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

$$\begin{aligned}\mathbb{E}(R_{T_l, T} | Y_{T_l}, T_l, T_{l+1}) &= p^+ \mathbb{E}(R_{T_l, T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &\quad + p^- \mathbb{E}(R_{T_l, T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1),\end{aligned}\tag{6.30}$$

where  $p^+ := \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} + 1 | Y_{T_l}, T_l, T_{l+1})$  and  $p^- := \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} - 1 | Y_{T_l}, T_l, T_{l+1})$ .<sup>6</sup>

With

$$\begin{aligned}\mathbb{E}(R_{T_l, T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) &= \mathbb{E}(R_{T_{l+1}, T} - \mathbb{1}_{\{T_{r-1}(l+1)} > T\}} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &= (Y_{T_l} + 1)e^{-\mu(T - T_{l+1})} - e^{-\mu(T - T_{l+1})} = Y_{T_l}e^{-\mu(T - T_{l+1})}\end{aligned}\tag{6.31}$$

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 (6.30) 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  $R_{T_l, T}$ :

$$\begin{aligned}\mathbb{E}(R_{T_l, T}^2 | Y_{T_l}, T_l, T_{l+1}) &= p^+ \mathbb{E}(R_{T_l, T}^2 | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &\quad + p^- \mathbb{E}(R_{T_l, T}^2 | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} - 1).\end{aligned}\tag{6.32}$$

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

$$\begin{aligned}\mathbb{E}(R_{T_l, T}^2 | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) &= \mathbb{E}((R_{T_{l+1}, T} - \mathbb{1}_{\{T_{r-1}(l+1)} > T\}})^2 | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &= \mathbb{E}\left(R_{T_{l+1}, T}^2 - 2R_{T_{l+1}, T}\mathbb{1}_{\{T_{r-1}(l+1)} > T\}} + \mathbb{1}_{\{T_{r-1}(l+1)} > T\}} \mid Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1, Y_{T_l} > 0\right)\end{aligned}\tag{6.33}$$

and further

$$\begin{aligned}\mathbb{E}(R_{T_{l+1}, T}^2 | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) &= \text{Var}(R_{T_{l+1}, T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &\quad + (\mathbb{E}(R_{T_{l+1}, T} | Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1))^2.\end{aligned}\tag{6.34}$$

<sup>6</sup> We have  $p^+ = \frac{\lambda}{\lambda + \mu}$  and  $p^- = \frac{\mu}{\lambda + \mu}$ . However, we do not need these explicit expressions for our purposes.

Given  $Y_{T_{l+1}}$ , let  $\mathfrak{L}_{l+1}$  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

$$\begin{aligned} & \text{Var}(R_{T_{l+1},T}|Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1) \\ &= \text{Var}\left(\sum_{j \in \mathfrak{L}_{l+1}} \mathbb{1}_{\{T_j^- > T\}} \mid Y_{T_l}, T_l, T_{l+1}, Y_{T_{l+1}} = Y_{T_l} + 1, Y_{T_l} > 0\right) \\ &= (Y_{T_l} + 1)(e^{-\mu(T-T_{l+1})} - e^{-2\mu(T-T_{l+1})}). \end{aligned}$$

For the second summand of (6.34), 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 (6.34) is equal to

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

For the remaining parts of (6.33), we have

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

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 (6.33) implies

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

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

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

Thus from (6.32) follows

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

□

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 below:

**6.3.19 Lemma**

For  $l \in \mathbb{N}$ , we have

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

*Proof:* By Jensen's inequality, we obtain

$$\begin{aligned} & \mathbb{E}\left(\left|\frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})}\right| \middle| Y_{T_l}, T_l, T_{l+1}\right) \\ & \leq \left(\mathbb{E}\left(\left(\frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})}\right)^2 \middle| Y_{T_l}, T_l, T_{l+1}\right)\right)^{\frac{1}{2}} \\ & = \left(\mathbb{1}_{\{Y_{T_l} > 1\}} \frac{\mathbb{E}((R_{T_l, T} - 1)^2 | Y_{T_l}, T_l, T_{l+1})}{(Y_{T_l} - 1)^2} - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} \mathbb{E}(R_{T_l, T} - 1 | Y_{T_l}, T_l, T_{l+1})\right. \\ & \quad \left. + e^{-2\mu(T-T_{l+1})}\right)^{\frac{1}{2}}. \end{aligned} \tag{6.35}$$

Lemma 6.3.18 implies

$$\begin{aligned} \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} \mathbb{E}(R_{T_l, T} - 1 | Y_{T_l}, T_l, T_{l+1}) &= 2 \cdot \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-2\mu(T-T_{l+1})} \\ &\quad - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} \end{aligned} \tag{6.36}$$

and

$$\begin{aligned} & \frac{1}{(Y_{T_l} - 1)^2} \mathbb{1}_{\{Y_{T_l} > 1\}} \mathbb{E}((R_{T_l, T} - 1)^2 | Y_{T_l}, T_l, T_{l+1}) \\ &= \frac{1}{(Y_{T_l} - 1)^2} \mathbb{1}_{\{Y_{T_l} > 1\}} \left(\mathbb{E}(R_{T_l, T}^2 | Y_{T_l}, T_l, T_{l+1}) - 2\mathbb{E}(R_{T_l, T} | Y_{T_l}, T_l, T_{l+1}) + 1\right) \\ &= \frac{1}{(Y_{T_l} - 1)^2} \mathbb{1}_{\{Y_{T_l} > 1\}} \left(Y_{T_l} e^{-\mu(T-T_{l+1})} - Y_{T_l} e^{-2\mu(T-T_{l+1})} + Y_{T_l}^2 e^{-2\mu(T-T_{l+1})}\right. \\ & \quad \left.- \mathbb{P}(Y_{T_{l+1}} = Y_{T_l} - 1 | Y_{T_l}, T_l, T_{l+1}) \left(2(Y_{T_l} - 1) e^{-2\mu(T-T_{l+1})} + e^{-\mu(T-T_{l+1})}\right)\right) \\ & \quad - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} + \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{(Y_{T_l} - 1)^2} \\ & \leq \frac{1}{(Y_{T_l} - 1)^2} \mathbb{1}_{\{Y_{T_l} > 1\}} \left(Y_{T_l} e^{-\mu(T-T_{l+1})} + Y_{T_l} (Y_{T_l} - 1) e^{-2\mu(T-T_{l+1})}\right) - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} \\ & \quad + \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{(Y_{T_l} - 1)^2} \\ &= \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{Y_{T_l} - 1} e^{-\mu(T-T_{l+1})} + \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{(Y_{T_l} - 1)^2} e^{-\mu(T-T_{l+1})} + \frac{Y_{T_l}}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-2\mu(T-T_{l+1})} \\ & \quad - \frac{2}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\mu(T-T_{l+1})} + \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{(Y_{T_l} - 1)^2} \\ & \leq \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-2\mu(T-T_{l+1})} + \frac{1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-2\mu(T-T_{l+1})} + \frac{1}{(Y_{T_l} - 1)^2} \mathbb{1}_{\{Y_{T_l} > 1\}} \\ & \leq \left(\frac{1}{Y_{T_l} - 1} + e^{-2\mu(T-T_{l+1})}\right) \mathbb{1}_{\{Y_{T_l} > 1\}}. \end{aligned} \tag{6.37}$$

By (6.36) and (6.37), we can bound the right-hand side of (6.35) from above by

$$\begin{aligned} \left( \frac{3}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-2\mu(T-T_{l+1})} + e^{-2\mu(T-T_{l+1})} \right)^{\frac{1}{2}} &\leq \left( \frac{3}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} + \mathbb{1}_{\{Y_{T_l} = 1\}} \right)^{\frac{1}{2}} \\ &\leq \left( \frac{6}{Y_{T_l}} \right)^{\frac{1}{2}}. \end{aligned}$$

□

The following purely analytical lemma is also used in the proof of Theorem 6.3.4.

### 6.3.20 Lemma

For all  $x \geq 0$ , we have

$$\sum_{k=2}^{\infty} \frac{(-x)^k}{k!} \leq \frac{x^2}{2}. \quad (6.38)$$

*Proof:* Let  $x \geq 0$ . The left-hand side of (6.38) is equal to  $e^{-x} - 1 + x$ . For  $x = 0$ , this expression is equal to  $\frac{x^2}{2}$ . By  $1 - e^{-x} \leq x$ , the derivative of the left-hand side of (6.38) is everywhere smaller than or equal to the derivative of the right-hand side of (6.38). Hence the statement follows. □

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.

### 6.3.21 Lemma

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

$$\begin{aligned} \mathbb{E} \left( \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} (T_{l+1} - T_l)^2 \mid Y_T > 0 \right) &\leq \frac{\mu T^2}{\lambda} e^{-(\lambda - \mu)T} + 60T^2 \frac{\lambda^3(\lambda + \mu)}{(\lambda - \mu)^4} e^{-(\lambda + \mu)T} + T^2 e^{-\frac{1}{2}\lambda T} \\ &\quad + \frac{2(\lambda - \mu)}{\lambda} \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) T^2 e^{-(\lambda - \mu)T} \\ &\quad + \left( \frac{3}{4}T^2 + \frac{T}{2(\lambda + \mu)} \right) \left( 1 + 2 \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) e^{-\frac{1}{4}(\lambda - \mu)T}. \end{aligned} \quad (6.39)$$

*Proof:* For the left-hand side of (6.39), we deduce

$$\begin{aligned} &\mathbb{E} \left( \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} (T_{l+1} - T_l)^2 \mid Y_T > 0 \right) \\ &\leq \mathbb{E} \left( \mathbb{1}_{\{T_{\kappa(T)} > T\}} \mathbb{1}_{\{T_{\mathcal{K}(T)} < T_{r(J_T)}\}} \sum_{l=1}^{\mathcal{M}_T - 1} \mathbb{1}_{\{T_l \geq T_{r(J_T)}\}} (T_{l+1} - T_l) \max_{r(J_T) \leq m \leq \mathcal{M}_T - 1} (T_{m+1} - T_m) \mid Y_T > 0 \right) \\ &\quad + \mathbb{E}(\mathbb{1}_{\{T_{\kappa(T)} \leq T\}} \mid Y_T > 0) T^2 + \mathbb{E}(\mathbb{1}_{\{T_{\mathcal{K}(T)} \geq T_{r(J_T)}\}} \mid Y_T > 0) T^2. \end{aligned} \quad (6.40)$$

Note that, given  $(Y_{T_k})_{k \in \mathbb{N}}$  and  $\mathcal{K}(T)$ , the inter-event times  $T_{m+1} - T_m$  are  $\text{Exp}((\lambda + \mu)Y_{T_m})$  distributed and independent for  $m > \mathcal{K}(T)$ . In order to derive an upper bound for the first conditional expectation on the right-hand side of (6.40), we introduce a sequence of random variables  $(U_m)_{m \in \mathbb{N}}$  such that, given

$(Y_{T_k})_{k \in \mathbb{N}}$  and  $\mathcal{K}(T)$ ,  $U_m \sim \text{Exp}((\lambda + \mu) \min_{\mathcal{K}(T) < k} Y_{T_k})$  are independent and identically distributed. Then, given  $(Y_{T_k})_{k \in \mathbb{N}}$  and  $\mathcal{K}(T)$ , we have  $T_{m+1} - T_m \leq_{st} U_m$  for  $\mathcal{K}(T) < m \leq \kappa(T)$  and obtain

$$\begin{aligned} & \mathbb{E} \left( \mathbb{1}_{\{T_{\kappa(T)} > T\}} \mathbb{1}_{\{T_{\mathcal{K}(T)} < T_{r(J_T)}\}} \sum_{l=1}^{\mathcal{M}_T-1} \mathbb{1}_{\{T_l \geq T_{r(J_T)}\}} (T_{l+1} - T_l) \max_{r(J_T) \leq m \leq \mathcal{M}_T-1} (T_{m+1} - T_m) \mid Y_T > 0 \right) \\ & \leq \frac{T}{2} \mathbb{E} \left( \mathbb{1}_{\{T_{\kappa(T)} > T\}} \mathbb{1}_{\{T_{\mathcal{K}(T)} < T_{r(J_T)}\}} \max_{r(J_T) \leq m \leq \mathcal{M}_T-1} (T_{m+1} - T_m) \mid Y_\infty > 0 \right) + \frac{T^2}{4} \mathbb{P}(Y_\infty = 0 \mid Y_T > 0). \end{aligned} \quad (6.41)$$

By Lemma 6.3.5, the second summand of the right-hand side is equal to

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

The first summand of the right-hand side of (6.41) is bounded from above by

$$\begin{aligned} & \frac{T}{2} \mathbb{E} \left( \mathbb{E} \left( \mathbb{1}_{\{T_{\kappa(T)} > T\}} \mathbb{1}_{\{T_{\mathcal{K}(T)} < T_{r(J_T)}\}} \max_{\mathcal{K}(T) < m \leq \kappa(T)} (T_{m+1} - T_m) \mid \mathcal{K}(T), (Y_{T_k})_{k \geq 1} \right) \mid Y_\infty > 0 \right) \\ & \leq \frac{T}{2} \mathbb{E} \left( \mathbb{E} \left( \max_{1 \leq m \leq \kappa(T)} U_m \mid \mathcal{K}(T), (Y_{T_k})_{k \geq 1} \right) \mid Y_\infty > 0 \right) \\ & = \frac{T}{2} \mathbb{E} \left( \frac{1}{(\lambda + \mu) \min_{\mathcal{K}(T) < k} Y_{T_k}} \sum_{l=1}^{\kappa(T)} \frac{1}{l} \mid Y_\infty > 0 \right), \end{aligned}$$

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

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

$$\begin{aligned} & \frac{T}{2} \mathbb{E} \left( \frac{1}{(\lambda + \mu) \min_{\mathcal{K}(T) < k} Y_{T_k}} \sum_{l=1}^{\kappa(T)} \frac{1}{l} \mid Y_\infty > 0 \right) \\ & \leq \frac{T}{2(\lambda + \mu)} \mathbb{E} \left( \frac{1}{\min_{\mathcal{K}(T) < k} Y_{T_k}} \mid Y_\infty > 0 \right) (\log(\kappa(T)) + 1) \\ & \leq \frac{T}{2(\lambda + \mu)} \mathbb{E} \left( \frac{1}{\min_{\mathcal{K}(T) < k} Y_{T_k}} \mid Y_\infty > 0 \right) \left( \frac{3}{2}(\lambda + \mu)T + 1 \right). \end{aligned} \quad (6.42)$$

For the conditional expectation in (6.42), we obtain for  $T \geq \frac{2 \log(2)}{\lambda - \mu}$

$$\begin{aligned} \mathbb{E} \left( \frac{1}{\min_{\mathcal{K}(T) < k} Y_{T_k}} \mid Y_\infty > 0 \right) & = \mathbb{E} \left( \frac{1}{\min_{T/2 \leq t} Y_t} \mid Y_\infty > 0 \right) \\ & \leq \left( 1 + 2 \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) e^{-\frac{1}{4}(\lambda - \mu)T}, \end{aligned} \quad (6.43)$$

where the last line follows from Lemma 6.3.15 with  $\delta = 1$  and  $\gamma = \frac{1}{4}$  and Lemma 6.3.16.

Thus we can conclude that the first summand of the right-hand side of (6.41) 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 (6.40), we can use the upper bounds from Lemma 6.3.12 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

$$\frac{60\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4} T^2 e^{-(\lambda+\mu)T} + T^2 e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda-\mu)}{\lambda} \left( \log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T \right) T^2 e^{-(\lambda-\mu)T}. \quad (6.44)$$

Altogether, we obtain (6.39) for  $T \geq (\frac{2}{\lambda-\mu} \log(2) \vee \frac{2(\log(4\lambda)-\log(\lambda-\mu))}{\lambda+\mu})$ .  $\square$

### Proof of the main theorem

In order to simplify notation, we introduce  $\mathbb{E}^*(\cdot) = \mathbb{E}(\cdot | Y_T > 0)$  and  $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot | Y_T > 0)$ .

In the following we prove Theorem 6.3.4. As before, we use Theorem 3.4.1 to obtain

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

In order to find an upper bound for  $\mathbb{E}^*|\Lambda_T - M|$ , we use the triangle inequality for the absolute value after plugging in the definitions of  $\Lambda_T$  and  $M$  given by (6.6) and (6.15), which yields

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

This is bounded from above by

$$\mathbb{E}^* \left( \frac{\alpha(1 - e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T - 1)\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{\mathfrak{B}_T} S_i \mathbb{1}_{\{T_i^- > T\}} \mathbb{1}_{\{Y_T > 1\}} \right) \quad (6.45)$$

$$+ \mathbb{E}^* \left( \frac{\alpha(1 - e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T - 1)\beta} S_{J_T} \sum_{\substack{i=1 \\ i \neq J_T}}^{\mathfrak{B}_T} \mathbb{1}_{\{T_i^- > T\}} \mathbb{1}_{\{Y_T > 1\}} \right) \quad (6.46)$$

$$+ \mathbb{E}^* \left| \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{\mathfrak{B}_T} (S_i - \mathbb{E}(S)) \mathbb{1}_{\{T_i^- > T\}} \sum_{l=r(i) \vee r(J_T)}^{\mathcal{M}_T-1} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \frac{1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} \right| \quad (6.47)$$

$$\begin{aligned} & + \mathbb{E}^* \left| \frac{\alpha}{\beta} \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \frac{1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} \sum_{\substack{i=1 \\ i \neq J_T}}^{r^{-1}(l)} (\mathbb{E}(S) + S_{J_T}) \mathbb{1}_{\{T_i^- > T\}} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \right. \\ & \quad \left. - \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} (e^{-(\beta+\mu)(T-T_{l+1})} - e^{-(\beta+\mu)(T-T_l)}) \right| \quad (6.48) \end{aligned}$$

$$\begin{aligned}
& + \mathbb{E}^* \left| \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} (e^{-(\beta+\mu)(T-T_{l+1})} - e^{-(\beta+\mu)(T-T_l)}) \right. \\
& \quad \left. - \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_\infty}) (1 - e^{-(\beta+\mu)A_{J_\infty}}) \right|, \tag{6.49}
\end{aligned}$$

where  $\mathcal{M}_T$  is the number of events up to time  $T$  and  $r^{-1}(l)$  is the number of births that occur not later than the  $l$ th event for all  $l \in \{1, \dots, \mathcal{M}_T\}$ , i.e.  $r^{-1} : \{1, \dots, \mathcal{M}_T\} \rightarrow \{1, \dots, \mathfrak{B}_T\}$ ,  $l \mapsto \sum_{i=1}^{\mathfrak{B}_T} \mathbb{1}_{\{r(i) \leq l\}}$ .

In the following, we deduce upper bounds for (6.45)-(6.49).

### Upper bound for (6.45) and (6.46)

We treat (6.45) similarly to the corresponding expression in the pure birth case, but condition on  $\mathfrak{B}_T$ ,  $\mathfrak{D}_T$ ,  $J_T$  and the information which nodes survive up to time  $T$ , and obtain

$$\mathbb{E}^* \left( \frac{\alpha(1 - e^{-\beta(T-T_{\mathcal{M}_T})})}{(Y_T - 1)\beta} \mathbb{1}_{\{Y_T > 1\}} \sum_{\substack{i=1 \\ i \neq J_T}}^{\mathfrak{B}_T} S_i \mathbb{1}_{\{T_i^- > T\}} \right) \leq \frac{\alpha}{\beta} \mathbb{E}(S) \mathbb{E}^*(1 - e^{-\beta(T-T_{\mathcal{M}_T})}). \tag{6.50}$$

Since we condition on  $J_T$ , the same expression is also obtained for (6.46).

Lemma 6.3.13 reveals an upper bound for the conditional expectation on the right-hand sides of (6.50): For  $T \geq \frac{1}{\lambda - \mu} \log(2)$ , we have

$$\mathbb{E}^*(1 - e^{-\beta(T-T_{\mathcal{M}_T})}) \leq \frac{2(\lambda - \mu)}{\lambda e^{(\lambda - \mu)T}} + \frac{\beta}{\lambda} \mathbb{E}^* \left( \frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \right).$$

Lemma 6.3.14 gives us an upper bound for the second summand of the upper bound from Lemma 6.3.13:

$$\mathbb{E}^* \left( \frac{1}{Y_T - 1} \mathbb{1}_{\{Y_T > 1\}} \right) \leq \frac{\lambda - \mu}{\lambda e^{(\lambda - \mu)T} - \mu} \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right).$$

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

$$\frac{2}{\lambda} \left( \lambda - \mu + \beta \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) \right) e^{-(\lambda - \mu)T}. \tag{6.51}$$

Plugging this expression in the right-hand side of (6.50) results in the following upper bound for the sum of (6.45) and (6.46):

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

### Upper bound for (6.47)

Recall that  $R_{T_l, T}$  is the number of nodes that are alive at time  $T_l$  and survive up to time  $T$ . We compute

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



$$\begin{aligned}
 &\leq \mathbb{E}^* \left( \left| \frac{\alpha}{\beta} \sum_{\substack{l=r(J_T) \\ R_{T_l, T} \geq 2}}^{\mathcal{M}_T-1} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \frac{\mathbb{1}_{\{Y_{T_l} > 1\}}}{R_{T_l, T} - 1} \sum_{\substack{i=1 \\ i \neq J_T}}^{r^{-1}(l)} (S_i - \mathbb{E}(S)) \mathbb{1}_{\{T_i^- > T\}} \right| \right) \\
 &\leq \mathbb{E}^* \left( \frac{\alpha}{\beta} \sum_{\substack{l=r(J_T) \\ R_{T_l, T} \geq 2}}^{\mathcal{M}_T-1} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} \right. \\
 &\quad \left. \cdot \mathbb{E} \left( \frac{1}{R_{T_l, T} - 1} \left| \sum_{\substack{i=1 \\ i \neq J_T}}^{r^{-1}(l)} (S_i - \mathbb{E}(S)) \mathbb{1}_{\{T_i^- > T\}} \right| \right) \middle| (Y_t)_{0 \leq t \leq T}, J_T \right) \right). \tag{6.52}
 \end{aligned}$$

Note in the second line that we can restrict the sum to  $R_{T_l, T} \geq 2$  because  $R_{T_l, T} \geq 1$  by  $l \geq r(J_T)$  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 (6.52) 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 (6.52) is smaller than or equal to

$$\mathbb{E}^* \left( \left| \frac{\alpha}{\beta} \sum_{\substack{l=r(J_T) \\ R_{T_l, T} \geq 2}}^{\mathcal{M}_T-1} (e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)}) \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} \frac{\sigma_S}{\sqrt{R_{T_l, T} - 1}} \right| \right). \tag{6.53}$$

Since  $e^{-\beta(T-T_{l+1})} - e^{-\beta(T-T_l)} \leq \beta(T_{l+1} - T_l)$  and  $R_{T_l, T} \leq Y_{T_l}$ , the expression (6.53) is smaller than or equal to

$$\begin{aligned}
 \mathbb{E}^* \left( \alpha \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \frac{\sigma_S}{\sqrt{Y_{T_l} - 1}} \mathbb{1}_{\{Y_{T_l} > 1\}} (T_{l+1} - T_l) \right) &\leq \mathbb{E}^* \left( \alpha \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \frac{\sigma_S}{\sqrt{Y_{T_l} - \frac{Y_{T_l}}{2}}} \mathbb{1}_{\{Y_{T_l} > 1\}} (T_{l+1} - T_l) \right) \\
 &\leq \sqrt{2} \alpha \sigma_S T \mathbb{E}^* \left( \max_{r(J_T) \leq k \leq \mathcal{M}_T-1} \frac{1}{\sqrt{Y_{T_k}}} \right) \\
 &\leq \sqrt{2} \alpha \sigma_S T \mathbb{E}^* \left( \mathbb{1}_{\{\mathcal{K}(T) < r(J_T)\}} \max_{r(J_T) \leq k \leq \mathcal{M}_T-1} \frac{1}{\sqrt{Y_{T_k}}} \right) \\
 &\quad + \sqrt{2} \alpha \sigma_S T \mathbb{E}^* \left( \mathbb{1}_{\{\mathcal{K}(T) \geq r(J_T)\}} \max_{r(J_T) \leq k \leq \mathcal{M}_T-1} \frac{1}{\sqrt{Y_{T_k}}} \right), \tag{6.54}
 \end{aligned}$$

where  $\mathcal{K}(T) = \max\{k : T_k \leq T/2\}$  (see Definition 6.3.11). We have

$$\begin{aligned}
 &\mathbb{E}^* \left( \mathbb{1}_{\{\mathcal{K}(T) < r(J_T)\}} \max_{r(J_T) \leq k \leq \mathcal{M}_T-1} \frac{1}{\sqrt{Y_{T_k}}} \right) \\
 &\leq \mathbb{E} \left( \mathbb{1}_{\{\mathcal{K}(T) < r(J_T)\}} \max_{r(J_T) \leq k \leq \mathcal{M}_T-1} \frac{1}{\sqrt{Y_{T_k}}} \middle| Y_\infty > 0 \right) \mathbb{P}^*(Y_\infty > 0) + 1 \cdot \mathbb{P}^*(Y_\infty = 0), \tag{6.55}
 \end{aligned}$$

where  $Y_\infty := \lim_{t \rightarrow \infty} Y_t$  as before.

By Lemma 6.3.5, we have for the second summand of the right-hand side of (6.55)

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

The first summand of the right-hand side of (6.55) is smaller than or equal to

$$\mathbb{E}\left(\max_{\mathcal{K}(T) < k} \frac{1}{\sqrt{Y_{T_k}}} \mid Y_\infty > 0\right) = \mathbb{E}\left(\max_{\frac{T}{2} \leq t} \frac{1}{\sqrt{Y_t}} \mid Y_\infty > 0\right).$$

By combining Lemma 6.3.15, where  $\delta = 1/2$  and  $\gamma = 1/6$ , and Lemma 6.3.16, we obtain for  $T \geq \frac{2}{\lambda - \mu} \log(2)$

$$\begin{aligned} \mathbb{E}\left(\max_{\frac{T}{2} \leq t} \frac{1}{\sqrt{Y_t}} \mid Y_\infty > 0\right) &\leq e^{-\frac{1}{6}(\lambda - \mu)T} + \mathbb{E}\left(\frac{1}{Y_{\frac{T}{2}}} \mid Y_\infty > 0\right) e^{\frac{1}{3}(\lambda - \mu)T} \\ &\leq \left(1 + 2 \log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right) e^{-\frac{1}{6}(\lambda - \mu)T}. \end{aligned} \quad (6.56)$$

Thus for  $T \geq \frac{2}{\lambda - \mu} \log(2)$ , the first summand of the right-hand side of (6.54) is smaller than or equal to

$$\sqrt{2}\alpha \frac{\mu}{\lambda} \sigma_S T e^{-(\lambda - \mu)T} + \sqrt{2}\alpha \sigma_S \left(1 + 2 \log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T\right) T e^{-\frac{1}{6}(\lambda - \mu)T}. \quad (6.57)$$

By Lemma 6.3.12(i), we obtain that for  $T \geq \frac{1}{\lambda - \mu} \log(2)$ , the second summand of the right-hand side of (6.54) is bounded from above by

$$\sqrt{2}\alpha \sigma_S T \left( e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda - \mu)}{\lambda} \left( \log\left(\frac{\lambda}{\lambda - \mu}\right) + (\lambda - \mu)T \right) e^{-(\lambda - \mu)T} \right). \quad (6.58)$$

Thus for  $T \geq \frac{2}{\lambda - \mu} \log(2)$ , the expression (6.47) is bounded from above by the sum of (6.57) and (6.58).

### Upper bound for (6.48)

For (6.48), we have

$$\begin{aligned} &\mathbb{E}^* \left| \frac{\alpha}{\beta} \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} \frac{1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} \sum_{\substack{i=1 \\ i \neq J_T}}^{r^{-1}(l)} (\mathbb{E}(S) + S_{J_T}) \mathbb{1}_{\{T_i^- > T\}} (e^{-\beta(T - T_{i+1})} - e^{-\beta(T - T_i)}) \right. \\ &\quad \left. - \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} (e^{-(\beta + \mu)(T - T_{i+1})} - e^{-(\beta + \mu)(T - T_i)}) \right| \\ &= \mathbb{E}^* \left| \frac{\alpha(\beta + \mu)}{\beta(\beta + \mu)} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} (e^{-\beta(T - T_{i+1})} - e^{-\beta(T - T_i)}) \right. \\ &\quad \left. - \frac{\alpha\beta}{\beta(\beta + \mu)} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} (e^{-(\beta + \mu)(T - T_{i+1})} - e^{-(\beta + \mu)(T - T_i)}) \right| \end{aligned} \quad (6.59)$$

since the second sum on the left-hand side of (6.59) has exactly  $R_{T_l, T} - 1$  non-zero summands. The right-hand side of (6.59) is equal to

$$\begin{aligned} &\mathbb{E}^* \left| \frac{\alpha}{\beta(\beta + \mu)} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T - 1} \left( (\beta + \mu) \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\beta(T - T_{i+1})} (1 - e^{-\beta(T_{i+1} - T_i)}) \right. \right. \\ &\quad \left. \left. - \beta e^{-(\beta + \mu)(T - T_{i+1})} (1 - e^{-(\beta + \mu)(T_{i+1} - T_i)}) \right) \right|. \end{aligned} \quad (6.60)$$

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

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

in order to obtain that (6.60) is smaller than or equal to

$$\begin{aligned} & \mathbb{E}^* \left| \frac{\alpha}{\beta(\beta + \mu)} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} (\beta + \mu)\beta \left( \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right) e^{-\beta(T-T_{l+1})} (T_{l+1} - T_l) \right| \\ & + \mathbb{E}^* \left| \frac{\alpha}{\beta(\beta + \mu)} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \left( \beta e^{-(\beta+\mu)(T-T_{l+1})} \sum_{k=2}^{\infty} \frac{(-(\beta + \mu)(T_{l+1} - T_l))^k}{k!} \right. \right. \\ & \quad \left. \left. - (\beta + \mu) \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} e^{-\beta(T-T_{l+1})} \sum_{k=2}^{\infty} \frac{(-\beta(T_{l+1} - T_l))^k}{k!} \right) \right| \\ & \leq \mathbb{E}^* \left( \alpha (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\infty} \mathbb{1}_{\{T_{l+1} \leq T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \right) \\ & + \mathbb{E}^* \left( \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \sum_{k=2}^{\infty} \frac{(-(\beta + \mu)(T_{l+1} - T_l))^k}{k!} \right) \\ & + \mathbb{E}^* \left( \frac{\alpha}{\beta} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} \sum_{k=2}^{\infty} \frac{(-\beta(T_{l+1} - T_l))^k}{k!} \right). \end{aligned} \quad (6.61)$$

Firstly, we consider the first summand of the right-hand side of (6.61). Since the social index  $S_{J_T}$  is independent of all other random variables appearing in (6.61), this summand is equal to

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

We derive

$$\begin{aligned} & \mathbb{E}^* \left( \sum_{l=r(J_T)}^{\infty} \mathbb{1}_{\{T_{l+1} < T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \right) \\ & \leq \mathbb{E}^* \left( \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{1}_{\{T_{l+1} < T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \right) \\ & \quad + T \mathbb{P}^*(\mathcal{K}(T) \geq r(J_T)). \end{aligned} \quad (6.62)$$

For  $\frac{1}{\lambda-\mu} \log(2)$ , the second summand of the right-hand side of (6.62) is smaller than or equal to

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

by Lemma 6.3.12(i). The first summand of the right-hand side of (6.62) is equal to

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

For the inner expectation, we have

$$\begin{aligned} & \mathbb{E}\left(\mathbb{1}_{\{T_{l+1} < T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \mid \mathcal{K}(T), Y_T > 0\right) \\ &= \mathbb{E}\left(\mathbb{1}_{\{T_{l+1} < T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \mid \mathcal{K}(T), Y_{T_l} > 0\right) \frac{\mathbb{P}(Y_{T_l} > 0 | \mathcal{K}(T))}{\mathbb{P}(Y_T > 0 | \mathcal{K}(T))}. \end{aligned} \quad (6.63)$$

For the fraction, we obtain for  $l \geq \mathcal{K}(T) + 1$

$$\frac{\mathbb{P}(Y_{T_l} > 0 | \mathcal{K}(T))}{\mathbb{P}(Y_T > 0 | \mathcal{K}(T))} \leq \frac{\mathbb{P}(Y_{T_{\mathcal{K}(T)}} > 0 | \mathcal{K}(T))}{\mathbb{E}(\mathbb{P}(Y_T > 0 | \mathcal{K}(T), Y_{T_{\mathcal{K}(T)}}) | \mathcal{K}(T))}. \quad (6.64)$$

Note that, given  $\mathcal{K}(T)$  and  $Y_{T_{\mathcal{K}(T)}}$ , we know that almost surely  $Y_{T_{\mathcal{K}(T)}}$  nodes are alive at time  $T/2$  since a jump at this time has probability zero. 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 2.2.2, the conditional probability  $\mathbb{P}(Y_T > 0 | \mathcal{K}(T), Y_{T_{\mathcal{K}(T)}})$  is equal to  $p_0(T/2)^{Y_{T_{\mathcal{K}(T)}}$ , which implies that (6.64) is equal to

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

Thus the right-hand side of (6.63) is smaller than or equal to

$$\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{T_{l+1} < T\}} \left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| (T_{l+1} - T_l) \mid Y_{T_l}, T_l, \mathcal{K}(T)\right) \mid \mathcal{K}(T), Y_{T_l} > 0\right) \frac{\lambda}{\lambda - \mu}. \quad (6.65)$$

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 (6.65) is equal to

$$\mathbb{E}\left(\mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) \mathbb{E}\left(\left| \frac{R_{T_l, T} - 1}{Y_{T_l} - 1} \mathbb{1}_{\{Y_{T_l} > 1\}} - e^{-\mu(T-T_{l+1})} \right| \mid Y_{T_l}, T_l, T_{l+1}\right) \mid \mathcal{K}(T), Y_{T_l} > 0\right) \frac{\lambda}{\lambda - \mu}. \quad (6.66)$$

By Lemma 6.3.19, summing over  $l$  and taking the expectation yields the following upper bound for the first summand of the right-hand side of (6.62):

$$\mathbb{E}^*\left(\sum_{l=\mathcal{K}(T)+1}^{\infty} \sqrt{6} \frac{\lambda}{\lambda - \mu} \mathbb{E}\left(\mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) Y_{T_l}^{-\frac{1}{2}} \mid \mathcal{K}(T), Y_{T_l} > 0\right)\right). \quad (6.67)$$

Obviously, (6.67) is equal to

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

We may conclude that for for  $T \geq \frac{1}{\lambda-\mu} \log(2)$ , the first summand of (6.61) is bounded from above by

$$\begin{aligned}
 & 2\alpha \mathbb{E}(S) \left( \sqrt{6} \frac{\lambda}{\lambda-\mu} \mathbb{E}^* \left( \mathbb{1}_{\{Y_\infty > 0\}} \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{E} \left( \mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) Y_{T_l}^{-\frac{1}{2}} \mid \mathcal{K}(T), Y_{T_l} > 0 \right) \right) \right. \\
 & \quad \left. + \sqrt{6} \frac{\lambda}{\lambda-\mu} \mathbb{E}^* \left( \mathbb{1}_{\{Y_\infty = 0\}} \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{E} \left( \mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) Y_{T_l}^{-\frac{1}{2}} \mid \mathcal{K}(T), Y_{T_l} > 0 \right) \right) \right) \\
 & \quad + T e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda-\mu)}{\lambda} \left( \log \left( \frac{\lambda}{\lambda-\mu} \right) + (\lambda-\mu)T \right) T e^{-(\lambda-\mu)T}. \tag{6.68}
 \end{aligned}$$

For the first outer expectation in (6.68), we have

$$\begin{aligned}
 & \mathbb{E}^* \left( \mathbb{1}_{\{Y_\infty > 0\}} \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{E} \left( \mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) Y_{T_l}^{-\frac{1}{2}} \mid \mathcal{K}(T), Y_{T_l} > 0 \right) \right) \\
 & \leq \mathbb{E} \left( \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{E} \left( \mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) Y_{T_l}^{-\frac{1}{2}} \mid \mathcal{K}(T) \right) \mid Y_\infty > 0 \right) \\
 & \leq \mathbb{E} \left( \max_{\frac{T}{2} \leq t} Y_t^{-\frac{1}{2}} \mathbb{E} \left( \sum_{l=\mathcal{K}(T)+1}^{\infty} \mathbb{1}_{\{T_{l+1} < T\}} (T_{l+1} - T_l) \mid \mathcal{K}(T) \right) \mid Y_\infty > 0 \right) \\
 & \leq \frac{T}{2} \mathbb{E} \left( \max_{\frac{T}{2} \leq t} Y_t^{-\frac{1}{2}} \mid Y_\infty > 0 \right).
 \end{aligned}$$

By Inequality (6.56), we obtain that for  $T \geq \frac{2}{\lambda-\mu} \log(2)$ , this expression is smaller than or equal to

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

For the second outer expectation in (6.68), we obtain

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

where the last equality follows from Lemma 6.3.5.

In conclusion, for  $T \geq \frac{2}{\lambda-\mu} \log(2)$ , the first summand of the right-hand side of (6.61) is bounded from above by

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

Since the social index  $S_{J_T}$  is independent of all other random variables appearing in (6.61), the second summand of the right-hand side of (6.61) is smaller than or equal to

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

By Lemma 6.3.20, this expression is bounded from above by

$$\frac{\alpha}{\beta + \mu} \mathbb{E}(S) \mathbb{E}^* \left( \sum_{l=r(J_T)}^{\mathcal{M}_T-1} (\beta + \mu)^2 (T_{l+1} - T_l)^2 \right). \quad (6.69)$$

For  $T \geq (\frac{2}{\lambda-\mu} \log(2) \vee \frac{2(\log(4\lambda) - \log(\lambda-\mu))}{\lambda+\mu})$ , Lemma 6.3.21 implies that (6.69) and hence also the second summand of the right-hand side of (6.61) is smaller than or equal to

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

For the third summand of the right-hand side of (6.61), we obviously obtain the same upper bound except that the factor  $\beta + \mu$  is replaced by  $\beta$ .

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 (6.48) is bounded from above by

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

### Upper bound for (6.49)

Recall that we arranged  $S_{J_T} = S_{J_\infty}$ . Since the sum in (6.49) telescopes, this implies

$$\begin{aligned} & \mathbb{E}^* \left( \left| \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_T}) \sum_{l=r(J_T)}^{\mathcal{M}_T-1} (e^{-(\beta+\mu)(T-T_{l+1})} - e^{-(\beta+\mu)(T-T_l)}) \right. \right. \\ & \quad \left. \left. - \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_\infty}) (1 - e^{-(\beta+\mu)A_{J_\infty}}) \right| \right) \\ & = \mathbb{E}^* \left( \left| \frac{\alpha}{\beta + \mu} (\mathbb{E}(S) + S_{J_\infty}) \left( e^{-(\beta+\mu)(T-T_{\mathcal{M}_T}}) - e^{-(\beta+\mu)(T-T_{r(J_T)})} - (1 - e^{-(\beta+\mu)A_{J_\infty}}) \right) \right| \right) \\ & \leq \frac{2\alpha}{\beta + \mu} \mathbb{E}(S) \left( \mathbb{E}^* \left( 1 - e^{-(\beta+\mu)(T-T_{\mathcal{M}_T})} \right) + \mathbb{E}^* \left( \left| e^{-(\beta+\mu)A_{J_T}(T)} - e^{-(\beta+\mu)A_{J_\infty}} \right| \right) \right), \quad (6.70) \end{aligned}$$

where the last inequality holds since  $S_{J_\infty}$  is independent of all other random variables appearing in (6.70).

By combining Lemma 6.3.13 and Lemma 6.3.14 as before, we obtain for the first conditional expectation on the right-hand side of (6.70) the upper bound

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

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

For the second conditional expectation on the right-hand side of (6.70), Corollary 4.3.3 implies

$$\mathbb{E}^* \left( \left| e^{-(\beta + \mu)A_{J_T}(T)} - e^{-(\beta + \mu)A_{J_\infty}} \right| \right) \leq \left( \frac{2\lambda}{\beta + \mu + \lambda} + \frac{2(\lambda - \mu)}{\lambda} \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) \right) e^{-(\lambda - \mu)T}$$

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

Altogether, we obtain that the right-hand side of (6.70) 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 (6.45)-(6.49), we have for  $T \geq \left( \frac{2}{\lambda - \mu} \log(2) \vee \frac{2(\log(4\lambda) - \log(\lambda - \mu))}{\lambda + \mu} \right)$

$$\begin{aligned} & d_{TV}(\text{MixPo}(\Lambda_T^*), \text{MixPo}(\mathbb{M}^*)) \\ & \leq \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) T e^{-\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( T e^{-\frac{1}{2}\lambda T} + \frac{2(\lambda - \mu)}{\lambda} \left( \log \left( \frac{\lambda}{\lambda - \mu} \right) + (\lambda - \mu)T \right) T e^{-(\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} \end{aligned}$$

$$\begin{aligned}
&= \left( \sqrt{2}\alpha\sigma_S + \sqrt{6}\alpha\frac{\lambda}{\lambda-\mu}\mathbb{E}(S) \right) \left( 1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T \right) T e^{-\frac{1}{6}(\lambda-\mu)T} \\
&+ \frac{\alpha}{2}(2\beta+\mu)\mathbb{E}(S) \left( \frac{3}{2}T + \frac{1}{\lambda+\mu} \right) \left( 1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T \right) T e^{-\frac{1}{4}(\lambda-\mu)T} \\
&+ \left( \sqrt{2}\alpha\sigma_S + 2\alpha\mathbb{E}(S) + \alpha(2\beta+\mu)\mathbb{E}(S)T \right) T e^{-\frac{1}{2}\lambda T} \\
&+ \left( \alpha\mathbb{E}(S) \left( \frac{4}{\beta\lambda} \left( \lambda - \mu + \beta \left( 1 + \frac{\beta+\lambda}{\beta+\mu} + \left( 1 + \frac{2\beta+\mu}{2}T \right) (\lambda-\mu)T \right) \right) \left( \log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T \right) \right) \right. \\
&\quad \left. + \sqrt{6}\frac{\lambda}{\lambda-\mu}T + \frac{(2\beta+\mu)\mu}{4\lambda}T^2 + \frac{4(\lambda-\mu)}{\lambda(\beta+\mu)} + \frac{4}{\beta+\mu}\frac{\lambda}{\beta+\lambda+\mu} \right) \\
&\quad + \sqrt{2}\alpha\sigma_S \left( \frac{\mu}{\lambda} + \frac{2(\lambda-\mu)}{\lambda} \left( \log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T \right) \right) T e^{-(\lambda-\mu)T} \\
&+ 60\alpha\frac{\lambda^3(\lambda+\mu)}{(\lambda-\mu)^4}(2\beta+\mu)\mathbb{E}(S)T^2e^{-(\lambda+\mu)T}. \tag{6.71}
\end{aligned}$$

In the remainder of the proof, we find a simpler upper bound by elementary calculations. In order to do so, we assume  $T \geq \frac{2\log(4(\lambda(\lambda-\mu)^{-1}))}{\lambda-\mu}$ . Then we have

$$\begin{aligned}
1 + 2\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T &\leq \frac{5}{2}(\lambda-\mu)T, \\
e^{-\frac{1}{4}(\lambda-\mu)T} &= e^{-\frac{1}{6}(\lambda-\mu)T} e^{-\frac{1}{12}(\lambda-\mu)T} \leq \left( 4^{-\frac{1}{6}} \wedge \frac{12}{(\lambda-\mu)T} \right) e^{-\frac{1}{6}(\lambda-\mu)T} \leq \left( \frac{2}{5} \wedge \frac{12}{(\lambda-\mu)T} \right) e^{-\frac{1}{6}(\lambda-\mu)T}, \\
e^{-\frac{1}{2}(\lambda-\mu)T} &= e^{-\frac{1}{6}(\lambda-\mu)T} e^{-\frac{1}{3}(\lambda-\mu)T} \leq 4^{-\frac{2}{3}} e^{-\frac{1}{6}(\lambda-\mu)T} \leq \frac{2}{5} e^{-\frac{1}{6}(\lambda-\mu)T}
\end{aligned}$$

and finally  $2 \leq (\lambda-\mu)T$ . Thus the first three lines of the right-hand side of (6.71) are smaller than or equal to

$$\alpha \left( \left( \frac{5\sqrt{6}}{2}\frac{\lambda}{\lambda-\mu} + \frac{2}{5} + \left( \beta + \frac{\mu}{2} \right) \left( \frac{229}{5(\lambda-\mu)} + \frac{2}{\lambda+\mu} \right) \right) \mathbb{E}(S) + \frac{27}{10}\sqrt{2}\sigma_S \right) (\lambda-\mu)T^2 e^{-\frac{1}{6}(\lambda-\mu)T}. \tag{6.72}$$

Note that we have the following inequalities:

$$\begin{aligned}
\log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T &\leq \frac{3}{2}(\lambda-\mu)T, \\
1 + \log\left(\frac{\lambda}{\lambda-\mu}\right) + (\lambda-\mu)T &\leq 2(\lambda-\mu)T, \\
e^{-(\lambda+\mu)T} &= e^{-2\mu T} e^{-(\lambda-\mu)T} \leq \frac{1}{2\mu T} e^{-(\lambda-\mu)T}.
\end{aligned}$$

Thus the last four lines of (6.71) are smaller than or equal to

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



$$+ \frac{4(\lambda - \mu)}{\lambda(\beta + \mu)} \mathbb{E}(S) + \frac{30\lambda^3(\lambda + \mu)}{\mu(\lambda - \mu)^4} (2\beta + \mu) \mathbb{E}(S) T + \frac{4}{\beta + \mu} \frac{\lambda}{\beta + \lambda + \mu} \mathbb{E}(S) \Big) e^{-(\lambda - \mu)T}. \quad (6.73)$$

We first consider those terms in (6.73) that depend on  $\mathbb{E}(S)$ . Since  $2 \leq (\lambda - \mu)T$ , grouping the terms where  $\beta$  dominates in the denominator, we have

$$\begin{aligned} & \frac{8\alpha\mathbb{E}(S)}{\beta} \frac{\lambda - \mu}{\lambda} e^{-(\lambda - \mu)T} + \frac{4\alpha\lambda\mathbb{E}(S)}{\beta(\beta + \lambda + \mu)} e^{-(\lambda - \mu)T} + \frac{6(\lambda - \mu)}{\beta} T e^{-(\lambda - \mu)T} \\ & \leq \frac{\alpha\mathbb{E}(S)}{\beta} \left( \frac{2(\lambda - \mu)^2}{\lambda} + 2(\lambda - \mu) + 6(\lambda - \mu) \right) T e^{-(\lambda - \mu)T} \\ & \leq \frac{\alpha\mathbb{E}(S)}{\beta} 10(\lambda - \mu) T e^{-(\lambda - \mu)T} \\ & \leq \frac{5\alpha\mathbb{E}(S)}{\beta} (\lambda - \mu)^2 T^2 e^{-(\lambda - \mu)T}; \end{aligned} \quad (6.74)$$

grouping the  $\beta$ -free terms in (6.73) yields

$$\begin{aligned} & \alpha\mathbb{E}(S) \left( 12 \frac{\lambda - \mu}{\lambda} T + 3 \frac{(\lambda - \mu)^2 \mu}{\lambda} T^3 + 3 \frac{(\lambda - \mu)^2}{\lambda} T^2 + \frac{\mu^2}{4\lambda} T^2 + \sqrt{6} \frac{\lambda}{\lambda - \mu} T + \frac{30\lambda^3(\lambda + \mu)}{(\lambda - \mu)^4} T \right) e^{-(\lambda - \mu)T} \\ & \leq 12\alpha\mathbb{E}(S) \frac{\lambda - \mu}{\lambda} T e^{-(\lambda - \mu)T} + \alpha\mathbb{E}(S) \left( \frac{3(\lambda - \mu)^2}{\lambda} (\mu T + 1) + \frac{\mu^2}{4\lambda} + \frac{\sqrt{6}\lambda}{2} + \frac{15\lambda^3(\lambda + \mu)}{(\lambda - \mu)^3} \right) T^2 e^{-(\lambda - \mu)T} \\ & \leq \alpha\mathbb{E}(S) \left( \frac{3(\lambda - \mu)^2}{\lambda} (\mu T + 3) + \frac{\mu^2}{4\lambda} + \frac{\sqrt{6}\lambda}{2} + \frac{15\lambda^3(\lambda + \mu)}{(\lambda - \mu)^3} \right) T^2 e^{-(\lambda - \mu)T}; \end{aligned} \quad (6.75)$$

and considering the terms in (6.73) where  $\beta$  dominates in the numerator, we obtain

$$\begin{aligned} & \alpha\beta\mathbb{E}(S) \frac{\mu}{2\lambda} T^2 e^{-(\lambda - \mu)T} + 16\alpha\beta\mathbb{E}(S) \frac{\lambda^3(\lambda + \mu)}{\mu(\lambda - \mu)^4} T e^{-(\lambda - \mu)T} + 6\alpha\beta\mathbb{E}(S) \frac{(\lambda - \mu)^2}{\lambda} T^3 e^{-(\lambda - \mu)T} \\ & \leq \alpha\beta\mathbb{E}(S) \left( \frac{\mu}{2\lambda} + \frac{8\lambda^3(\lambda + \mu)}{\mu(\lambda - \mu)^3} + 6\alpha\beta \frac{(\lambda - \mu)^2}{\lambda} T \right) T^2 e^{-(\lambda - \mu)T}. \end{aligned} \quad (6.76)$$

For the one term in (6.73) that depends on  $\sigma_S$ , the inequality  $2 \leq (\lambda - \mu)T$  implies

$$\begin{aligned} \sqrt{2} \left( \frac{\mu}{\lambda} T + \frac{3(\lambda - \mu)^2}{\lambda} T^2 \right) \sigma_S e^{-(\lambda - \mu)T} & \leq \sqrt{2} \left( \frac{\mu(\lambda - \mu)}{2\lambda} + \frac{3(\lambda - \mu)^2}{\lambda} \right) \sigma_S T^2 e^{-(\lambda - \mu)T} \\ & = \sqrt{2} \frac{6\lambda^2 - 5\mu\lambda + \mu^2}{2\lambda} \sigma_S T^2 e^{-(\lambda - \mu)T} \\ & \leq 3\sqrt{2}\lambda\sigma_S T^2 e^{-(\lambda - \mu)T}. \end{aligned} \quad (6.77)$$

Combining (6.74)-(6.77), we obtain the following upper bound for the right-hand side of (6.73):

$$\begin{aligned} & \alpha\beta \left( \left( \frac{1}{2} + \frac{\mu}{4\beta} \right) \frac{\mu}{\lambda} + \left( \frac{8}{\mu} + \frac{4}{\beta} \right) \frac{\lambda^3(\lambda + \mu)}{(\lambda - \mu)^3} + \left( 6T + \frac{\sqrt{6}}{2} + \frac{3}{\beta} (\mu T + 3) + \frac{5\lambda}{\beta^2} \right) \lambda \right) \mathbb{E}(S) T^2 e^{-(\lambda - \mu)T} \\ & + 3\sqrt{2}\alpha\lambda\sigma_S T^2 e^{-(\lambda - \mu)T}. \end{aligned} \quad (6.78)$$

The total upper bound for the theorem is the sum of (6.72) and (6.78), and this sum is obviously of the desired order.  $\square$

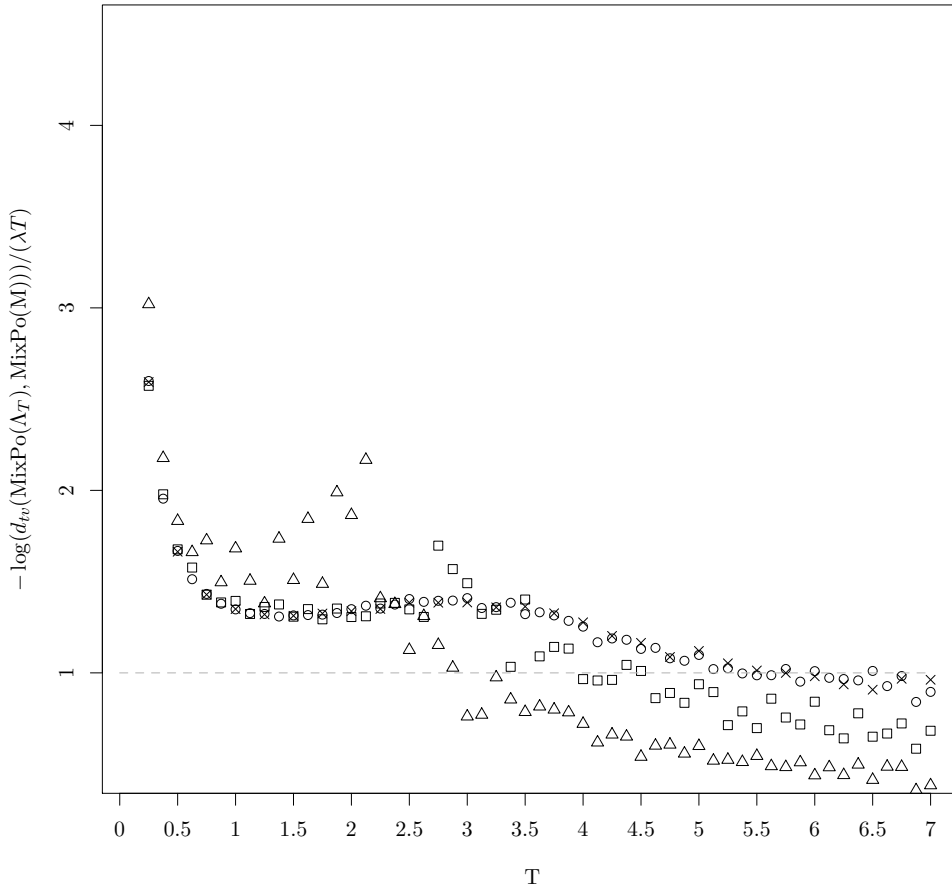


Figure 6.1: Estimated  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/(\lambda T)$  for  $S = \alpha = \beta = \lambda = 1$  and  $\mu = 0$  based on  $10^6$  (crosses), 500,000 (circles), 5,000 (squares) and 500 (triangles) simulated realizations of a linear birth and death process, respectively

### 6.3.3 Sharpness of the convergence rates

In this subsection, we examine how close our convergence rates are to the actual ones by simulating several quantities. For simplicity, we restrict ourselves to the pure birth case.

We set  $\alpha = \beta = \lambda = 1$  and begin with the case where the social index  $S$  is constant and equal to 1. In Figure 6.1, we consider the total variation distance between the degree distribution  $\text{MixPo}(\Lambda_T)$  and the asymptotic degree distribution  $\text{MixPo}(M)$ . In order to estimate the total variation distance, we consider Corollary 3.1.6, which yields

$$d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) = \frac{1}{2} \sum_{k=0}^{\infty} |f(k) - g(k)|,$$

where  $f$  and  $g$  denote the probability mass functions of the  $\text{MixPo}(\Lambda_T)$  and the  $\text{MixPo}(M)$  distribution, respectively. We computed the corresponding relative frequencies in order to approximate the probability mass functions. Note that the very natural estimator we used here is biased such that we need a huge number of iterations in order to obtain reliable results (see Figure 6.1). Note further

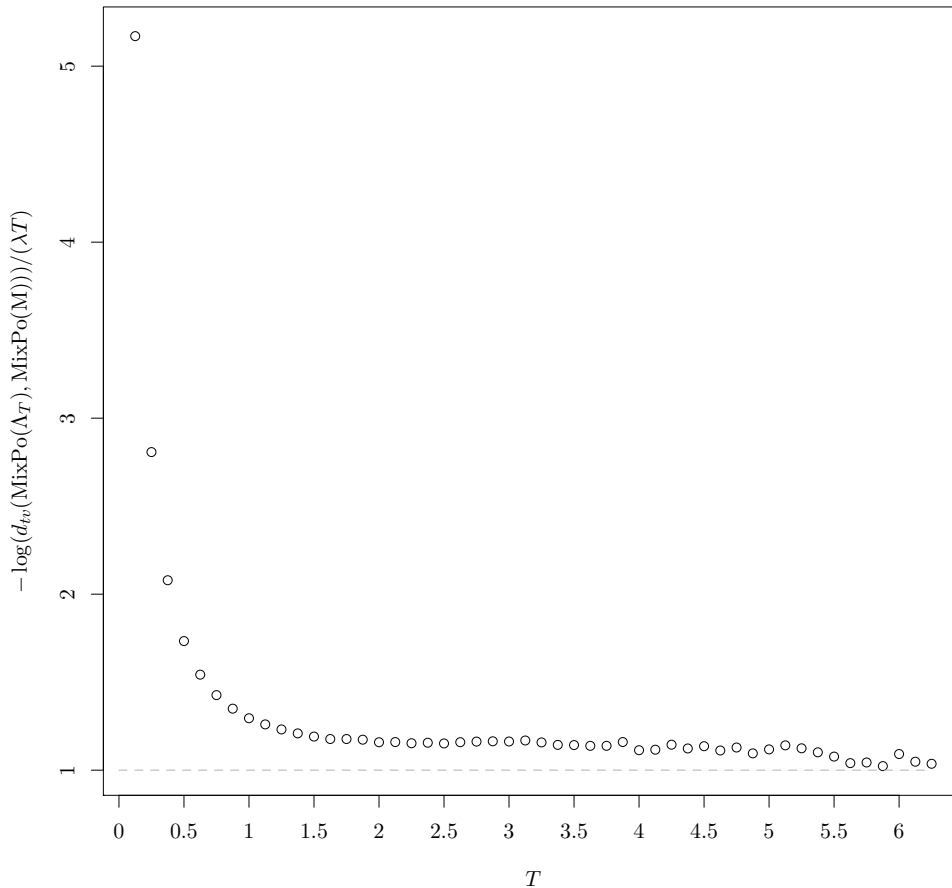


Figure 6.2: Estimated  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/(\lambda T)$  for  $S \sim \text{Exp}(1)$ ,  $\alpha = \beta = \lambda = 1$  and  $\mu = 0$  based on 250,000 simulated realizations of a linear birth and death process

that most of the differences  $f(k) - g(k)$  are very small for large  $T$ , which can easily lead to numerical problems.

Since the standard deviation  $\sigma_S$  is 0 in this case, Theorem 6.3.2 gives us the rate  $Te^{-\lambda T}$ . In Figure 6.1, we consider the left-hand side in Theorem 6.3.2. More precisely, we consider the expression  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/T$ , which is of the order  $\lambda + O(\log(T)/T)$  if our rate is sharp. Indeed Figure 6.1 confirms that  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/(\lambda T)$  approaches a value close to 1, i.e. our rate is close to the actual one in this case. The gap between 1 and the value simulated (using  $10^6$  iterations) is due to a small simulation error or terms of lower order.

In order to examine the case where  $\sigma_S > 0$ , we consider  $S \sim \text{Exp}(1)$ . Theorem 6.3.2 gives us the rate  $\sqrt{T}e^{-\frac{\lambda}{2}T}$ . In Figure 6.2, we consider the expression  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/T$  again, which is of the order  $\lambda/2 + O(\log(T)/T)$  if our rate is sharp. However, Figure 6.1 suggests that  $-\log(d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)))/(\lambda T)$  approaches a value close to 1 again, i.e. our rate does not seem to be sharp in this case.

### 6.3.22 Remark

Note that, due to the problems mentioned above, the simulations are very time-consuming and require

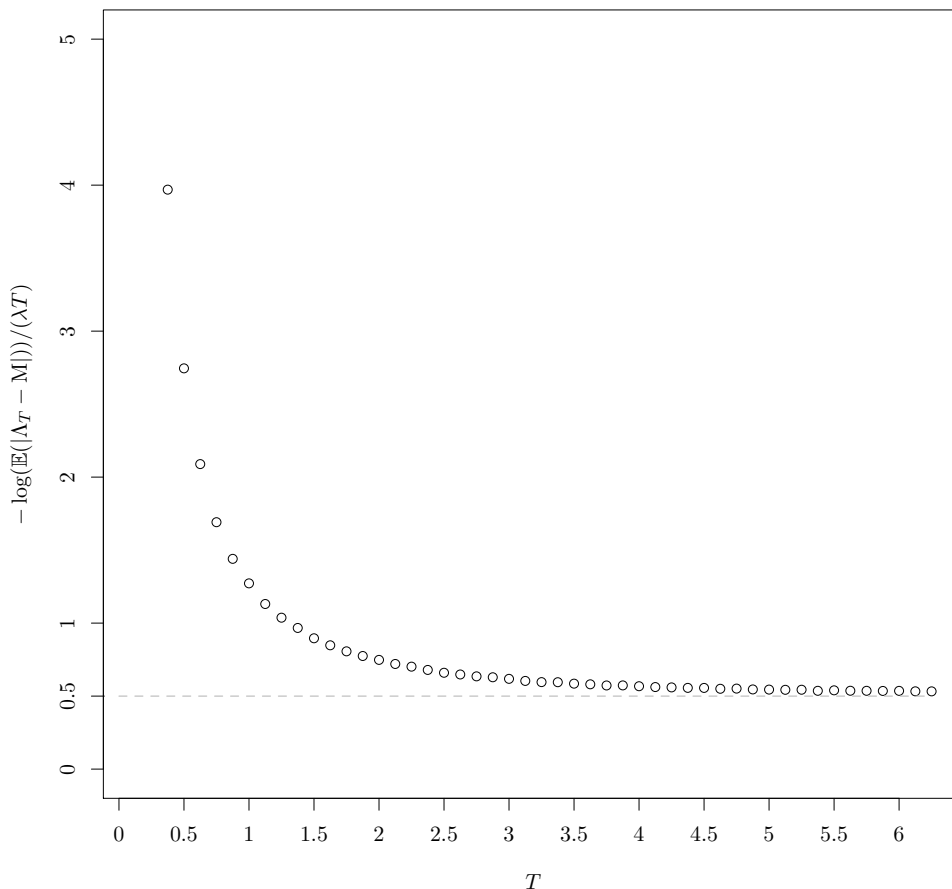


Figure 6.3: Simulated  $-\log(\mathbb{E}(|\Lambda_T - M|))/(\lambda T)$  for  $S \sim \text{Exp}(1)$ ,  $\alpha = \beta = \lambda = 1$  and  $\mu = 0$  based on 250,000 simulated realizations of a linear birth and death process

a high computing capacity, which makes our theoretical results even more valuable. Note further that the simulations for exponentially distributed social indices are computationally considerably more costly than in the case where the social indices are deterministic, in particular because the expression for  $\Lambda_T$  given by (6.5) can be simplified in the latter case such that we do not need to simulate all ages (or birth times). As a consequence, we used a relatively small number of iterations for Figure 6.2, namely 250,000.

In the following, we examine which step of our derivation of our convergence rate leads to the different order. We bounded the total variation distance from above by  $\mathbb{E}(|\Lambda_T - M|)$ . In Figure 6.3, we consider the quantity  $-\log(\mathbb{E}(|\Lambda_T - M|))/T$ , which is approximately of the order  $\lambda$  if the rate that is given by  $\mathbb{E}(|\Lambda_T - M|)$  is sharp. However, Figure 6.3 suggests that  $-\log(\mathbb{E}(|\Lambda_T - M|))/(\lambda T)$  approaches a value close to  $1/2$ , i.e. the corresponding rate is close to our rate but not to the actual one.

Theorem 3.4.1 implies that we also could have used  $\mathbb{E}(|\sqrt{\Lambda_T} - \sqrt{M}|)$  instead of  $\mathbb{E}(|\Lambda_T - M|)$  as upper bound for the total variation distance  $d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M))$ . However, Figure 6.4 suggests that we would have obtained the same rate. Note that it is also plausible from a theoretical point of view that  $\mathbb{E}(|\sqrt{\Lambda_T} - \sqrt{M}|)$  does not give us a better rate since we have that  $\mathbb{E}(|\Lambda_T - M|)$  is equal to

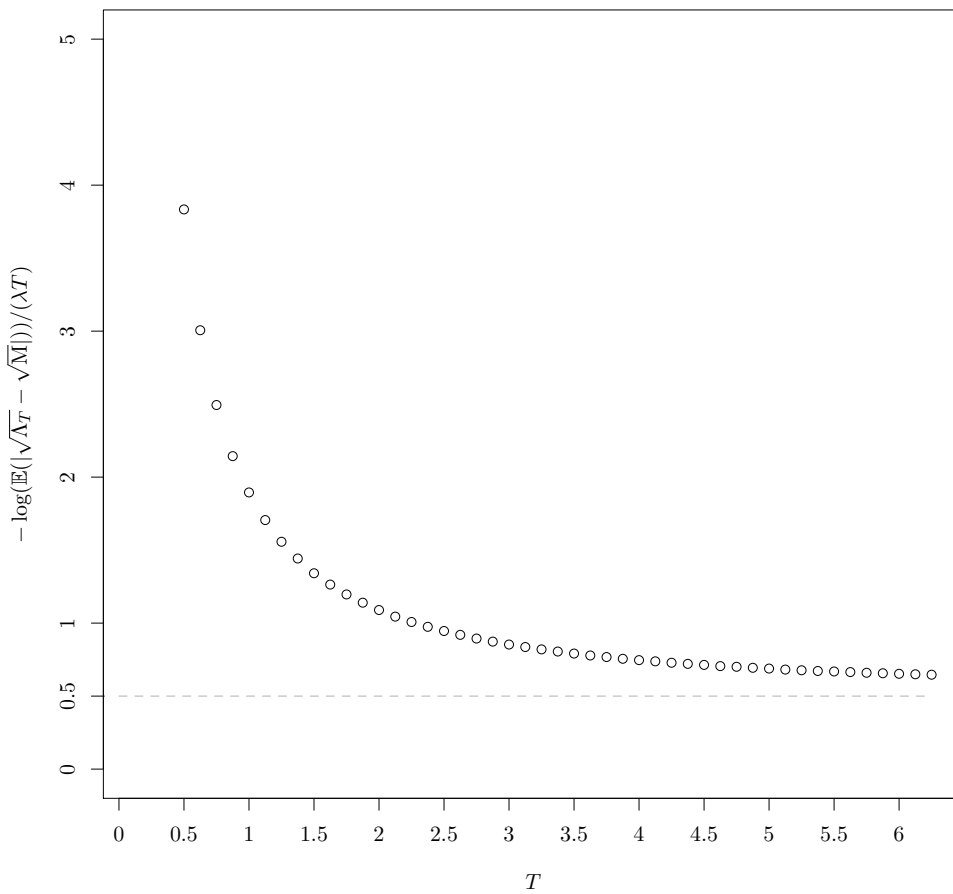


Figure 6.4: Estimated  $-\log(\mathbb{E}(|\sqrt{\Lambda_T} - \sqrt{M}|))/(\lambda T)$  for  $S \sim \text{Exp}(1)$ ,  $\alpha = \beta = \lambda = 1$  and  $\mu = 0$  based on 250,000 simulated realizations of a linear birth and death process

$\mathbb{E}(|\sqrt{\Lambda_T} - \sqrt{M}| \cdot (\sqrt{\Lambda_T} + \sqrt{M}))$ . Note further that our proof would have become more difficult if we used  $\mathbb{E}(|\sqrt{\Lambda_T} - \sqrt{M}|)$  instead of  $\mathbb{E}(|\Lambda_T - M|)$  as upper bound.

In order to examine further how our rate could be improved theoretically, we also consider the minimal coupling of  $\Lambda_T$  and  $M$ , i.e. the expression

$$\inf_{\substack{\tilde{\Lambda}: \tilde{\Lambda} \stackrel{\mathcal{D}}{=} \Lambda \\ \tilde{M}: \tilde{M} \stackrel{\mathcal{D}}{=} M}} \mathbb{E}(|\tilde{\Lambda}_T - \tilde{M}|). \tag{6.79}$$

Obviously, it is smaller than or equal to  $\mathbb{E}(|\Lambda_T - M|)$ , and, by Remark 3.4.2, it is an upper bound for  $d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M))$ . In order to be able to simulate the quantity (6.79) more effectively, we use the following theoretical background. By Example 3.1.3, (6.79) is the Wasserstein distance between  $\mathcal{L}(\Lambda_T)$  and  $\mathcal{L}(M)$ . We also know from Example 3.1.3 that this Wasserstein distance can be expressed as

$$\int_{-\infty}^{\infty} |\mathbb{P}(\Lambda_T \leq x) - \mathbb{P}(M \leq x)| dx.$$

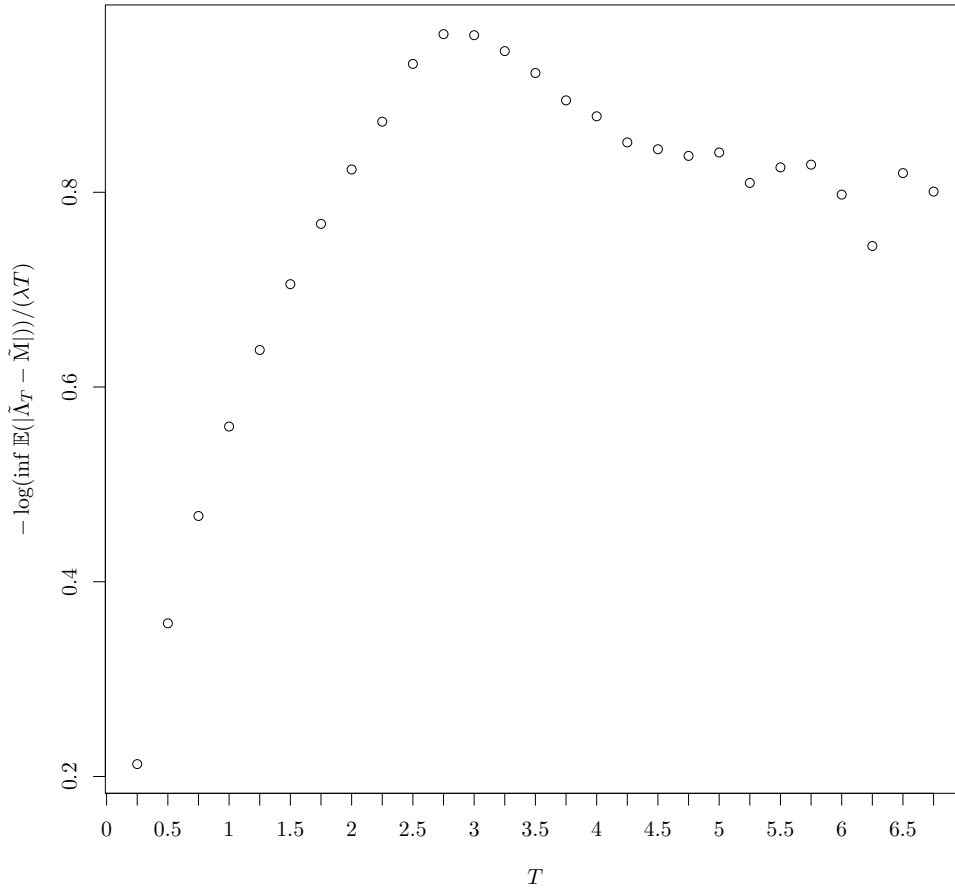


Figure 6.5: Simulated  $-\log(\inf \mathbb{E}(|\tilde{\Lambda}_T - \tilde{M}|)) / (\lambda T)$  for  $S \sim \text{Exp}(1)$ ,  $\alpha = \beta = \lambda = 1$  and  $\mu = 0$  based on 500,000 simulated realizations of a linear birth and death process, where we take the infimum over all random variables  $\tilde{\Lambda}$  and  $\tilde{M}$  with  $\tilde{\Lambda} \stackrel{\mathcal{D}}{=} \Lambda$  and  $\tilde{M} \stackrel{\mathcal{D}}{=} M$

In order to simulate this expression, we use the corresponding empirical distribution functions. Note that we need a large number of iterations in order to approximate the integral in a reasonable way.

Figure 6.5 suggests that the rate corresponding to (6.79) is smaller than one, i.e. that this expression decreases slower than the total variation distance. This indicates that the application of our mixed Poisson approximation result does not lead to a sharp upper bound here. However, since the class of mixed Poisson distributions is very large, it is difficult to find a better universal result, and even in our special case, we do not expect to find a more precise approximation since the mixing distribution of the  $\text{MixPo}(\Lambda_T)$  is very complex.

Note that Figure 6.5, which is based on 500,000 simulations of a linear birth and death process, is not very satisfying since it does not allow to draw a clear conclusion about the rate. Unfortunately, we cannot obtain a very convincing results here due to computational limitations.

We have reasons to doubt that our coupling is the optimal one since the situation in Figure 6.3 and Figure 6.5 looks different. However, our coupling is a very natural one, which allows us to derive an explicit upper bound.

### 6.3.23 Remark

In the pure birth case, we can simulate the linear birth and death process efficiently by using the distribution of the ages from Section 4.2. If nodes can die, we also need to consider the death times, which makes the simulations very time-consuming. Therefore, we do not examine this case numerically here. However, we note that a positive death rate  $\mu$  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 well be larger than the one stated in Theorem 6.3.4.

## 6.4 Negligibility of multiple edges

In this section, we show that multiple edges are negligible (Lemma 6.4.4 below) and use this result to prove Corollary 1.0.3 from the introduction, which states our main result for the case where multiple edges are ignored.

First we consider the social index of a node that is connected to the node  $J_T$  by an incoming edge.

### 6.4.1 Lemma

Given  $Y_T = n$  for  $n \in \mathbb{N} \setminus \{1\}$ , the cumulative distribution function  $F_{\tilde{S}}$  of the social index  $\tilde{S}$  of a node  $i_1$  that is connected to  $J_T$  at time  $T$  by an edge that was created by  $i_1$  is given by

$$F_{\tilde{S}}(s) = \int_0^s \mathbb{E} \left( \frac{s_1}{\frac{s_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right) \mathbb{P}^S(ds_1) = \mathbb{E} \left( \frac{S_1 \mathbb{1}_{[0,s]}(S_1)}{\frac{S_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right)$$

for  $s \geq 0$ .

*Proof:* We condition on  $Y_T = n$  for  $n \in \mathbb{N} \setminus \{1\}$ . Note that the conditional probability that  $s_1$  is the social index of a node connected to  $J_T$  at time  $T$  by an incoming edge given  $S_1 = s_1, \dots, S_n = s_n$  and  $J_T = j_T \in \{2, \dots, n\}$  is  $s_1 (\sum_{i \neq j_T} s_i)^{-1}$ . Since the social indices are identically distributed, we thus obtain by Bayes' Theorem

$$\begin{aligned} F_{\tilde{S}}(s) &= \left( \int_{[0,s] \times [0,\infty)^{n-2}} \frac{s_1}{\sum_{i=1}^{n-1} s_i} \mathbb{P}^S(ds_1) \dots \mathbb{P}^S(ds_{n-1}) \right) \left( \int_{[0,\infty)^{n-1}} \frac{s_1}{\sum_{i=1}^{n-1} s_i} \mathbb{P}^S(ds_1) \dots \mathbb{P}^S(ds_{n-1}) \right)^{-1} \\ &= \left( \int_0^s \mathbb{E} \left( \frac{S_1}{\sum_{i=1}^{n-1} S_i} \middle| S_1 = s_1 \right) \mathbb{P}^S(ds_1) \right) \left( \mathbb{E} \left( \frac{S_1}{\sum_{i=1}^{n-1} S_i} \right) \right)^{-1}. \end{aligned} \quad (6.80)$$

Since we have

$$1 = \mathbb{E} \left( \sum_{j=1}^{n-1} \frac{S_j}{\sum_{i=1}^{n-1} S_i} \right) = \sum_{j=1}^{n-1} \mathbb{E} \left( \frac{S_j}{\sum_{i=1}^{n-1} S_i} \right) = (n-1) \mathbb{E} \left( \frac{S_1}{\sum_{i=1}^{n-1} S_i} \right),$$

the right-hand side of (6.80) is equal to

$$\int_0^s \mathbb{E} \left( (n-1) \frac{S_1}{\sum_{i=1}^{n-1} S_i} \middle| S_1 = s_1 \right) \mathbb{P}^S(ds_1) = \int_0^s \mathbb{E} \left( \frac{s_1}{\frac{s_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right) \mathbb{P}^S(ds_1).$$

□

### 6.4.2 Remark

Since

$$\frac{S_1 \mathbb{1}_{[0,s]}(S_1)}{\frac{S_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \leq 1$$

almost surely and  $\frac{S_1}{n-1} \rightarrow 0$  and  $\frac{1}{n-1} \sum_{i=2}^{n-1} S_i \rightarrow \mathbb{E}(S)$  almost surely by the strong law of large number as  $n \rightarrow \infty$ , Lemma 6.4.1 and Fubini's Theorem imply that the cumulative distribution function  $F_{\tilde{S}}$  from Lemma 6.4.1 converges to

$$\frac{\mathbb{E}(S \mathbb{1}_{[0,s]}(S))}{\mathbb{E}(S)}$$

as  $n \rightarrow \infty$ . Note that this is just the cumulative distribution function of the so-called size-biased distribution by its definition (see e.g. (1.2.3) in [vdH16]). Thus the distribution of the social index of a node  $i_1$  that is connected to  $J_T$  at time  $T$  by an edge that was created by  $i_1$  converges weakly to the size-biased distribution of the social index distribution. Since the size-biased distribution dominates the original one stochastically (see e.g. Section 2.3 of [vdH16]), we may conclude that the neighbours that are connected by incoming edges have asymptotically in average a higher social index than  $\mathbb{E}(S)$ .<sup>7</sup>

Since the distribution of the social index of a neighbour that is connected by an *outgoing* edge is obviously just the social index distribution, the distribution of the social index of a neighbour picked uniformly at random is a mixture of the distribution given by the cumulative distribution function  $F_{S_{i_1}}$  from Lemma 6.4.1 and the social index distribution. We may conclude that this distribution converges weakly to a mixture of the size biased distribution of the social index distribution and the social index distribution, which is a result that was already stated in Subsection 3.3 of [BLT11].

### 6.4.3 Corollary

The expected value of the social index  $\tilde{S}$  of a node  $i_1$  that is connected to  $J_T$  at time  $T$  by an edge that was created by  $i_1$  is bounded from above by

$$\frac{2\sigma_S^2}{c^2 \mathbb{E}(S)} + \frac{\mathbb{E}(S^2)}{\mathbb{E}(S)(1-c)}$$

for any  $c \in (0, 1)$ .

*Proof:* We condition on  $Y_T = n \in \mathbb{N} \setminus \{1\}$ . By Lemma 6.4.1, the conditional expected value of the social index  $\tilde{S}$  of a node  $i_1$  that is connected to  $J_T$  at time  $T$  by an edge that was created by  $i_1$  is then equal to

$$\int_0^\infty s_1 \mathbb{E} \left( \frac{s_1}{\frac{s_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right) \mathbb{P}^S(ds_1) = \mathbb{E} \left( \frac{S_1^2}{\frac{S_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right),$$

and we have

$$\mathbb{E} \left( \frac{S_1^2}{\frac{S_1}{n-1} + \frac{1}{n-1} \sum_{i=2}^{n-1} S_i} \right) \leq \mathbb{E} \left( \mathbb{1}_{\{\frac{1}{n-2} \sum_{i=2}^{n-1} S_i \leq \mathbb{E}(S)(1-c)\}} (n-1) S_1 \right)$$

<sup>7</sup> This result is similar to the result that neighbours have in average a higher degree than a node picked uniformly at random (see e.g. Theorem 1.2 in [vdH16]).



$$\begin{aligned}
& + \mathbb{E} \left( \mathbb{1}_{\left\{ \frac{1}{n-2} \sum_{i=2}^{n-1} S_i > \mathbb{E}(S)(1-c) \right\}} \frac{(n-1)S_1^2}{(n-2)\mathbb{E}(S)(1-c)} \right) \\
& \leq \mathbb{P} \left( \frac{1}{n-2} \sum_{i=2}^{n-1} S_i \leq \mathbb{E}(S)(1-c) \right) (n-1)\mathbb{E}(S_1) + \frac{2\mathbb{E}(S_1^2)}{\mathbb{E}(S)(1-c)},
\end{aligned}$$

where we use the convention  $\frac{0}{0} := 0$ . By Chebyshev's inequality, we have

$$\begin{aligned}
\mathbb{P} \left( \frac{1}{n-2} \sum_{i=2}^{n-1} S_i \leq \mathbb{E}(S)(1-c) \right) & \leq \mathbb{P} \left( \left| \frac{1}{n-2} \sum_{i=2}^{n-1} S_i - \mathbb{E}(S) \right| \geq c\mathbb{E}(S) \right) \leq \frac{1}{(c\mathbb{E}(S))^2} \text{Var} \left( \frac{1}{n-2} \sum_{i=2}^{n-1} S_i \right) \\
& \leq \frac{1}{n-2} \frac{\sigma_S^2}{(c\mathbb{E}(S))^2},
\end{aligned}$$

which yields the desired result.  $\square$

The following lemma states that multiple edges are negligible.

#### 6.4.4 Lemma

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

*Proof.* Let  $D_T$  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  $\rho_1 < \dots < \rho_{D_T}$  be the birth times of these edges. Condition on  $J_T, D_T, \rho_1 < \dots < \rho_{D_T}$  and  $(Y_t)_{0 \leq t \leq T}$ . Let  $B_k$  be the event that  $J_T$  creates an outgoing edge at time  $\rho_k$  that is a multiple edge up to time  $T$ . Then the (conditional) probability of  $B_k$  is smaller than or equal to

$$\frac{D_T - 1}{Y_{\rho_k} - 1}.$$

Note that  $\bigcup_{k=1}^{D_T} B_k$  is the event that  $J_T$  has at least one outgoing edge that is a multiple edge at time  $T$ . By subadditivity, we have

$$\begin{aligned}
& \mathbb{P} \left( \bigcup_{k=1}^{D_T} B_k \mid J_T, D_T, \rho_1, \dots, \rho_{D_T}, (Y_t)_{0 \leq t \leq T}, Y_T > 0 \right) \\
& \leq \min \left( \sum_{k=1}^{D_T} \mathbb{P}(B_k \mid J_T, D_T, \rho_1, \dots, \rho_{D_T}, (Y_t)_{0 \leq t \leq T}, Y_T > 0), 1 \right) \\
& \leq \min \left( \sum_{k=1}^{D_T} \frac{D_T - 1}{Y_{\rho_k} - 1}, 1 \right) \\
& \leq \min \left( \frac{D_T^2}{\min_{T - A_{J_T}(T) \leq t \leq T} Y_t - 1}, 1 \right).
\end{aligned}$$

Taking the expectation, we obtain

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{k=1}^{D_T} B_k \mid Y_T > 0\right) &\leq \mathbb{E}\left(\min\left(\frac{D_T^2}{\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1}, 1\right) \mid Y_T > 0\right) \\
&\leq \mathbb{P}(D_T \geq e^{\frac{1}{12}(\lambda-\mu)T} \mid Y_T > 0) + \mathbb{P}\left(\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1 \leq e^{\frac{1}{3}(\lambda-\mu)T} \mid Y_T > 0\right) \\
&\quad + \mathbb{E}\left(\min\left(\frac{e^{\frac{1}{6}(\lambda-\mu)T}}{e^{\frac{1}{3}(\lambda-\mu)T}}, 1\right) \mid Y_T > 0\right). \tag{6.81}
\end{aligned}$$

Writing  $D_\infty$  for a random variable having the asymptotic degree distribution  $\text{MixPo}(M^*)$  with  $M^*$  defined at the beginning of Subsection 6.3.2, we obtain by conditioning on  $M^*$  that the second moment  $\mathbb{E}(D_\infty^2)$  is equal to

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

Thus Theorem 6.3.4 and the Markov inequality imply

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

For the second summand of the right-hand side of (6.81), we obtain

$$\begin{aligned}
\mathbb{P}\left(\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1 \leq e^{\frac{1}{3}(\lambda-\mu)T} \mid Y_T > 0\right) &\leq \mathbb{P}\left(\max_{\frac{T}{2} \leq t \leq T} \frac{1}{Y_t} > \frac{1}{e^{\frac{1}{3}(\lambda-\mu)T} + 1} \mid Y_T > 0\right) \\
&\quad + \mathbb{P}\left(A_{J_T}(T) > \frac{T}{2} \mid Y_T > 0\right). \tag{6.82}
\end{aligned}$$

By Lemma 6.3.12, the second summand of the right-hand side is of the order  $O(e^{-\frac{1}{2}(\lambda-\mu)T})$  as  $T \rightarrow \infty$ . With  $Y_\infty = \lim_{t \rightarrow \infty} Y_t$ , we have

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

Thus the first summand of the right-hand side of (6.82) is smaller than or equal to

$$\mathbb{P}\left(\max_{T/2 \leq t \leq T} \frac{1}{Y_t} > \frac{1}{e^{\frac{1}{3}(\lambda-\mu)T} + 1} \mid Y_\infty > 0\right) + \mathbb{P}(Y_\infty = 0 \mid Y_T > 0). \tag{6.83}$$

The second summand is smaller than or equal to  $\frac{\mu}{\lambda}e^{-(\lambda-\mu)T}$  by Lemma 6.3.5. By Corollary 6.3.9 and the inequality (2) in Theorem 6.14 on page 99 in [Yeh95], the first summand of (6.83) is bounded from above by

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

where the first equality follows from Lemma 6.3.16.

We may conclude that the probability that  $J_T$  has at least one outgoing edge that is a multiple edge at time  $T$  is of the order  $O(Te^{-\frac{1}{6}(\lambda-\mu)T})$

Now we consider incoming edges. By conditioning on  $J_T$ ,  $(S_i)_{i \in \mathbb{N}}$  and  $(Y_t)_{0 \leq t \leq T}$ , we obtain that the probability for the event  $\tilde{B}_i^{(1)}$  that node  $i$  creates an edge that connects  $i$  to  $J_T$  at time  $T$  is smaller than or equal to

$$\begin{aligned} \mathbb{E}\left(1 - \exp\left(-\alpha S_i A_{J_T} \frac{1}{\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1}\right) \mid Y_T > 0\right) &\leq \mathbb{E}\left(\alpha S_i A_{J_T} \frac{1}{\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1} \mid Y_T > 0\right) \\ &\leq \mathbb{P}\left(\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1 \leq e^{\frac{1}{3}(\lambda-\mu)T} \mid Y_T > 0\right) + \mathbb{P}\left(A_{J_T} > \frac{T}{2} \mid Y_T > 0\right) + \alpha \mathbb{E}(S) \frac{T}{2} e^{-\frac{1}{3}(\lambda-\mu)T}. \end{aligned}$$

Since we showed above that the first two summands are of the order  $O(Te^{-\frac{1}{6}(\lambda-\mu)T})$ , the right-hand side is of the order  $O(Te^{-\frac{1}{3}(\lambda-\mu)T})$ .

We condition on  $\tilde{B}_i^{(1)}$  now and denote the birth time of the edge  $(i, J_T)$  corresponding to  $\tilde{B}_i^{(1)}$  by  $\eta_i$ . Then we have for the conditional probability of the event  $\tilde{B}_i^{(2)}$  that  $i$  creates another edge  $(i, J_T)$  in the time interval  $(\eta_i, T] \subset (T - A_{J_T}(T), T]$  that survives up to time  $T$

$$\begin{aligned} \mathbb{P}(\tilde{B}_i^{(2)} \mid \tilde{B}_i^{(1)}, Y_T > 0) &\leq \mathbb{E}\left(1 - \exp\left(-\alpha \tilde{S}_i A_{J_T} \frac{1}{\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1}\right) \mid \tilde{B}_i^{(1)}, Y_T > 0\right) \\ &\leq \mathbb{E}\left(\alpha \tilde{S}_i A_{J_T} \frac{1}{\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1} \mid Y_T > 0\right) \\ &\leq \mathbb{P}\left(\min_{T-A_{J_T}(T) \leq t \leq T} Y_t - 1 \leq e^{\frac{1}{3}(\lambda-\mu)T} \mid Y_T > 0\right) + \mathbb{P}\left(A_{J_T} > \frac{T}{2} \mid Y_T > 0\right) + \alpha \mathbb{E}(\tilde{S}) \frac{T}{2} e^{-\frac{1}{3}(\lambda-\mu)T}, \end{aligned}$$

where  $\tilde{S}$  denotes the social index of a node connected to  $J_T$  at time  $T$  by an incoming edge. By Corollary 6.4.3, the right-hand side is of the order  $O(Te^{-\frac{1}{3}(\lambda-\mu)T})$ .

For simplicity, we denote the  $Y_T$  nodes alive at time  $T$  by  $1, \dots, Y_T$  now. For the probability that  $J_T$  has at least two incoming edges from the same node at time  $T$ , we then obtain by subadditivity

$$\begin{aligned} \mathbb{P}\left(\bigcup_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \tilde{B}_i^{(1)} \cap \tilde{B}_i^{(2)} \mid Y_T > 0\right) &\leq \mathbb{E}\left(\min\left(\sum_{i=1}^{Y_T-1} \mathbb{P}(\tilde{B}_i^{(1)} \cap \tilde{B}_i^{(2)} \mid (Y_t)_{0 \leq t \leq T}), 1\right) \mid Y_T > 0\right) \\ &= \mathbb{E}(\min((Y_T - 1)\mathbb{P}(\tilde{B}_1^{(1)} \cap \tilde{B}_1^{(2)} \mid (Y_t)_{0 \leq t \leq T}), 1) \mid Y_T > 0) \\ &\leq e^{\frac{7}{6}(\lambda-\mu)T} \mathbb{P}(\tilde{B}_1^{(1)} \cap \tilde{B}_1^{(2)} \mid Y_T > 0) + \mathbb{P}(Y_T - 1 > e^{\frac{7}{6}(\lambda-\mu)T} \mid Y_T > 0) \end{aligned}$$

By the Markov inequality, the second summand of the right-hand side is of the order  $O(e^{-\frac{1}{6}(\lambda-\mu)T})$ . For the first summand, we have

$$\begin{aligned} e^{\frac{7}{6}(\lambda-\mu)T} \mathbb{P}(\tilde{B}_1^{(1)} \cap \tilde{B}_1^{(2)} | Y_T > 0) &= e^{\frac{7}{6}(\lambda-\mu)T} \mathbb{P}(\tilde{B}_1^{(1)} | Y_T > 0) \mathbb{P}(\tilde{B}_1^{(2)} | \tilde{B}_1^{(1)}, Y_T > 0) \\ &= e^{\frac{7}{6}(\lambda-\mu)T} O(Te^{-\frac{1}{3}(\lambda-\mu)T}) O(Te^{-\frac{1}{3}(\lambda-\mu)T}) = O(T^2 e^{-\frac{1}{6}(\lambda-\mu)T}). \end{aligned}$$

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

### Proof of Corollary 1.0.3

Recall that  $\tilde{\nu}_t$  denotes the distribution of the number of neighbours,  $\nu_t$  the degree distribution at time  $t$  and  $\nu$  the asymptotic degree distribution. Lemma 6.4.4 implies that  $d_{TV}(\tilde{\nu}_t, \nu_t) = O(t^2 e^{-\frac{1}{6}(\lambda-\mu)t})$  as  $t \rightarrow \infty$ . From Theorem 1.0.2, we know that  $d_{TV}(\nu_t, \nu) = O(t^2 e^{-\frac{1}{6}(\lambda-\mu)t})$  as  $t \rightarrow \infty$ . Thus the triangle inequality yields the desired result.  $\square$

## 6.5 Heavy tails and power laws of the asymptotic degree distribution

Recall that the asymptotic degree distribution is

$$\text{MixPo}\left(\frac{\alpha}{\beta + \mu}(S + \mathbb{E}(S))(1 - e^{-(\beta+\mu)A})\right),$$

where  $A \sim \text{Exp}(\lambda)$ ,  $S$  has the social index distribution, and  $A$  and  $S$  are independent. In this section, we show that this asymptotic degree distribution can be heavy-tailed (Subsection 6.5.1) and exhibit power law behaviour (Subsection 6.5.2). In order to do so, we choose a Pareto distribution, i.e. a power law distribution (cf. Remark 5.2.4), as distribution of the social indices. This choice does not only lead to the desired results, but can also “be naturally justified by arguing that power laws appear rather generically in many contexts when one ranks, for example, people according to their incomes or cities according to their population, etc.” ([CCRM02]).

### 6.5.1 Heavy tails of the asymptotic degree distribution

Note that the random variable  $Z := 1 - e^{-(\beta+\mu)A}$  is  $\text{Beta}(1, \frac{\lambda}{\beta+\mu})$  distributed (see e.g. Subsection 3.3 of [BLT11]). For the social index, we choose the  $\text{Pareto}(\theta - 1, b)$  distribution, i.e. the distribution with Lebesgue density  $\frac{(\theta-1)b^{\theta-1}}{x^\theta} \mathbb{1}_{\{x \geq b\}}$ , where  $b > 0$  and  $\theta > 3$ , and show that the mixed Poisson distribution has heavy tails. Note that we chose the parameter  $\theta$  such that the first two moments of  $S$  exist.

In order to do so, we firstly apply the general Theorem 3.3.7 in order to show that  $\text{MixPo}(\frac{\alpha}{\beta+\mu}SZ)$  and the distribution of  $\frac{\alpha}{\beta+\mu}SZ$  are tail equivalent (cf. Definition 3.3.4). Secondly, we show that the distribution of  $\frac{\alpha}{\beta+\mu}SZ$  has power law tails. Finally, we conclude that the asymptotic degree distribution  $\text{MixPo}(\frac{\alpha}{\beta+\mu}(S + \mathbb{E}(S))Z)$  has heavy tails.

Theorem 3.3.7 states that the mixing distribution and the corresponding mixed Poisson distribution are tail equivalent under some assumptions. These assumptions involve the density and the survival function of the mixing distribution, which we derive now for the mixing distribution of  $\text{MixPo}(\frac{\alpha}{\beta+\mu}SZ)$ .

Firstly, we condition on  $Z = z \in [0, 1]$ . Let  $\eta = \frac{\alpha}{\beta+\mu}$ . We obtain for the conditional survival function  $1 - F_{\eta SZ|Z=z}$  of  $\eta SZ$

$$1 - F_{\eta SZ|Z=z}(x) = \frac{b^{\theta-1}\eta^{\theta-1}z^{\theta-1}}{x^{\theta-1}}$$

for  $x \geq \eta bz$  since  $S$  is  $\text{Pareto}(\theta - 1, b)$  distributed. By differentiation of  $F_{\eta SZ|Z=z}$ , we obtain that the conditional density is given by

$$f_{\eta SZ|Z=z}(x) = \frac{(\theta - 1)b^{\theta-1}\eta^{\theta-1}z^{\theta-1}}{x^\theta} \quad (6.84)$$

for  $x \geq \eta bz$ .

Using the density of the beta distribution, Equation (6.84) yields

$$f_{\eta SZ}(x) = \int_0^1 \frac{1}{\text{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} \frac{(\theta-1)b^{\theta-1}\eta^{\theta-1}z^{\theta-1}}{x^\theta} \mathbb{1}_{\{z \leq \frac{x}{\eta b}\}} dz,$$

where  $f_{\eta SZ}$  is the unconditional density of  $\eta SZ$  and  $\text{B}$  denotes the beta function. Thus for  $x > \eta b$ , we have

$$f_{\eta SZ}(x) = \frac{1}{x^\theta} \int_0^1 \frac{1}{\text{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} (\theta-1)b^{\theta-1}\eta^{\theta-1}z^{\theta-1} dz. \quad (6.85)$$

Analogously, we obtain for the unconditional survival function

$$1 - F_{\eta SZ}(x) = \int_0^1 \frac{1}{\text{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} \frac{b^{\theta-1}\eta^{\theta-1}z^{\theta-1}}{x^{\theta-1}} \mathbb{1}_{\{z \leq \frac{x}{\eta b}\}} dz,$$

and for  $x > \eta b$ , we have

$$1 - F_{\eta SZ}(x) = \frac{1}{x^{\theta-1}} \int_0^1 \frac{1}{\text{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} b^{\theta-1}\eta^{\theta-1}z^{\theta-1} dz. \quad (6.86)$$

### 6.5.1 Remark

Since the integral (6.86) does not depend on  $x$ , it is clear that the prerequisite of Proposition 3.3.3 is fulfilled. Thus we obtain that there exist constants  $\tilde{c}_1 < \tilde{c}_2$  such that

$$\tilde{c}_1 m^{1-\theta} \leq 1 - G(m) \leq \tilde{c}_2 m^{1-\theta}, \quad (6.87)$$

where  $G$  is the cumulative distribution function of the  $\text{MixPo}(\eta SZ)$  distribution and  $m \in \mathbb{N}_0$ . In particular, (6.87) implies that the  $\text{MixPo}(\eta SZ)$  distribution is heavy-tailed. However, it is not difficult either to obtain this result by applying Theorem 3.3.7, which we do now.

Equations (6.85) and (6.86) imply

$$\lim_{x \rightarrow \infty} \frac{x f_{\eta SZ}(x)}{F_{\eta SZ}(x)} = \frac{C_1}{C_2} = c > 0, \quad (6.88)$$

where

$$C_1 = \int_0^1 \frac{1}{\mathbf{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} (\theta-1) b^{\theta-1} \eta^{\theta-1} z^{\theta-1} dz \in (0, \infty)$$

and

$$C_2 = \int_0^1 \frac{1}{\mathbf{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} b^{\theta-1} \eta^{\theta-1} z^{\theta-1} dz \in (0, \infty).$$

Equation (6.88) is just the first von Mises condition, which yields that the distribution of  $\eta SZ$  and  $\text{MixPo}(\eta SZ)$  are tail equivalent by Theorem 3.3.7, i.e. that

$$\lim_{x \rightarrow \infty} \frac{1 - F_{\eta SZ}(x)}{1 - G(x)} = 1, \quad (6.89)$$

where  $G$  is the cumulative distribution function of  $\text{MixPo}(\eta SZ)$ .

Note that (6.86) yields that the distribution of  $\eta SZ$  is heavy-tailed if we choose  $\text{Pareto}(\theta-1, b)$  with  $b > 0$  and  $\theta > 3$  as distribution of the social indices. Thus the tail equivalence stated in (6.89) implies that the  $\text{MixPo}(\eta SZ)$  distribution has heavy tails. Since the  $\text{MixPo}(\eta SZ)$  distribution is stochastically dominated by the asymptotic degree distribution  $\text{MixPo}(\eta(S + \mathbb{E}(S))Z)$ , this yields that the latter also has heavy tails.

We would also like to show that the asymptotic degree distribution even exhibits power law behaviour (in the weak sense). In order to do so, we use another approach in the following subsection.

### 6.5.2 Power law behaviour of the asymptotic degree distribution

We use the same assumptions as before. However, we do not consider the distribution of  $\eta SZ$  first, but directly the mixing distribution, i.e. the distribution of  $\eta(S + \mathbb{E}(S))Z$ , which leads to the following theorem.

#### 6.5.2 Theorem

Let the social indices be distributed according to the  $\text{Pareto}(\theta-1, b)$  distribution with  $b > 0$  and  $\theta > 3$ . Then the asymptotic degree distribution in the Britton-Lindholm model obeys a power law with exponent  $\theta$  in the weak sense.

*Proof:* Analogously to the derivation of  $f_{\eta SZ}$ , we obtain for the density  $f_{\eta(S+\mathbb{E}(S))Z}$  of  $\eta(S + \mathbb{E}(S))Z$

$$f_{\eta(S+\mathbb{E}(S))Z}(x) = \int_0^1 \frac{1}{\mathbf{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} \frac{(\theta-1) b^{\theta-1} \eta^{\theta-1} z^{\theta-1}}{(x - \eta \mathbb{E}(S) z)^\theta} \mathbb{1}_{\{z \leq \frac{x}{\eta(b+\mathbb{E}(S))}\}} dz.$$

Let  $x \geq \eta b + \eta \mathbb{E}(S)$ . Then we obtain

$$f_{\eta(S+\mathbb{E}(S))Z}(x) = \int_0^1 \frac{1}{\mathbf{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} \frac{(\theta-1) b^{\theta-1} \eta^{\theta-1} z^{\theta-1}}{(x - \eta \mathbb{E}(S) z)^\theta} dz. \quad (6.90)$$

On the one hand, the right-hand side of (6.90) is larger than or equal to

$$\int_0^1 \frac{1}{\mathbb{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} (\theta-1) b^{\theta-1} \eta^{\theta-1} z^{\theta-1} dz \frac{1}{x^\theta} = C_1 \frac{1}{x^\theta}. \quad (6.91)$$

On the other hand, since  $\eta \mathbb{E}(S)z \leq \eta \mathbb{E}(S)$  for all  $z \in [0, 1]$ , we have that the right-hand side of (6.90) is smaller than or equal to

$$\int_0^1 \frac{1}{\mathbb{B}(1, \frac{\lambda}{\beta+\mu})} (1-z)^{\frac{\lambda}{\beta+\mu}-1} (\theta-1) b^{\theta-1} \eta^{\theta-1} z^{\theta-1} dz \frac{1}{(x - \eta \mathbb{E}(S))^\theta} = C_1 \frac{1}{(x - \eta \mathbb{E}(S))^\theta}. \quad (6.92)$$

Let  $D_\infty \sim \text{MixPo}(\eta(S + \mathbb{E}(S))Z)$  and the function  $g$  be defined by  $g(x) = C_1 x^{-\theta}$ . Then (6.91) and (6.92) imply that  $f_{\eta(S+\mathbb{E}(S))Z} \sim g$ , where the operator  $\sim$  is defined in Definition 3.3.4. Thus Theorem 3.3.8 yields  $\mathbb{P}(D_\infty = k) \sim C_1 k^{-\theta}$ , i.e.

$$\lim_{k \rightarrow \infty} k^\theta \mathbb{P}(D_\infty = k) = C_1.$$

This implies that for all  $\tilde{\varepsilon} > 0$ , we can find a  $k_0 \in \mathbb{N}$  such that for all  $k > k_0$

$$(1 - \tilde{\varepsilon}) \frac{C_1}{k^\theta} \leq \mathbb{P}(D_\infty = k) \leq (1 + \tilde{\varepsilon}) \frac{C_1}{k^\theta}.$$

Thus we have for the cumulative distribution function  $F$  of  $D_\infty$  and all  $m \geq k_0$

$$\begin{aligned} 1 - F(m) &= \sum_{k=m+1}^{\infty} \mathbb{P}(D_\infty = k) \leq (1 + \tilde{\varepsilon}) C_1 \sum_{k=m+1}^{\infty} \frac{1}{k^\theta} \leq (1 + \tilde{\varepsilon}) C_1 \int_{m+1}^{\infty} (x-1)^{-\theta} dx \\ &= (1 + \tilde{\varepsilon}) C_1 \int_m^{\infty} x^{-\theta} dx = (1 + \tilde{\varepsilon}) C_1 \frac{1}{\theta-1} m^{1-\theta} \end{aligned}$$

and on the other hand

$$\begin{aligned} 1 - F(m) &= \sum_{k=m+1}^{\infty} \mathbb{P}(D_\infty = k) \geq (1 - \tilde{\varepsilon}) C_1 \sum_{k=m+1}^{\infty} \frac{1}{k^\theta} \geq (1 - \tilde{\varepsilon}) C_1 \int_{m+1}^{\infty} x^{-\theta} dx \\ &= (1 - \tilde{\varepsilon}) C_1 \frac{1}{\theta-1} (m+1)^{1-\theta}. \end{aligned}$$

Letting  $\tilde{\varepsilon} \rightarrow 0$  yields

$$\lim_{m \rightarrow \infty} \frac{1 - F(m)}{m^{1-\theta}} = \frac{C_1}{\theta-1}. \quad (6.93)$$

Let

$$\varphi(m) := \frac{1 - F(m)}{m^{1-\theta}}.$$

Then (6.93) implies

$$\lim_{m \rightarrow \infty} \frac{\varphi(cm)}{\varphi(m)} = \lim_{m \rightarrow \infty} \frac{1 - F(cm)}{(cm)^{1-\theta}} \frac{m^{1-\theta}}{1 - F(m)} = \frac{C_1}{\theta-1} \frac{\theta-1}{C_1} = 1$$

for all  $c > 0$ . Thus  $\varphi$  is a slowly varying function, and we have

$$1 - F(m) = \varphi(m)m^{1-\theta}.$$

Consequently, the asymptotic degree distribution obeys a power law with exponent  $\theta$  in the weak sense.  $\square$

## 6.6 Further discussion of the asymptotic degree distribution

Recall that the asymptotic degree distribution is  $\text{MixPo}(M^*)$ . Note that we have obtained a rigorous proof of this result as a by-product of the proofs of Theorem 6.3.1, Theorem 6.3.2 and Theorem 6.3.4. Inspired by the derivation of the asymptotic degree distribution in the Erdős-Rényi model, we can give the following heuristic, but intuitive argument for the degree distribution being a mixed Poisson distribution.

### 6.6.1 Remark

Let the population size  $Y_T = y_T$  and the age  $a$  as well as the social index  $s$  of the node  $J_T$  picked uniformly at random at time  $T$  be given. Assume that  $T$  and  $y_T$  are large. Then the probability that  $J_T$  is connected to some other node  $i$  is approximately  $f(a, s)/(y_T - 1)$  for some function  $f$  since the probability that a certain node becomes the “second” node of a newly created edge is  $\frac{1}{y_T - 1}$  during the lifetime of  $J_T$  if we ignore that the population size may change between the birth time of  $J_T$  and  $T$ . Thus the degree of  $J_T$  is approximately  $\text{Bin}(y_T - 1, f(a, s)/(y_T - 1))$  distributed, which can be approximated by the  $\text{Po}(f(a, s))$  distribution by the Poisson limit theorem. Lifting the conditioning leads to a mixed Poisson distribution.

Note that the class of asymptotic degree distributions we obtained here is a rich one due to the high variability in the choice of the model parameters (including the distribution of the social indices).

Since we know the asymptotic degree distribution, the model allows us to find a random network for a given (asymptotic) degree distribution. This is to some extent, though not quite as flexible, as in the configuration model. On the other hand, the configuration cannot provide a realistic model for the evolution of the network.

We could now try to model real-world networks where a mixed Poisson distributions of the form  $\text{MixPo}(M^*)$  might be realistic as degree distributions by using the Britton-Lindholm model. If we were satisfied with any network model that has (approximately) the correct asymptotic degree distribution, it would be sufficient to consider the pure birth case only since the class of mixed Poisson distributions that we can obtain in the limit in this case is obviously the same as in the general case. However, the model where nodes can die gives us a more realistic description of the evolution of many real-world networks.

We could estimate the parameters of our model (including the distribution of the social indices) using real-world network data and derive the asymptotic degree distribution for the estimated parameters, which gives a prediction for the degree distribution of the real-world network at large finite times. The approximation error can be bounded using Theorem 6.3.2 and Theorem 6.3.4.



Furthermore, if a lack of data does not allow us to estimate all parameters, we could try to estimate the asymptotic degree distribution, which might give us further information about the parameters. For example, in the pure birth case or if the per-capita death rate  $\mu$  is known in the general case, we can (at least theoretically) identify the random variable  $M$  from the asymptotic degree distribution due to the identifiability of mixed Poisson distributions given by Proposition 3.3.2. Then  $M$  provides us with more information about the parameters.

Finally, we consider a case where it is possible to derive the asymptotic degree distribution more explicitly.

### 6.6.2 Example

Recall that the random variable  $Z := 1 - e^{-(\beta+\mu)A}$  is  $\text{Beta}(1, \frac{\lambda}{\beta+\mu})$  distributed and  $\text{MixPo}(SZ)$  is the asymptotic degree distribution we are interested in. Assume that  $\lambda = \beta + \mu$ . Then  $Z$  is uniformly distributed on  $[0, 1]$  since the  $\text{Beta}(1, 1)$  distribution is the uniform distribution on  $[0, 1]$ . Furthermore, let  $S = s$  be deterministic for some  $s > 0$ . The random variable  $SZ$  is then uniformly distributed on  $[0, s]$ . Thus the asymptotic degree distribution  $\text{MixPo}(SZ)$  is a *Poisson-uniform distribution* (also called Poisson-rectangular distribution) considered e.g. in Section 2.8. of [Grz06]. By (2.39) in [Grz06], the corresponding probability mass function is

$$\frac{1}{s} \left( 1 - e^{-s} \sum_{l=0}^k \frac{s^l}{l!} \right).$$

Note that the Poisson-uniform distribution is applied in linguistics in order to describe the lengths of words (see e.g. [Grz06]).

## 6.7 The pure birth case with immortal edges

Up to now, we only considered the case where the edges die. Now we assume that  $\mu = \beta = 0$ , i.e. both nodes and edges are *immortal*. In order to obtain the degree distribution, we treat the outgoing and incoming edges separately again.

### 6.7.1 The number of outgoing edges in the pure birth case with immortal edges

We can derive the distribution of the number of outgoing edges analogously to the procedure in Subsection 6.2.1. We condition on  $J_T = j_T$ ,  $S_{J_T} = s_{j_T}$ ,  $A_{J_T} = a_{j_T}$  and  $A_2 = a_2$ . Then the number of outgoing edges evolves according to a Poisson process of rate  $\alpha s_{j_T}$  started in zero at time  $T - a_{\max(j_T, 2)}$ . Consequently, the number of outgoing edges at time  $T$  is  $\text{Po}(\alpha s_{j_T} a_{\max(j_T, 2)})$  distributed by Remark 2.1.4. Thus lifting the conditioning leads to the  $\text{MixPo}(\alpha S_{J_T} A_{\max(J_T, 2)})$  distribution. Note that this distribution converges to the

$$\text{MixPo}(\alpha SA)$$

distribution as  $T \rightarrow \infty$ , where  $S$  and  $A$  are defined as before.

### 6.7.2 The number of incoming edges in the pure birth case with immortal edges

We proceed analogously to Subsection 6.2.2. We condition on  $(Y_t)_{0 \leq t \leq T} = (y_t)_{0 \leq t \leq T}$ , the social indices  $(S_k)_{k \in \mathbb{N}} = (s_k)_{k \in \mathbb{N}}$ ,  $J_T = j_T$  and the ages  $T = a_1 > \dots > a_{y_T}$ . Let  $i \in \{1, \dots, y_T\} \setminus \{j_T\}$  be a node, and consider a time interval  $[T - a_l, T - a_{l+1})$ , where  $l \geq i \vee 2$ . The number of edges created by  $i$  that connect  $i$  to  $j_T$  evolves according to a Poisson process of rate  $\alpha s_i \frac{1}{l-1}$ . Thus the total number of edges created by  $i$  in the interval  $[T - a_l, T - a_{l+1})$  that connect  $i$  to  $j_T$  has distribution

$$\text{Po}\left(\frac{\alpha s_i}{l-1}(a_l - a_{l+1})\right).$$

As in Subsection 6.2.2, we use an induction argument in order to obtain that the number of incoming edges has distribution

$$\text{Po}\left(\sum_{\substack{i=1 \\ i \neq j_T}}^{y_T} \frac{\alpha s_i a_{y_T}}{y_T - 1} + \sum_{\substack{i=1 \\ i \neq j_T}}^{y_T} \sum_{l=i \vee j_T}^{y_T-1} \frac{\alpha s_i (a_l - a_{l+1})}{l-1}\right).$$

Thus lifting the conditioning leads to the distribution

$$\text{MixPo}\left(\sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{\alpha S_i A_{Y_T}}{Y_T - 1} + \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{\alpha S_i (A_l - A_{l+1})}{l-1}\right).$$

Note that this distribution converges to

$$\text{MixPo}(\alpha \mathbb{E}(S)A)$$

as  $T \rightarrow \infty$  (cf. Subsection 6.3.1), where  $S$  and  $A$  are defined as before.

### 6.7.3 The degree distribution in the pure birth case with immortal edges

By adding the random parameters of the distributions of outgoing and incoming edges, we obtain that the mixed Poisson distribution describing the degree of a node picked uniformly at random in the pure birth case with immortal edges has the random parameter

$$\alpha S_{J_T} A_{\max(J_T, 2)} + \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{\alpha S_i A_{Y_T}}{Y_T - 1} + \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{\alpha S_i (A_l - A_{l+1})}{l-1}.$$

Note that this distribution converges to the

$$\text{MixPo}(\alpha(S + \mathbb{E}(S))A)$$

distribution as  $T \rightarrow \infty$ , where  $S$  and  $A$  are defined as before.

#### 6.7.1 Remark

It is easy to see that we can obtain the same convergence rates as in the case where edges can die analogously to the procedure in Subsection 6.3.1. Indeed the proof becomes even slightly easier if edges cannot die.

For the time being, we condition on  $S = s$ . We then have  $\alpha(S + \mathbb{E}(S))A \sim \text{Exp}(\frac{\lambda}{\alpha(s + \mathbb{E}(S))})$  since  $A \sim \text{Exp}(\lambda)$ . For  $X \sim \text{MixPo}(\alpha(s + \mathbb{E}(S))A)$ , we thus obtain

$$\mathbb{P}(X = k) = \tilde{\lambda} \int_0^{\infty} e^{-\tilde{\lambda}x} \frac{x^k}{k!} e^{-x} dx = \left( \frac{\frac{1}{\tilde{\lambda}}}{\frac{1}{\tilde{\lambda}} + 1} \right)^k \frac{1}{\frac{1}{\tilde{\lambda}} + 1}$$

for  $k \in \mathbb{N}_0$ , where  $\tilde{\lambda} = \frac{\lambda}{\alpha(s + \mathbb{E}(S))}$  (cf. page 213 in [JKK05]). Thus for a deterministic social index, the asymptotic degree distribution is a geometric distribution. Lifting the conditioning, we see that for any social index distribution, the asymptotic degree distribution is a mixture of geometric distributions.

### 6.7.2 Remark

- (i) Let  $X^{(out)}$  be distributed according to the asymptotic distribution of outgoing edges. Then by the same reasoning as above, we obviously obtain that, conditioned on  $S = s$ , the asymptotic distribution of outgoing edges is given by

$$\mathbb{P}(X^{(out)} = k) = \tilde{\lambda}_1 \int_0^{\infty} e^{-\tilde{\lambda}_1 x} \frac{x^k}{k!} e^{-x} dx = \left( \frac{\frac{1}{\tilde{\lambda}_1}}{\frac{1}{\tilde{\lambda}_1} + 1} \right)^k \frac{1}{\frac{1}{\tilde{\lambda}_1} + 1} \quad (6.94)$$

for  $k \in \mathbb{N}_0$ , where  $\tilde{\lambda}_1 = \frac{\lambda}{\alpha s}$ . Thus the unconditional asymptotic distribution of outgoing edges is the same mixture of geometric distribution that we obtained as asymptotic degree distribution for a special case of the SHR model in Section 5.4 if we choose  $\lambda = \alpha$ .

Let  $\lambda = \alpha$ , and let  $S$  be a random variable such that  $S+1 \sim \text{Pareto}(\theta-1, 1)$  where  $\theta > 2$ . Then we have  $U := \log(S+1) \sim \text{Exp}(\theta-1)$  by a well-known property of the Pareto distribution. By (6.94), the asymptotic distribution of the out-degree is then a mixture of geometric distributions with mixing distribution  $\mathcal{L}(1/(S+1)) = \mathcal{L}(e^{-U})$ . By Remark 3.2 in [Rei09], this mixture is the Yule-Simon distribution with parameter  $\theta-1$  shifted by one, which obeys a power law with exponent  $\theta$  in the weak sense by Remark 5.4.9. Note that this is the asymptotic distribution of the in-degree in the classical preferential attachment model defined in Definition 5.4.1 (cf. Theorem 5.4.4) if we choose  $\theta = 3$ . Note further that we obtain the asymptotic degree distribution of the continuous-time preferential attachment model defined in Definition 5.4.13 for general  $\theta$ . Thus we can obtain a continuous-time fitness model that has essentially the same asymptotic degree distribution as those preferential attachment models, i.e. a distribution that obeys a power law with exponent  $\theta$ , by slightly changing how the edges are created, for example as follows. We could create new edges similar to the procedure in Definition 5.4.13, i.e. first create half-edges instead of complete edges that are joined together in the order in which they are created, and thereby remove the incoming edges in the original model.

- (ii) Let  $X^{(in)}$  be distributed according to the asymptotic distribution of incoming edges. Then we obtain that the asymptotic distribution of the number of incoming edges is given by

$$\mathbb{P}(X^{(in)} = k) = \tilde{\lambda}_2 \int_0^{\infty} e^{-\tilde{\lambda}_2 x} \frac{x^k}{k!} e^{-x} dx = \left( \frac{\frac{1}{\tilde{\lambda}_2}}{\frac{1}{\tilde{\lambda}_2} + 1} \right)^k \frac{1}{\frac{1}{\tilde{\lambda}_2} + 1}$$

for  $k \in \mathbb{N}_0$ , where  $\tilde{\lambda}_2 = \frac{\lambda}{\alpha \mathbb{E}(S)}$ . Note that we obtain the same geometric distribution that is the asymptotic degree distribution in the uniform attachment model if we choose  $\lambda$ ,  $\alpha$  and  $S$  such that  $\tilde{\lambda}_2 = 1$ , e.g.  $\alpha = \lambda$  and  $S$  such that  $\mathbb{E}(S) = 1$  (cf. Section 5.4). Obviously, for all admissible parameters we have a distribution that can be obtained as asymptotic degree distribution in the SHR model from Section 5.4.

Note further that we obtain the same asymptotic distribution as in the case where  $m > 1$  edges are added per time step in the model considered in [BRS<sup>+</sup>01] (cf. Remark 5.4.18) if we choose  $\lambda = 1/m$  and  $\alpha$  and  $S$  such that  $\alpha \mathbb{E}(S) = 1$ , where  $1/m$  can be interpreted as discrete-time birth rate of the node population.<sup>8</sup>

Furthermore, we obtain the same distribution as in the model considered in [CHK<sup>+</sup>01] with parameter  $\delta$  (cf. Remark 5.4.18) if we choose  $\lambda = 1$  and  $\alpha$  and  $S$  such that  $\alpha \mathbb{E}(S) = 2\delta$ . This is plausible since  $2\delta$  can be seen as discrete-time birth rate of the edges, which should be closely related to the "average" birth rate  $\alpha \mathbb{E}(S)$  of the edges in the model by Britton and Lindholm.

---

<sup>8</sup> However, in the model from [BRS<sup>+</sup>01] the birth rate of the node population is constant whereas this rate is linear in the model by Britton and Lindholm.

# 7 A spatial network model

## 7.1 Model and main result

In this chapter, we extend our model by the spatial aspect. We assume that nodes are located in some Euclidean space, and the rates at which edges are created depend on the distances between the nodes. Spatial models are important for certain real-world networks such as social networks. Note that the spatial component can obviously lead to a higher clustering, represented by a higher clustering coefficient (cf. Remark 5.1.7).

For the sake of simplicity, we stick to the pure birth case, where nodes cannot die. Let each node  $i$  be equipped with a *position* that is denoted by  $\mathcal{P}(i)$  and uniformly distributed on a  $m$ -dimensional *unit hypercube*  $\mathcal{C}$  independent from all other random variables.<sup>9</sup> Since we do not want to treat edge effects, we define the *metric*

$$d : \mathcal{C} \times \mathcal{C} \rightarrow [0, \infty), (x, y) \mapsto \min \left( \|x - y\|, \min_{\mathfrak{R} \subset \{1, \dots, m\}} (\|x + e_{\mathfrak{R}} - y\|), \min_{\mathfrak{R} \subset \{1, \dots, m\}} (\|y + e_{\mathfrak{R}} - x\|) \right),$$

where  $e_{\mathfrak{R}} = \sum_{k \in \mathfrak{R}} e_k$  and  $e_k$  denotes the  $k$ th standard unit vector, i.e. the vector whose entries are all zero except of the  $k$ th one that is one.

Note that one can think of the nodes being placed on a  $m$ -dimensional torus.

We assume that so-called *potential edges* are created in the same way (including the same rates) as the edges in the non-spatial case. However, each potential edge between nodes  $i$  and  $j$  is deleted instantaneously with probability  $1 - \psi(d(\mathcal{P}(i), \mathcal{P}(j)))$  (independently from all other events) for some measurable function  $\psi : [0, 1/2] \rightarrow [0, 1]$ . All potential edges that are not deleted become *actual* edges instantaneously.

Since the above procedure can be interpreted as thinning, each living node  $i$  creates edges that connect  $i$  to some other living node  $j$  at rate

$$\frac{S_i \psi(d(\mathcal{P}(i), \mathcal{P}(j)))}{l - 1}$$

at time  $t$  if  $l$  nodes are alive at this time (cf. Section 6.2). Analogously to the procedure in the

<sup>9</sup> Note that the positions of nodes are uniformly distributed in space in most spatial preferential attachment models as well; see e.g. Subsection IV.D. of [Bar11].

non-spatial case, we thus obtain the following distribution for the number of incoming edges:

$$\begin{aligned} \text{MixPo} & \left( \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} S_i \psi(d(\mathcal{P}(i), \mathcal{P}(J_T))) \frac{1}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \right. \\ & \left. + \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} S_i \psi(d(\mathcal{P}(i), \mathcal{P}(J_T))) \frac{1}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right). \end{aligned} \quad (7.1)$$

For the outgoing edges, the  $\psi$ -term necessitates the analogue node-by-node analysis as for the incoming edges, yielding

$$\begin{aligned} \text{MixPo} & \left( \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} S_{J_T} \psi(d(\mathcal{P}(i), \mathcal{P}(J_T))) \frac{1}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \right. \\ & \left. + \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} S_{J_T} \psi(d(\mathcal{P}(i), \mathcal{P}(J_T))) \frac{1}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right). \end{aligned} \quad (7.2)$$

### 7.1.1 Definition and Remark

Let  $\bar{\psi} = \mathbb{E}(\psi(d(U_1, U_2))) \leq 1$ , where  $U_1$  and  $U_2$  are independent and uniformly distributed on  $\mathcal{C}$ . Note that  $\mathcal{L}(d(U_1, U_2)) = \mathcal{L}(d(U_1, u_2))$  and thus  $\bar{\psi} = \mathbb{E}(\psi(d(U_1, u_2)))$  for any fixed  $u_2 \in \mathcal{C}$ . Note further that  $\mathcal{L}(d(U_1, U_2)) = \mathcal{L}(d(U_1, u_2)) = \mathcal{L}(U)$  in the one-dimensional case for all fixed  $u_2 \in \mathcal{C}$  and a random variable  $U$  that is uniformly distributed on  $[0, 1/2]$  so that  $\bar{\psi} = \mathbb{E}(\psi(U))$ .

The asymptotic degree distribution is now given by the following corollary, which follows directly from Theorem 7.1.5 below.

### 7.1.2 Corollary

Let  $S_{J_\infty}$  and  $A_{J_\infty}$  be defined as in the non-spatial case. Then

$$\text{MixPo} \left( \frac{\alpha \bar{\psi}}{\beta} (S_{J_\infty} + \mathbb{E}(S)) (1 - e^{-\beta A_{J_\infty}}) \right) \quad (7.3)$$

is the asymptotic degree distribution.

In order to simplify notation, let  $\Lambda_T$  be the random parameter of the degree distribution, i.e. the sum of the random parameters of the distributions (7.1) and (7.2), and let  $M$  be the random parameter of the asymptotic degree distribution (7.3). Then the following main theorem of this section yields the corresponding convergence rate.

### 7.1.3 Theorem

We have

$$\begin{aligned} & d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) \\ & \leq \frac{4\alpha}{\lambda} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E} \left( \frac{1}{\sqrt{Y_T}} \right) + \frac{4\alpha}{\lambda} \bar{\psi} \mathbb{E}(S) \mathbb{E} \left( \frac{1}{Y_T} \right) + \mathbb{E}(S) \frac{\alpha \lambda}{\beta(\beta + \lambda)} \frac{1}{e^{\lambda T} - 1}, \end{aligned}$$

where  $C(S, \psi_{1,2}^{d,\mathcal{P}}) = \sqrt{\text{Var}(S_1 \psi_{1,2}^{d,\mathcal{P}})} + \mathbb{E}(S) \sqrt{\text{Var}(\psi_{1,2}^{d,\mathcal{P}})}$  with  $\psi_{1,2}^{d,\mathcal{P}} = \psi(d(\mathcal{P}(1), \mathcal{P}(2)))$ .

### 7.1.4 Definition and Remark

Let  $\sigma_\psi^2 = \text{Var}(\psi_{1,2}^{d,\mathcal{P}})$ , where  $\psi_{1,2}^{d,\mathcal{P}} = \psi(d(\mathcal{P}(1), \mathcal{P}(2)))$ . Then we have  $\sigma_\psi^2 = \text{Var}(\psi(d(U_1, U_2)))$  for  $U_1$  and  $U_2$  being independent and uniformly distributed on  $\mathcal{C}$ . Note that we can use

$$\text{Var}(S_1 \psi_{1,2}^{d,\mathcal{P}}) \leq \sigma_S^2 \sigma_\psi^2 + (\mathbb{E}(S))^2 \sigma_\psi^2 + \sigma_S^2 \bar{\psi}^2 \leq \mathbb{E}(S^2) \sigma_\psi^2 + \sigma_S^2 \bar{\psi}^2$$

in order to bound  $C(S, \psi_{1,2}^{d,\mathcal{P}})$  in the upper bound from Theorem 7.1.3.

Note further that we have  $\sigma_\psi^2 = \text{Var}(U)$  for  $U$  being uniformly distributed on  $[0, 1/2]$  in the one-dimensional case (cf. Definition and Remark 7.1.1).

The following theorem gives the corresponding convergence rate.

### 7.1.5 Theorem

For  $T \geq \log(2)/\lambda$ , we have

$$d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) \leq \frac{\sqrt{32}\alpha}{\sqrt{\lambda}} C(S, \psi_{1,2}^{d,\mathcal{P}}) \sqrt{T} e^{-\frac{1}{2}\lambda T} + 4\alpha \bar{\psi} \mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T},$$

where  $C(S, \psi_{1,2}^{d,\mathcal{P}})$  is defined as in Theorem 7.1.3 above, and for  $T \rightarrow \infty$ , the right-hand side is of the order  $O(\sqrt{T} e^{-\frac{1}{2}\lambda T})$  if  $\sigma_S$  is finite.

*Proof:* The inequality follows directly from Theorem 6.3.1, Proposition 2.2.4 and Corollary 2.2.5.  $\square$

### 7.1.6 Remark

Note that a node  $J_T$  picked uniformly at random can only have a multiple edge if there have been at least two potential edges up to time  $T$  that connected  $J_T$  to the same node. Since the potential edges are created in the same way as the (ordinary) edges in the non-spatial case, the negligibility of multiple edges follows from the corresponding result in the case considered in Lemma 6.4.4. More precisely, the probability that  $J_T$  has at least one multiple edge at time  $T$  is of the order  $O(T^2 e^{-\frac{1}{6}(\lambda - \mu)T})$  in the spatial case as well.

## 7.2 Proof of the main theorem

In order to simplify the notation, we set  $\psi_{i,j}^{d,\mathcal{P}} = \psi(d(\mathcal{P}(i), \mathcal{P}(j)))$  for  $i, j \in \mathbb{N}$ . As in the non-spatial case, we apply Theorem 3.4.1 and obtain

$$\begin{aligned} d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) &\leq \mathbb{E}(|\Lambda_T - M|) \\ &\leq \mathbb{E} \left| \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} (S_i + S_{J_T}) \psi_{i,J_T}^{d,\mathcal{P}} \frac{1}{Y_T - 1} (1 - e^{-\beta A_{Y_T}(T)}) \right| \end{aligned} \quad (7.4)$$

$$+ \mathbb{E} \left| \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T - 1} (S_i + S_{J_T}) \psi_{i,J_T}^{d,\mathcal{P}} \frac{1}{l - 1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) - \frac{\alpha \bar{\psi}}{\beta} (S_{J_\infty} + \mathbb{E}(S)) (1 - e^{-\beta A_{J_\infty}}) \right|. \quad (7.5)$$

**Upper bound for (7.4)**

By conditioning on  $Y_T, J_T$  and the conditional independence of node positions, social indices and ages, we obtain analogously to the non-spatial case (compare the derivations of the upper bounds for (6.9) in the pure birth and for (6.45) and (6.46) in the general case) that the expectation (7.4) is equal to

$$2\bar{\psi} \frac{\alpha}{\beta} \mathbb{E} \left( \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \frac{\mathbb{E}(S)}{Y_T - 1} (1 - \mathbb{E}(e^{-\beta A_{Y_T}(T)} | Y_T)) \right).$$

The outer expectation is just the right-hand side of (6.9), which we already know to be bounded from above by (6.10), i.e. by

$$\frac{\beta}{\lambda} \mathbb{E}(S) \mathbb{E} \left( \frac{1}{Y_T} \right).$$

**Upper bound for (7.5)**

The expectation in (7.5) is bounded from above by

$$\mathbb{E} \left| \frac{\alpha}{\beta} \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{(S_i + S_{J_T}) \psi_{i, J_T}^{d, \mathcal{P}} - (\mathbb{E}(S) + S_{J_\infty}) \bar{\psi}}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \quad (7.6)$$

$$+ \frac{\alpha}{\beta} \mathbb{E} \left| (S_{J_\infty} + \mathbb{E}(S)) \bar{\psi} (1 - e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{l=i \vee J_T}^{Y_T-1} \frac{(\mathbb{E}(S) + S_{J_\infty}) \bar{\psi}}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \quad (7.7)$$

Expression (7.6) corresponds to the second term on the right-hand side of (6.8) and the expression (7.7) to the combined third and fourth term. Since we arranged  $S_{J_T} = S_{J_\infty}$ , the term (7.6) is bounded from above by

$$\begin{aligned} & \mathbb{E} \left| \frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l \left( (S_i + S_{J_\infty}) \psi_{i, J_T}^{d, \mathcal{P}} - (\mathbb{E}(S) + S_{J_\infty}) \bar{\psi} \right) (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \\ & \leq \mathbb{E} \left| \frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l \left( S_i \psi_{i, J_T}^{d, \mathcal{P}} - \mathbb{E}(S) \bar{\psi} \right) (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \\ & + \mathbb{E} \left| \frac{\alpha}{\beta} S_{J_\infty} \sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l \left( \psi_{i, J_T}^{d, \mathcal{P}} - \bar{\psi} \right) (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right|. \end{aligned} \quad (7.8)$$

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

$$\begin{aligned} & \mathbb{E} \left( \mathbb{E} \left( \frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l |S_i \psi_{i, J_T}^{d, \mathcal{P}} - \mathbb{E}(S) \bar{\psi}| (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right) \middle| (Y_t)_{0 \leq t \leq T}, J_T \right) \\ & \leq \mathbb{E} \left( \frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \mathbb{E} \left( \left| \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l (S_i \psi_{i, J_T}^{d, \mathcal{P}} - \mathbb{E}(S) \bar{\psi}) \right| \middle| (Y_t)_{0 \leq t \leq T}, J_T \right) \right). \end{aligned} \quad (7.9)$$



Since the random variables  $(S_i, \psi_{i,J_T}^{d,\mathcal{P}})_{i \neq J_T}$ , are stochastically independent from  $(Y_t)_{0 \leq t \leq T}$ ,  $J_T$  and are identically distributed with expectation  $\mathbb{E}(S)\bar{\psi}$ , we obtain analogously to the non-spatial case that (7.9) is bounded from above by

$$\mathbb{E}\left(\frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)} | Y_T) \frac{1}{\sqrt{l-1}} \sqrt{\text{Var}(S_1 \psi_{1,2}^{d,\mathcal{P}})}\right).$$

The second summand of the right-hand side of (7.8) is smaller than or equal to

$$\begin{aligned} & \mathbb{E}\left(\mathbb{E}\left(\frac{\alpha}{\beta} S_{J_\infty} \sum_{l=J_T \vee 2}^{Y_T-1} \left| \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l (\psi_{i,J_T}^{d,\mathcal{P}} - \bar{\psi}) \right| (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \middle| Y_T, J_T\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}(S_{J_\infty}) \mathbb{E}\left(\frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \left| \frac{1}{l-1} \sum_{i=1, i \neq J_T}^l (\psi_{i,J_T}^{d,\mathcal{P}} - \bar{\psi}) \right| (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \middle| Y_T, J_T\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}(S_{J_\infty}) \frac{\alpha}{\beta} \sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}\left(\left| \frac{1}{l-1} \sum_{i=2}^l (\psi_{i,1}^{d,\mathcal{P}} - \bar{\psi}) \right|\right) \mathbb{E}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)} | Y_T)\right), \end{aligned}$$

where the last equality holds since the sum  $\sum_{i=1, i \neq J_T}^l |\psi_{i,J_T}^{d,\mathcal{P}} - \bar{\psi}|$  is (stochastically) independent from  $J_T$  and the positions of the nodes are independent from  $(Y_t)_{0 \leq t \leq T}$  and  $J_T$ . With the same argument as above, we obtain that the right-hand side is bounded from above by

$$\mathbb{E}\left(\frac{\alpha}{\beta} \mathbb{E}(S) \sqrt{\text{Var}(\psi_{1,J_T}^{d,\mathcal{P}})} \sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)} | Y_T) \frac{1}{\sqrt{l-1}}\right).$$

Altogether, this yields that (7.6) is smaller than or equal to

$$\begin{aligned} & \frac{\alpha}{\beta} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}(e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)} | Y_T) \frac{1}{\sqrt{l-1}}\right) \\ &= \frac{\alpha}{\beta} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}(e^{-\beta A_{l+1}(T)} (1 - e^{-\beta(A_l(T) - A_{l+1}(T))}) | Y_T) \frac{1}{\sqrt{l-1}}\right) \\ &\leq \frac{\alpha}{\beta} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E}\left(\sum_{l=J_T \vee 2}^{Y_T-1} \mathbb{E}(1 - e^{-\beta(A_l(T) - A_{l+1}(T))} | Y_T) \frac{1}{\sqrt{l-1}}\right), \end{aligned} \tag{7.10}$$

where  $C(S, \psi_{1,2}^{d,\mathcal{P}}) = \sqrt{\text{Var}(S_1 \psi_{1,2}^{d,\mathcal{P}})} + \mathbb{E}(S) \sqrt{\text{Var}(\psi_{1,2}^{d,\mathcal{P}})}$ . Now we observe that the right-hand side of this expression is equal to the right-hand side of (6.11) if we substitute the constant factor  $\sigma_S$  by the constant factor in the right-hand side of (7.10). Thus we only need to substitute this factor in the upper bound that we derived for the right-hand side of (6.11) in order to obtain the following upper bound for the right-hand side of (7.10):

$$\frac{4\alpha}{\lambda} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E}\left(\frac{1}{\sqrt{Y_T}}\right).$$

For the expectation in (7.7), we have

$$\begin{aligned} & \mathbb{E} \left| (S_{J_\infty} + \mathbb{E}(S)) \bar{\psi} (1 - e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{\substack{l=i \vee J_T \\ l \neq J_T}}^{Y_T-1} \frac{(\mathbb{E}(S) + S_\infty) \bar{\psi}}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right| \\ & \leq 2\bar{\psi} \mathbb{E} \left| \mathbb{E}(S) (1 - e^{-\beta A_{J_\infty}}) - \sum_{\substack{i=1 \\ i \neq J_T}}^{Y_T} \sum_{\substack{l=i \vee J_T \\ l \neq J_T}}^{Y_T-1} \frac{\mathbb{E}(S)}{l-1} (e^{-\beta A_{l+1}(T)} - e^{-\beta A_l(T)}) \right|. \end{aligned}$$

Note that the (outer) expectation in this expression is the left-hand side of (6.12), which we already showed to be bounded from above by

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

### Conclusion

Altogether, we obtain that  $d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M))$  is bounded from above by

$$\frac{4\alpha}{\lambda} C(S, \psi_{1,2}^{d,\mathcal{P}}) \mathbb{E} \left( \frac{1}{\sqrt{Y_T}} \right) + \frac{4\alpha}{\lambda} \bar{\psi} \mathbb{E}(S) \mathbb{E} \left( \frac{1}{Y_T} \right) + 2\mathbb{E}(S) \bar{\psi} \frac{\alpha\lambda}{\beta(\beta + \lambda)} \frac{1}{e^{\lambda T} - 1}.$$

## 7.3 Examples

### 7.3.1 The "hardcore" case

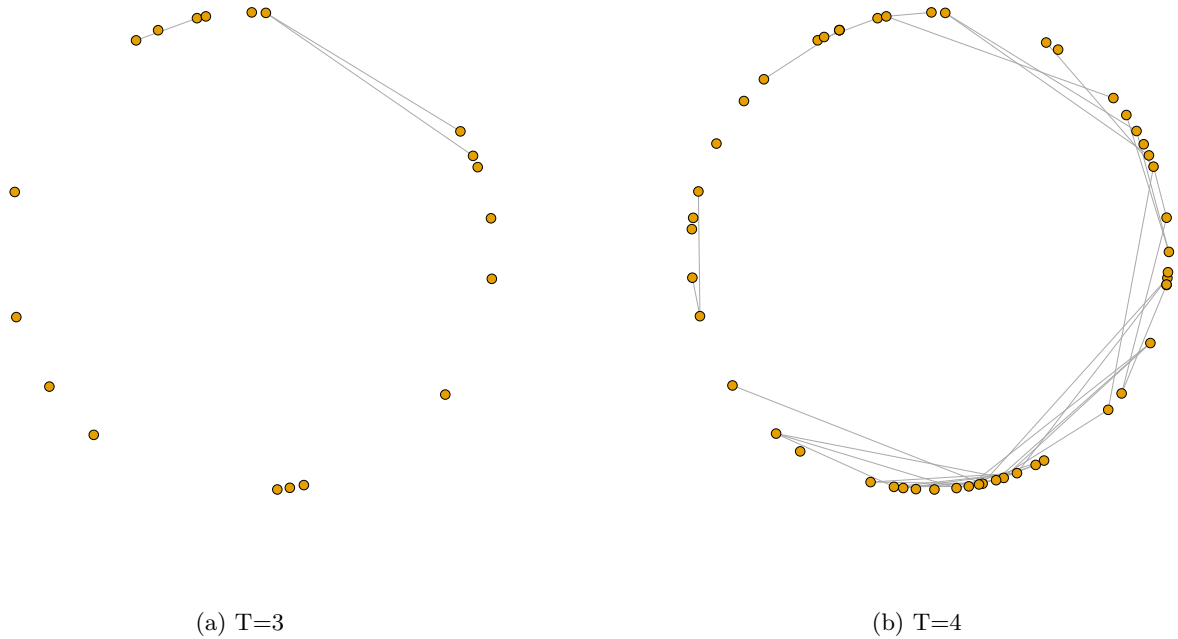


Figure 7.1: Simulated realizations of the one-dimensional "hardcore" case of the spatial model with  $\alpha = 4$ ,  $\beta = \lambda = 1$  and  $\mathcal{R} = 0.2$  for  $T = 3$  and  $T = 4$  and deterministic social indices.

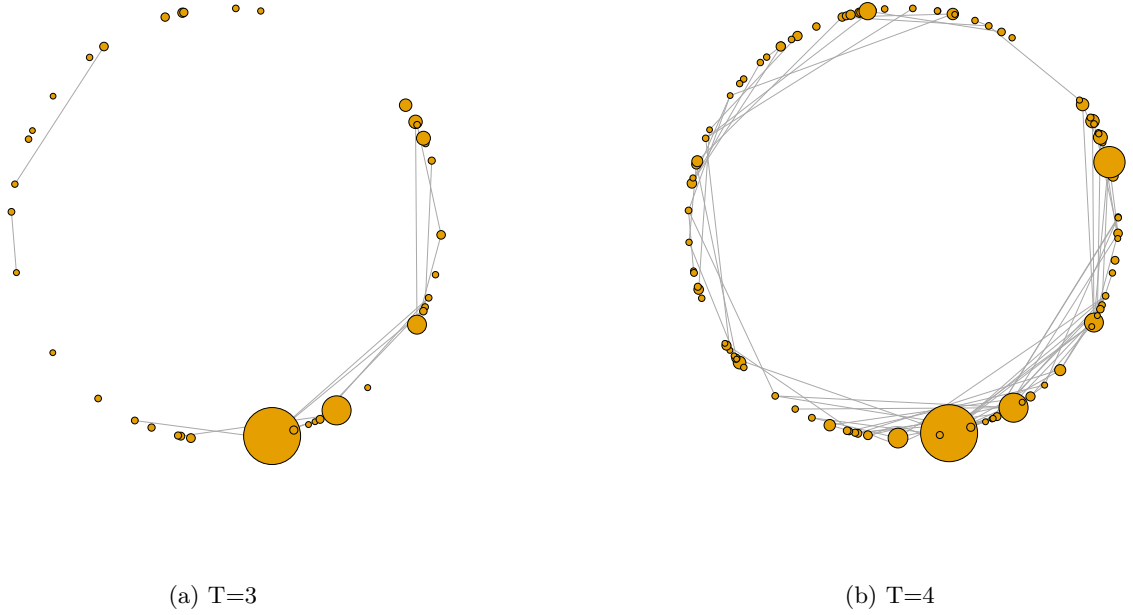


Figure 7.2: Simulated realizations of the one-dimensional “hardcore” case of the spatial model with  $\alpha = 4$ ,  $\beta = \lambda = 1$  and  $\mathcal{R} = 0.2$  for  $T = 3$  and  $T = 4$  and Pareto(3, 2/3) distributed social indices.

First we consider the case where edges appear if and only if the distance between the corresponding nodes is not larger than some fixed value  $0 < \mathcal{R} \leq 1/2$ , i.e. the function  $\psi$  is given by

$$\psi(x) = \mathbb{1}_{\{\bar{x}: |\bar{x}| \leq \mathcal{R}\}}(x).$$

In order to apply Theorem 7.1.5, we need  $\bar{\psi}$  and  $\sigma_\psi$ , which are given by the following lemma.

### 7.3.1 Lemma

For the  $m$ -dimensional case with  $\psi(x) = \mathbb{1}_{\{\bar{x}: |\bar{x}| \leq \mathcal{R}\}}(x)$ , we have

$$\bar{\psi} = \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2} + 1)} \mathcal{R}^m = \begin{cases} \frac{\pi^{\frac{m}{2}}}{(\frac{m}{2})!} \mathcal{R}^m & \text{if } m \text{ is even,} \\ \frac{2(\frac{m-1}{2})!(4\pi)^{\frac{m-1}{2}}}{m!} \mathcal{R}^m & \text{if } m \text{ is odd} \end{cases}$$

and

$$\sigma_\psi^2 = \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2} + 1)} \mathcal{R}^m \left( 1 - \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2} + 1)} \mathcal{R}^m \right) = \begin{cases} \frac{\pi^{\frac{m}{2}}}{(\frac{m}{2})!} \mathcal{R}^m \left( 1 - \frac{\pi^{\frac{m}{2}}}{(\frac{m}{2})!} \mathcal{R}^m \right) & \text{if } m \text{ is even,} \\ \frac{2(\frac{m-1}{2})!(4\pi)^{\frac{m-1}{2}}}{m!} \mathcal{R}^m \left( 1 - \frac{2(\frac{m-1}{2})!(4\pi)^{\frac{m-1}{2}}}{m!} \mathcal{R}^m \right) & \text{if } m \text{ is odd.} \end{cases}$$

*Proof:* Note that  $\bar{\psi} = \mathbb{E}(\psi(d(U_1, U_2))) = \mathbb{E}(\psi(d(U_1, u_2))) = \mathbb{E}(\mathbb{1}_{\{\psi(d(U_1, u_2)) \leq \mathcal{R}\}})$ , where  $U_1$  and  $U_2$  are independent and uniformly distributed on  $\mathcal{C}$  and  $u_2 \in \mathcal{C}$ , is the probability that the distance of a random node to a fixed one is less or equal to  $\mathcal{R}$ . This probability is equal to the volume of a  $m$ -dimensional ball with radius  $\mathcal{R}$ . Thus we obtain the desired expressions for  $\bar{\psi}$ .

Alternatively, we can compute  $\mathbb{E}(\psi(d(U_1, U_2)))$  directly. For the sake of simplicity, we only state the computation for  $m = 1$  here.<sup>10</sup> In this case, we have

$$\begin{aligned}
\mathbb{E}(\psi(d(U_1, U_2))) &= \mathbb{E}(\mathbb{1}_{\{d(U_1, U_2) \leq \mathcal{R}\}}) \\
&= \int_0^1 \int_0^1 \mathbb{1}_{\{|x-y| \leq \mathcal{R}\}} dx dy \\
&= \int_0^{\mathcal{R}} \left( \int_0^{y+\mathcal{R}} 1 dx + \int_{y+1-\mathcal{R}}^1 1 dx \right) dy + \int_{\mathcal{R}}^{1-\mathcal{R}} \int_{y-\mathcal{R}}^{y+\mathcal{R}} 1 dx dy + \int_{1-\mathcal{R}}^1 \left( \int_{y-\mathcal{R}}^1 1 dx + \int_0^{y-1+\mathcal{R}} 1 dx \right) dy \\
&= \int_0^{\mathcal{R}} 2\mathcal{R} dy + \int_{\mathcal{R}}^{1-\mathcal{R}} 2\mathcal{R} dy + \int_{1-\mathcal{R}}^1 2\mathcal{R} dy \\
&= \int_0^1 2\mathcal{R} dy = 2\mathcal{R}.
\end{aligned}$$

Note that  $\mathbb{E}(\psi(d(U_1, U_2))) = \mathbb{E}((\psi(d(U_1, U_2)))^2)$  since  $\psi$  is an indicator function. This yields the expressions for

$$\sigma_\psi^2 = \text{Var}(\psi(d(U_1, U_2))) = E((\psi(d(U_1, U_2)))^2) - (\mathbb{E}(\psi(d(U_1, U_2))))^2 = \bar{\psi} - \bar{\psi}^2.$$

□

Combining Corollary 7.1.2 and Lemma 7.3.1, we obtain the following corollary.

### 7.3.2 Corollary

For  $\psi(x) = \mathbb{1}_{\{\tilde{x}: |\tilde{x}| \leq \mathcal{R}\}}(x)$ , we have that

$$\text{MixPo} \left( \frac{\alpha \pi^{\frac{m}{2}}}{\beta \Gamma(\frac{m}{2} + 1)} \mathcal{R}^m (S_{J_\infty} + \mathbb{E}(S)) (1 - e^{-\beta A_{J_\infty}}) \right)$$

is the asymptotic degree distributions in the  $m$ -dimensional case.

Theorem 7.1.5, Definition and Remark 7.1.4 and Lemma 7.3.1 yield the following corollary, which provides convergence rates.

### 7.3.3 Corollary

In the  $m$ -dimensional case with  $\psi(x) = \mathbb{1}_{\{\tilde{x}: |\tilde{x}| \leq \mathcal{R}\}}(x)$ , we have

$$\begin{aligned}
&d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) \\
&\leq \frac{\sqrt{32}\alpha}{\sqrt{\lambda}} \left( \sqrt{\sigma_S^2 + \mathbb{E}(S^2)} + \mathbb{E}(S) \right) \left( \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2} + 1)} \right)^{\frac{1}{2}} \mathcal{R}^{\frac{m}{2}} \sqrt{T} e^{-\frac{1}{2}\lambda T} \\
&\quad + 4\alpha \frac{\pi^{\frac{m}{2}}}{\Gamma(\frac{m}{2} + 1)} \mathcal{R}^m \mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T}
\end{aligned}$$

for  $T \geq \log(2)/\lambda$ .

<sup>10</sup> Note that we do not use the simpler form of  $\psi(d(U_1, U_2))$  given in Definition and Remark 7.1.1 here for illustrative reasons.

**7.3.4 Remark**

Since decreasing  $\mathcal{R}$  can be interpreted as *enlarging* the hypercube  $\mathcal{C}$  by rescaling, it is also interesting to consider the limit of the upper bound in Corollary 7.3.3 for  $\mathcal{R} \rightarrow 0$ . In order to obtain interesting behaviour, we let  $\alpha$ , i.e. the birth rates of the edges, increase as  $\mathcal{R}$  decreases in an appropriate way in the following corollary, which follows immediately from Corollary 7.3.3.

**7.3.5 Corollary**

In the  $m$ -dimensional case with  $\psi(x) = \mathbb{1}_{\{\tilde{x}:|\tilde{x}|\leq\mathcal{R}\}}(x)$ , we set  $\alpha = \Gamma(\frac{m}{2}+1)\pi^{-m/2}\mathcal{R}^{-m}$ . For  $T \geq \log(2)/\lambda$ , we then have

$$\begin{aligned} & d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(\beta^{-1}(S_{J_\infty} + \mathbb{E}(S))(1 - e^{-\beta A_{J_\infty}}))) \\ & \leq \frac{\sqrt{32}}{\sqrt{\lambda}} \left( \sqrt{\sigma_S^2 + \mathbb{E}(S^2)} + \mathbb{E}(S) \right) \left( \frac{\Gamma(\frac{m}{2} + 1)}{\pi^{\frac{m}{2}}} \right)^{\frac{1}{2}} \mathcal{R}^{-\frac{m}{2}} \sqrt{T} e^{-\frac{1}{2}\lambda T} + 4\mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T}. \end{aligned}$$

**7.3.6 Remark**

We may now let  $\mathcal{R} = \mathcal{R}(T)$  converge to zero as  $T \rightarrow \infty$  in such way that the upper bound from Corollary 7.3.5 still converges to zero as  $T \rightarrow \infty$ . As a result, we can obtain a approximation for the degree distribution at some finite time  $T$  and an upper bound for the corresponding approximation error for small  $\mathcal{R}$ . Since a small  $\mathcal{R}$  corresponds to a large hypercube (see Remark 7.3.4 above), this result might be interesting for applications.

**7.3.2 Edge probability depending linearly on distance**

For the sake of simplicity, we stick to the one-dimensional case here. We consider the case where the probability for a potential edge to be deleted is proportional to two times the distance between the corresponding nodes, i.e. the function  $\psi$  is given by

$$\psi(x) = 1 - 2x.$$

In order to obtain an upper bound for  $d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M))$  from Theorem 7.1.5, we need the following lemma.

**7.3.7 Lemma**

For the one-dimensional case with  $\psi(x) = 1 - x$ , we have

$$\bar{\psi} = \frac{1}{2} \text{ and } \sigma_\psi^2 = \frac{1}{12}.$$

*Proof:* Let  $U_1, U_2$  and  $U$  be defined as in Definition and Remark 7.1.1. We then have

$$\bar{\psi} = \mathbb{E}(\psi(d(U_1, U_2))) = \mathbb{E}(\psi(U)) = 1 - 2\mathbb{E}(U) = \frac{1}{2}.$$

Furthermore, we obtain

$$\mathbb{E}((\psi(d(U_1, U_2)))^2) = \mathbb{E}(\psi(U)^2) = \mathbb{E}((1 - 2U)^2)$$

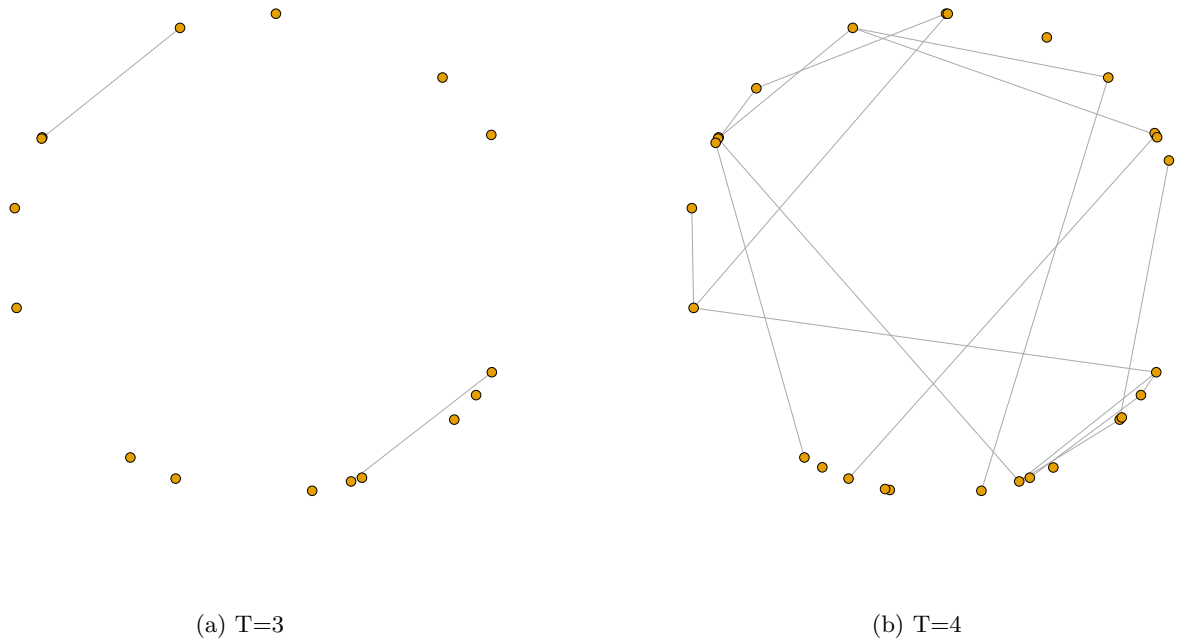


Figure 7.3: Simulated realizations of the case where the edge probability depends linearly on the distance of the spatial model with  $\alpha = 4$  and  $\beta = \lambda = 1$  for  $T = 3$  and  $T = 4$  and deterministic social indices.

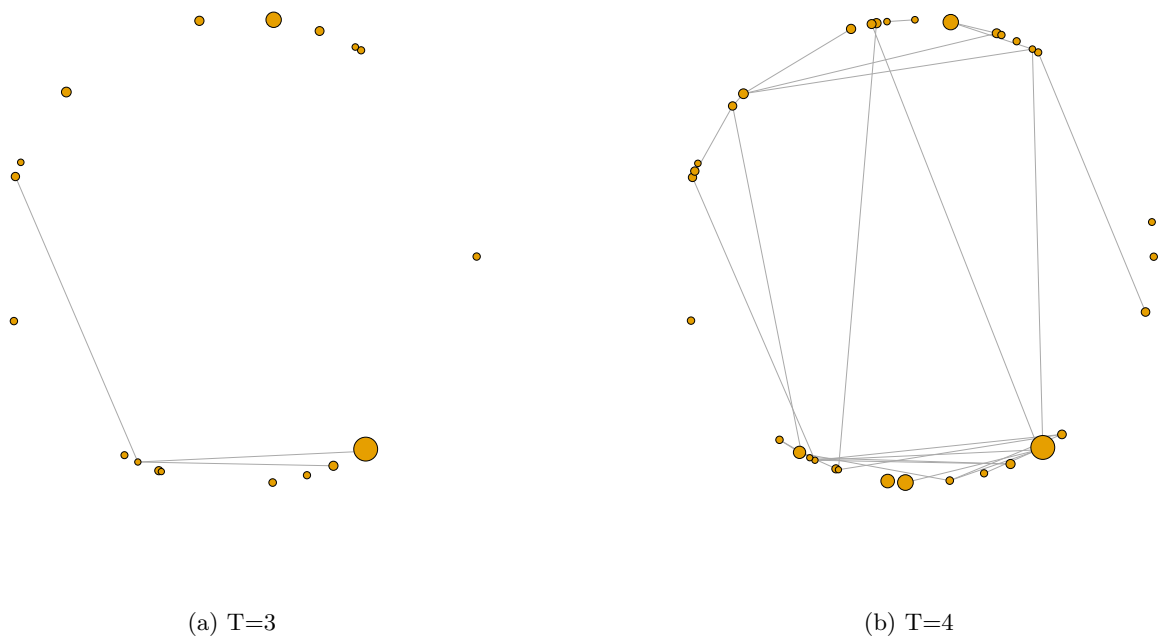


Figure 7.4: Simulated realizations of the case where the edge probability depends linearly on the distance of the spatial model with  $\alpha = 4$  and  $\beta = \lambda = 1$  for  $T = 3$  and  $T = 4$  and Pareto(3, 2/3) distributed social indices.

and compute

$$\mathbb{E}((1 - 2U)^2) = 2 \int_0^{\frac{1}{2}} (1 - 2x)^2 dx = 2 \left[ x - 2x^2 + \frac{4x^3}{3} \right]_0^{\frac{1}{2}} = \frac{1}{3}.$$

Finally, this implies

$$\sigma_\psi^2 = \text{Var}(\psi(d(U_1, U_2))) = E((\psi(d(U_1, U_2)))^2) - (\mathbb{E}(\psi(d(U_1, U_2))))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

□

Lemma 7.3.7 allows us to state the following corollary, which reveals an explicit expression for the asymptotic degree distribution and follows directly from Corollary 7.1.2 and Lemma 7.3.7.

### 7.3.8 Corollary

For the one-dimensional case with  $\psi(x) = 1 - x$ , we have that

$$\text{MixPo} \left( \frac{\alpha}{2\beta} (S_{J_\infty} + \mathbb{E}(S)) (1 - e^{-\beta A_{J_\infty}}) \right)$$

is the asymptotic degree distribution.

Theorem 7.1.5, Definition and Remark 7.1.4 and Lemma 7.3.1 yield the following corollary, which provides convergence rates.

### 7.3.9 Corollary

For the one-dimensional case with  $\psi(x) = 1 - x$ , we have

$$\begin{aligned} d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(M)) &\leq \frac{\sqrt{8}\alpha}{\sqrt{3}\lambda} \left( \sqrt{3\sigma_S^2 + \mathbb{E}(S^2)} + \mathbb{E}(S) \right) \sqrt{T} e^{-\frac{1}{2}\lambda T} \\ &\quad + 2\alpha \mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T} \end{aligned}$$

for  $T \geq \log(2)/\lambda$ .

### 7.3.3 Gaussian kernels

Finally, we would like to consider the case where  $\psi$  is a *Gaussian kernel*, i.e. the density of the standard normal distribution renormalized so that  $\psi(0) = 1$ . More precisely, we define

$$\psi(x) = e^{-\frac{x^2}{\mathcal{R}}} \tag{7.11}$$

for  $x \in [0, 1/2]$  and some  $\mathcal{R} > 0$ .

### 7.3.10 Lemma

For the one-dimensional case with  $\psi(x) = e^{-x^2/\mathcal{R}}$ , we have

$$\bar{\psi} = \sqrt{\pi\mathcal{R}} \left( 2\Phi \left( \frac{1}{\sqrt{2\mathcal{R}}} \right) - 1 \right) \text{ and } \sigma_\psi^2 = \sqrt{2\pi\mathcal{R}} \left( 2\Phi \left( \frac{1}{\sqrt{2\mathcal{R}}} \right) - 1 \right) - \pi\mathcal{R} \left( 2\Phi \left( \frac{1}{\sqrt{2\mathcal{R}}} \right) - 1 \right)^2,$$

where  $\Phi$  denotes the cumulative distribution function of the standard normal distribution.

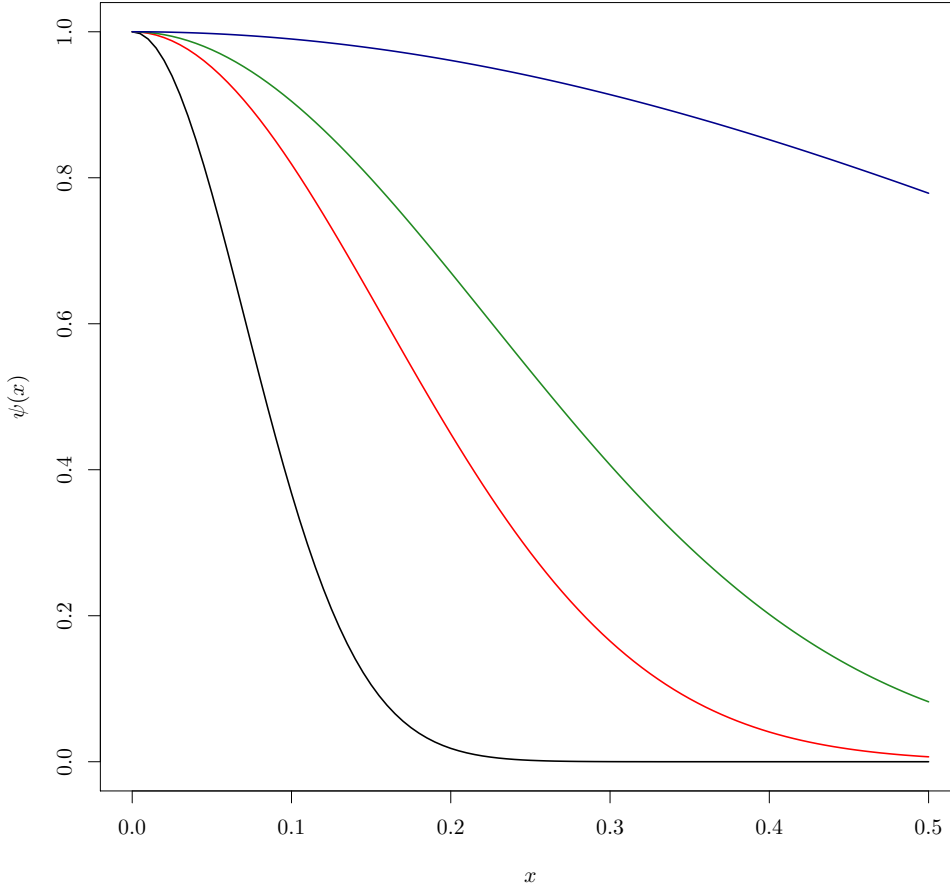


Figure 7.5: The Gaussian kernel  $\psi$  defined by (7.11) for  $\mathcal{R} = 0.01$  (black line),  $\mathcal{R} = 0.05$  (red line),  $\mathcal{R} = 0.1$  (green line) and  $\mathcal{R} = 1$  (blue line)

*Proof:* Let  $U$  be defined as in Definition and Remark 7.1.1. We then have

$$\bar{\psi} = \mathbb{E}(\psi(U)) = 2 \int_0^{\frac{1}{2}} e^{-\frac{x^2}{\mathcal{R}}} dx = 2\Phi\left(\frac{1}{\sqrt{2\mathcal{R}}}\right)\sqrt{\pi\mathcal{R}} - 2\Phi(0)\sqrt{\pi\mathcal{R}} = \sqrt{\pi\mathcal{R}}\left(2\Phi\left(\frac{1}{\sqrt{2\mathcal{R}}}\right) - 1\right).$$

Analogously, we obtain

$$\mathbb{E}((\psi(U))^2) = 2 \int_0^{\frac{1}{2}} e^{-\frac{2x^2}{\mathcal{R}}} dx = 2\Phi\left(\frac{1}{\sqrt{\mathcal{R}}}\right)\sqrt{\frac{\pi\mathcal{R}}{2}} - 2\Phi(0)\sqrt{\frac{\pi\mathcal{R}}{2}} = \sqrt{2\pi\mathcal{R}}\left(2\Phi\left(\frac{1}{\sqrt{\mathcal{R}}}\right) - 1\right).$$

By Definition and Remark 7.1.4, we thus have

$$\sigma_{\psi}^2 = \text{Var}(\psi(U)) = \mathbb{E}((\psi(U))^2) - (\mathbb{E}(\psi(U)))^2 = \sqrt{2\pi\mathcal{R}}\left(2\Phi\left(\frac{1}{\sqrt{\mathcal{R}}}\right) - 1\right) - \pi\mathcal{R}\left(2\Phi\left(\frac{1}{\sqrt{2\mathcal{R}}}\right) - 1\right)^2.$$

□

The following corollary gives us an explicit expression for the asymptotic degree distribution and is an immediate consequence of Corollary 7.1.2 and Lemma 7.3.10.



**7.3.11 Corollary**

For the one-dimensional case with  $\psi(x) = e^{-x^2/\mathcal{R}}$ , we have that

$$\text{MixPo}\left(\sqrt{\pi\mathcal{R}}\left(2\Phi\left(\frac{1}{\sqrt{2\mathcal{R}}}\right) - 1\right)\frac{\alpha}{\beta}(S_{J_\infty} + \mathbb{E}(S))(1 - e^{-\beta A_{J_\infty}})\right)$$

is the asymptotic degree distribution.

The following corollary reveals convergence rates.

**7.3.12 Corollary**

For the one-dimensional case with  $\psi(x) = e^{-x^2/\mathcal{R}}$ , we have

$$\begin{aligned} & d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(\text{M})) \\ & \leq \frac{\sqrt{32}\alpha}{\lambda} \left( (\sigma_S^2 \pi \mathcal{R} + \mathbb{E}(S^2) \sqrt{2\pi\mathcal{R}})^{\frac{1}{2}} + \mathbb{E}(S) (2\pi\mathcal{R})^{\frac{1}{4}} \right) \sqrt{T} e^{-\frac{1}{2}\lambda T} + 4\alpha \sqrt{\pi\mathcal{R}} \mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T} \end{aligned}$$

for  $T \geq \log(2)/\lambda$ .

*Proof:* Using

$$\Phi\left(\frac{1}{\sqrt{2\mathcal{R}}}\right) \leq 1,$$

the statement can easily be deduced from Theorem 7.1.5 and Lemma 7.3.10.  $\square$

Finally, we consider the case where  $\alpha$  depends on  $\mathcal{R}$  in a reasonable way for the reasons stated in Remark 7.3.4.

**7.3.13 Corollary**

In the one-dimensional case with  $\psi(x) = e^{-x^2/\mathcal{R}}$ , we set  $\alpha = (\pi\mathcal{R})^{-\frac{1}{2}}$ . For  $T \geq \log(2)/\lambda$ , we then have

$$\begin{aligned} & d_{TV}(\text{MixPo}(\Lambda_T), \text{MixPo}(\text{M})) \\ & \leq \frac{\sqrt{32}}{\lambda} \left( \left( \sigma_S^2 + \frac{\sqrt{2}\mathbb{E}(S^2)}{\sqrt{\pi\mathcal{R}}} \right)^{\frac{1}{2}} + \sqrt{2}\mathbb{E}(S)(\pi\mathcal{R})^{-\frac{1}{4}} \right) \sqrt{T} e^{-\frac{1}{2}\lambda T} + 4\mathbb{E}(S) \left( 2T + \frac{\lambda}{\beta(\beta + \lambda)} \right) e^{-\lambda T}. \end{aligned}$$

**7.3.14 Remark**

Of course we could consider more than one dimension, analogously. In this case, we could for example apply the density of the multivariate normal distribution.



## 8 Concluding remarks

We end this thesis by summarizing the most important contributions of this thesis in a few sentences, discussing our findings briefly and giving a short outlook.

In this thesis, we mainly focused on a loop-free version of the original dynamic random network model by Britton and Lindholm. We placed the model among other, well-known network models. We found several interesting similarities to other models and pointed out differences. We motivated its applicability, in particular by proving that its asymptotic degree distribution can exhibit power law behaviour.

In large parts of this thesis, we were concerned with the derivation of convergence rates for the degree distribution in this model. As a by-product, we have obtained a rigorous proof of the convergence itself. We have proved several results along the way that might also be of interest in other contexts. Most remarkably, we found the age distribution of a linear birth and death process. This result might be very useful for applications since the linear birth and death process is a very popular population model and the age is a very natural and interesting quantity. In order to prove an explicit formula for the cumulative distribution function of the age, we applied contour processes. Some of the techniques that we used might also be helpful for the derivation of related results.

For the pure birth case with deterministic social indices, this network model can be seen as a variant of a continuous-time uniform attachment model. For the discrete-time uniform attachment model, we have the convergence rate  $O(1/n)$  (cf. Theorem 5.4.17). Since  $n$  is the number of nodes, we would thus expect the rate  $O(\mathbb{E}(1/Y_T))$  for the pure birth case with deterministic social indices. Indeed, since we have  $\sigma_S = 0$  for deterministic social indices, we obtain a rate of this order by Theorem 6.3.1.

Since we have the convergence rate  $O(\log(n)/n)$  for the classical discrete-time preferential attachment models (cf. Theorem 5.4.4) and the expected population size  $\mathbb{E}(Y_t)$  increases exponentially in  $t$ , it is plausible that we still obtain an exponential rate if the social indices are random.

Note that rates for the convergence towards the asymptotic degree distribution in models that include fitness do not appear to have been considered before. Those rates can be very interesting for applications since they allow us to judge the approximation of the degree distribution at finite time  $t$  by the asymptotic one, which is considerably easier to deal with, for arbitrary  $t$ .

Note further that we have the convergence rate  $O(1/n)$  also for a sequence of Erdős-Rényi graphs by Remark 5.3.3 corresponding to the rate  $O(\mathbb{E}(1/Y_T))$  for the pure birth case with deterministic social indices in the (loop-free version of the) Britton-Lindholm model. For the modified generalized random graph model with independent and identically distributed non-deterministic weights, we obtained

the rate  $O(\sqrt{1/n})$ , which corresponds to the rate  $O(\mathbb{E}(\sqrt{1/Y_T}))$  for the pure birth case with non-deterministic social indices.

In Chapter 7, we also obtained exponential rates for an extension of the model, where positions of the nodes were considered in the attachment procedure.

Some of the techniques used in this thesis in order to find the convergence rates might also be useful for extended or completely different network models. In particular, it would be interesting to derive rates for a model that allows a combination of fitness and preferential attachment.

Apart from that, the model could be extended in various ways. Since many real-world networks are very complex, one might want to consider different types, e.g. sexes, of nodes or covariates that influence the attachment procedure. It would also be interesting to consider dependent social indices since in some real-world networks popularity is partly influenced by neighbours or inherited.

In the spatial version, uniformly distributed positions of nodes might not be particularly realistic in many contexts. One could use a more complex point process instead. Furthermore, the spatial model could of course be extended to the case where nodes can die.

Finally, it would of course be interesting to apply (the loop-free version of) the model by Britton and Lindholm to data from real-world networks.

# A Appendix

## A.1 R code for the introduction

### Simulation of the Britton Lindholm model without loops

```
library(igraph);
qpareto=function(u, a, b) b/(1-u)^(1/a)
rpareto=function(n, a, b) qpareto(runif(n),a,b) #Pareto distribution
lambda=1; alpha=3; beta=1;
M=1;
T=5;
degrees =rep(0,M);
for(m in seq(1, M,by=1))
{
  E=vector();
  S=list();
  Y=1;
  S=1;
  tbirth=0;
  while (tbirth[length(tbirth)]<T)
  {
    h=rexp(1,1)/(Y*lambda);
    tbirth=c(tbirth,tbirth[Y]+h);
    if (tbirth[Y+1]<T)
    {
      Y=Y+1;
      S=c(S,1);
      if(Y>2)
      {
        for(k in seq(1, Y-1,by=1))
        {
          number_edges=rpois(1, alpha*S[k]*(tbirth[Y]-tbirth[Y-1]));
          for(j in seq(0, number_edges,by=1))
          {
            K=sample(setdiff(1:(Y-1),k))[1];
            if (j>0 )
```

```

        {
            edge_birth_time=runif(1,tbirth[Y-1], tbirth[Y]);
            if(T-edge_birth_time < (rexp(1,1)/beta))
                {E=c(E,c(K,k));}
        }
    }
}
}
else #last interval
{
if(Y>2)
{
for(k in seq(1, Y,by=1))
{
number_edges=rpois(1, alpha*S[k]*(T-tbirth[Y]));
for(j in seq(0, number_edges,by=1))
{
K=sample(setdiff(1:Y,k))[1];
if (j>0 )
{
edge_birth_time=runif(1,tbirth[Y], T);
if(T-edge_birth_time < (rexp(1,1)/beta))
    {E=c(E,c(K,k));}
}
}
}
}
}
}
g=graph( E, n=Y );
}
plot.igraph(as.undirected(g, mode="each"), vertex.size=3*S, vertex.label=NA)

```

## A.2 R code for Chapter 4

### Simulation of the contour process

```

n=500
T=1.5
lambda=2
mu=1

```

```
# The ages of the first Y_T-1 individuals
rages=vector()
for(i in 1:n)
{
  ages=vector()
  x=0
  end=0
  slope=1
  while(end==0)
  {
    if(slope==1)
    {
      xn=x+rexp(1,mu)
      slope=-1
      if(xn>T)
        {x=T}
      else
        {x=xn}
    }
    else
    {
      xn=x-rexp(1,lambda)
      slope=1
      if(x==T)
        {ages=c(ages,T-xn)}
      if(xn<0)
      {
        x=0
        end=1
      }
      else
        {x=xn}
    }
  }
  if(length(ages)>1)
  {
    ages=ages[-length(ages)]
    rages=c(rages, ages)
  }
}
cdfexp=function(x) 1-(exp(-lambda*x)-exp(-(lambda-mu)*T)*exp(-mu*x))
```

```

                                                    / (1-exp(-(lambda-mu)*T))
plot(ecdf(rages),xlim=range(0:2),xlab='t',lwd=1.5,cex.lab=1.4,0cex.axis=1.2)
curve(cdfexp,add=TRUE , xlim=range(0:1.5), lwd=1.5, col='gray', from=0, to=T)

# The age of the last individual
n=5000
rages=vector()
for(i in 1:n)
{
  age=0
  x=0
  end=0
  slope=1
  while(end==0)
  {
    if(slope==1)
    {
      xn=x+rexp(1,mu)
      slope=-1
      if(xn>T)
        {x=T}
      else
        {x=xn}
    }
    else
    {
      xn=x-rexp(1,lambda)
      slope=1
      if(x==T)
        {age=min(T-xn,T)}
      if(xn<0)
      {
        x=0
        end=1
      }
      else
        {x=xn}
    }
  }
  if(age>0)
    {rages=c(rages, age)}
}

```



```

}
cdfexp=function(x) (lambda*(1-exp(-mu*x))-mu*(1-exp(-lambda*x)))/(lambda-mu)
plot(ecdf(rages),xlim=range(0:2),xlab='t',ylab=' ',main=NULL)
curve(cdfexp,add=TRUE,xlim=range(0:1.5), lwd=2, col='gray', from=0, to=T)

```

### Simulation of a linear birth and death process

```

lbd= function(a1,a2,x0,Tmax) #simulates a linear birth and death process
{
  a=a1+a2
  tdeath=rexp(1)/a
  x=rep(0,1)
  time=rep(0,1)
  x[1]=x0
  time[1]=0
  i=1
  while(time[length(time)]<Tmax )
  {i=i+1; tdeath=cbind(tdeath,rexp(1)/a);
  if (x[i-1]==0)
  {x=x[1:i-1]
  time=time[1:i-1]
  return(list(x=as.vector(x),time=as.vector(time)))
  }
  else
  {
  if(runif(1)>a1/a)
  {x=cbind(x,x[i-1]-1)}
  else
  {x=cbind(x,x[i-1]+1)}
  time=cbind(time,time[i-1]+tdeath[i-1]/x[i-1])
  }
  }
  x=x[1:length(x)-1]
  time=time[1:length(time)-1]
  return(list(x=as.vector(x),time=as.vector(time)))
}

```

### Simulation of the age distribution

```

N=5000
T=1.5
lambda=2; mu=1

```

```

ages=vector()
for(n in seq(1, N,by=1))
{
  p=lbd(lambda,mu,1,T)
  if(p$x[length(p$x)]>0)
  {
    s=1
    zg=1
    zt=0
    gzp=rep(0,length(p$x))
    tzp=rep(0,length(p$x))
    if(length(p$x)>1)
    {
      for(k in seq(2, length(p$x),by=1))
      {
        if (p$x[k]-p$x[k-1]==1)
        {
          zg=zg+1
          gzp[zg]=p$time[k]
          s=union(s,zg)
        }
        else
        {
          zt=zt+1
          tzp[zt]=p$time[k]
          r=sample(1:length(s),1,replace=TRUE)
          s=setdiff(s,s[r])
        }
      }
    }
    {
      age=rep(0,length(s))
      for(i in seq(1, length(s),by=1))
      {age[i]=T-gzp[s[i]]}
      ages=c(ages,age[sample.int(length(age), 1)])
    }
  }
}

cdfage=function(x) as.numeric(x > T)+as.numeric(x <= T)*((1-(lambda-mu)
/(lambda*exp((lambda-mu)*T)-lambda)*log((lambda*exp((lambda-mu)*T)-mu)/(lambda-mu)))
*(1-(exp(-lambda*x)-exp(-(lambda-mu)*T)*exp(-mu*x))/(1-exp(-(lambda-mu)*T)))
+ (lambda-mu)/(lambda*exp((lambda-mu)*T)-lambda)*log((lambda*exp((lambda-mu)*T)-mu)

```

```

/(lambda-mu))*(lambda*(1-exp(-mu*x))-mu*(1-exp(-lambda*x)))/(lambda-mu)
cdfexp=function(x) 1-exp(-lambda*x)
plot(ecdf(ages),xlim=range(0:2),xlab='t',ylab=' ',main=NULL)
curve(cdfage,add=TRUE,col='gray',xlim=range(0:1.5), from=0, to=T)
curve(cdfexp,add=TRUE,col='gray',xlim=range(0:2), from=0, to=2.5,lty=3)

```

## A.3 R code for Chapter 6

### Simulation of the convergence rate

```

#inverse of the cdf of the Exp(lambda) distribution truncated at t:
itexp=function(u, m, t) { -log(1-u*(1-exp(-t*m)))/m }
#Exp(lambda) distribution truncated at t:
rtexp=function(n, lambda, t) { itexp(runif(n), lambda, t) }
#inverse of the cdf of A_2 given Y_T=yt:
inv= function(u, yt, lambda, t) {log(1-u^(1/(yt-1))*(1-exp(lambda*t)))/lambda}
#distribution of A_2 given Y_T=yt, i.e. of the minimum of yt-1 iid truncated
#Exp(lambda) distributed random variables:
ra2=function(n, yt, lambda, t) { inv(runif(n), yt, lambda, t) }
alpha=1; beta=1; my.lambda=1
my.mu=0 #constant
my.start=1
maxT=7
step=0.125
E=rep(0,maxT/step)
n=500
logtv=seq(step, maxT,by=step)
#Asymptotic distribution
S=1
Z=rexp(n,my.lambda)
Z=1-exp(-beta*Z)
M=alpha/beta*2*Z
asymptotic_degree=rpois(n,M)
for(T in seq(step, maxT,by=step) )
{
  logq=log(exp(my.lambda*T)-1)-my.lambda*T #Y_T~Geo(1-q)
  degree=vector()
  for (k in seq(1, n,by=1) )
  {
    state=as.numeric(ceiling(-rexp(1,1)/logq))#~Geo(1-q); here we use
    #ceiling(X)~Geo(p) for X~Exp(-log(1-p))
    while(is.na(state))

```

```

    {state=as.numeric(ceiling(-rexp(1,1)/logq))}
  J_T=ceiling(runif(1,0,state))
  if(J_T>1)
  {
    age=rtexp(1, my.lambda, T) #age ist am Ende A_max(J_T,2)
  }
  else if(state>1)
  {
    age=ra2(1, state, my.lambda, T)
  }
  if(state==1)
  {Lambda_T=0}
  else
  {Lambda_T= 2*alpha/beta*(1-exp(-beta*age))}
  if(Lambda_T==0)
  {degree=c(degree,0)}
  else
  {degree=c(degree,rpois(1,Lambda_T))}
}

sum4=0
N=50
for(k in seq(0, N,by=1) )
{sum4=sum4+abs(sum(degree==k)/n-sum(asymptotic_degree==k)/n)}
logtv[T/step]= log(sum4/2)/(-my.lambda*T)
E[T/step]=sum4/2
}

onef=function(x) 1
plot(logtv,xaxt="n",xlab="T",)
axis(1,at=0:(maxT/step),labels=seq(0, maxT,by=step))
axis(2,at=1,labels=1)
onefv=Vectorize(onef)
curve(onefv,add=TRUE, lwd=1.5, col='gray',lty=2,from=0)

```

### Simulation of $\mathbb{E}(|\Lambda_T - M|)$

```

itexp=function(u, m, t) { -log(1-u*(1-exp(-t*m)))/m }
rtexp=function(n, lambda, t) { itexp(runif(n), lambda, t) }
inv= function(u, yt, lambda, t) {log(1-u^(1/(yt-1))*(1-exp(lambda*t)))/lambda}
ra2=function(n, yt, lambda, t) { inv(runif(n), yt, lambda, t) }
alpha=1

```

```

my.lambda=1
my.mu=0 #constant
my.start=1
beta=1
maxT=6.5
step=0.125
n=250000
E=vector()
for(T in seq(step, maxT,by=step) )
{
  print(T)
  M=vector()
  Lambda_T=vector()
  logq=log(exp(my.lambda*T)-1)-my.lambda*T #Y_T~Geo(1-q)
  for (k in seq(1, n,by=1) )
  {
    prob=1-(exp(my.lambda*T)-1)/exp(my.lambda*T) #Y_T~Geo(prob)
    state=rgeom(1, prob)+1
    if(state==1)
    {times=0}
    else
    {
      times=T-rtexp(state-1, my.lambda, T)
      times=c(0,sort(times))
    }
    S=rexp(state,rate=1)
    J_T=sample(1:state,1,replace=TRUE)
    sum1=sum(S[-J_T])/(state-1)*(1-exp(-beta*(T-times[length(times)])))
    if(state<3)
    {
      sum2=0
    }
    else
    {
      factr=(exp(-beta*(T-times[-c(1,2)]))- exp(-beta*(T-times[-c(1,state)])))
        /(1:(state-2))
      summands=outer(S,factr)
      which=outer(1:state,2:(state-1),function(i,l){pmax(i,J_T)<=l&i !=J_T})
      sum2=sum(summands[which])
    }
    if(state==1)

```

```

{
  Lambda_T=c(Lambda_T,0)
  M=c(M,0)
}
else
{
  age=T-times[max(2,J_T)]
  Lambda_T= c(Lambda_T,alpha/beta*S[J_T]*(1-exp(-beta*age))
              +alpha/beta*sum1+alpha/beta*sum2)
  Z= -log(1-(1-exp(-my.lambda*age))/(1-exp(-my.lambda*T)))/my.lambda
  M=c(M,alpha/beta*(S[J_T]+1)*(1-exp(-beta*Z)))
}
}
E=c(E,sum(abs(M-Lambda_T))/n)
#E=c(E,sum(abs(sqrt(M)-sqrt(Lambda_T)))/n)
}
logE= log(E)/(-my.lambda*seq(step, maxT,by=step))
onef=function(x) 1/2
plot(logE,ylim= c(0,3),xaxt="n",xlab="T")
axis(1,at=0:(maxT/step),labels=seq(0, maxT,by=step))
axis(2,at=1/2,labels=1/2)
onefv=Vectorize(onef)
curve(onefv,add=TRUE, lwd=1.5, col='gray',lty=2,from=0)

```

### Simulation of (6.79)

```

itexp=function(u, m, t) { -log(1-u*(1-exp(-t*m)))/m }
rtexp=function(n, lambda, t) { itexp(runif(n), lambda, t) }
inv= function(u, yt, lambda, t) {log(1-u^(1/(yt-1))*(1-exp(lambda*t)))/lambda}
ra2=function(n, yt, lambda, t) { inv(runif(n), yt, lambda, t) }
alpha=1; beta=1; my.lambda=1; my.mu = 0
my.start = 1
maxT=7
step=0.25
n=500000
E=vector()
S=rexp(n,rate=1)
asymptotic_degree=vector()
Z=rexp(n,my.lambda)
Z=1-exp(-beta*Z)
M=alpha/beta*(1+S)*Z
for(T in seq(step, maxT,by=step) )

```

```

{
  Lambda_T=vector()
  for (k in seq(1, n,by=1) )
  {
    prob=1-(exp(my.lambda*T)-1)/exp(my.lambda*T) #Y_T~Geo(prob)
    state=rgeom(1, prob)+1
    print(k)
    if(state==1)
    {times=0}
    else
    {
      times=T-rtexp(state-1, my.lambda, T)
      times=c(0,sort(times))
    }
    S=rexp(state,rate=1)
    J_T=sample(1:state,1,replace=TRUE)
    sum1=sum(S[-J_T])/(state-1)*(1-exp(-beta*(T-times[length(times)])))
    if(state<3)
    {
      sum2=0
    }
    else
    {
      factr=(exp(-beta*(T-times[-c(1,2)])))
        - exp(-beta*(T-times[-c(1,state)])))/(1:(state-2))
      summands=outer(S,factr)
      which=outer(1:state,2:(state-1),function(i,l){pmax(i,J_T)<=l&i!= J_T})
      sum2=sum(summands[which])
    }
    if(state==1)
    {
      Lambda_T=c(Lambda_T,0)
    }
    else
    {
      age=T-times[max(2,J_T)]
      Lambda_T= c(Lambda_T,alpha/beta*S[J_T]*(1-exp(-beta*age))
        + alpha/beta*sum1+alpha/beta*sum2)
    }
  }
}
E=c(E,mean(abs(sort(Lambda_T)-sort(M))))

```

```

}
logE= log(E)/(-my.lambda*seq(step, maxT,by=step))
onef=function(x) 1
plot(logE,xaxt="n",xlab="T")
axis(1,at=0:(maxT/step),labels=seq(0, maxT,by=step))
axis(2,at=1,labels=1)
onefv=Vectorize(onef)
curve(onefv,add=TRUE, lwd=1.5, col='gray',lty=2,from=0)

```

## A.4 R code for Chapter 7

### Simulation of a spatial model

```

library(igraph);
d=function(x,y) min(abs(x-y), abs(x+1-y), abs(y+1-x))
psi=function(x,y) 1-2*d(x,y)
alpha=3; beta=1; lambda=1;
M=1;
l=1;
T=4;
degrees =rep(0,M);
for(m in seq(1, M,by=1)) {
  E=vector();
  S=list();
  Pos1= runif(1,0,1);
  N=c(1,Pos1);
  Y=length(N)/2;
  S=rpareto(1,3,2/3);
  tbirth=0;
  while (tbirth[length(tbirth)]<T)
  {
    h=rexp(1,1)/(Y*lambda);
    tbirth=c(tbirth,tbirth[Y]+h);
    if(tbirth[length(tbirth)-1]<3*T/4 && tbirth[length(tbirth)]>3*T/4)
    {N1=N
      Y1=Y
      E1=E}
    if (tbirth[Y+1]<T)
    {
      Posi= runif(1,0,1);
      Y=Y+1;
      N=c(N,c(Y,Posi));
    }
  }
}

```



```

S=c(S,rpareto(1,3,2/3));
if(Y>2)
{
  for(k in seq(1, Y-1,by=1))
  {number_edges=rpois(1, alpha*S[k]*(tbirth[Y]-tbirth[Y-1]));
  for(j in seq(0, number_edges,by=1))
  {K=sample(setdiff(1:(Y-1),k))[1];
  if (j>0 && (psi(N[2*K],N[2*k]) >= runif(1,0,1)))
  { print(psi(N[2*K],N[2*k]))
  print(N[2*K])
  print(N[2*k])
  edge_birth_time=runif(1,tbirth[Y-1], tbirth[Y]);
  if(T-edge_birth_time < (rexp(1,1)/beta))
  {E=c(E,c(K,k));}
  }
  }
}
}
else #last interval
{if(Y>2)
{
  for(k in seq(1, Y,by=1))
  {number_edges=rpois(1, alpha*S[k]*(T-tbirth[Y]));
  for(j in seq(0, number_edges,by=1))
  {K=sample(setdiff(1:Y,k))[1];
  if (j>0 && (psi(N[2*K],N[2*k]) >= runif(1,0,1)))
  {
    print(psi(N[2*K],N[2*k]))
    edge_birth_time=runif(1,tbirth[Y], T);
    if(T-edge_birth_time < (rexp(1,1)/beta))
    {E=c(E,c(K,k));}
  }
  }
}
}
}
}
g1=graph( E1, n=Y1 );
g=graph( E, n=Y );
}

```

```
M=matrix( , nrow=Y, ncol=2, byrow=TRUE)
for(i in seq(1, Y,by=1))
{M[i,1]=sin(N[2*i]*(2*pi)); M[i,2]=cos(N[2*i]*(2*pi)); }
M1=matrix( , nrow=Y1, ncol=2, byrow=TRUE)
for(i in seq(1, Y1,by=1))
{M1[i,1]=sin(N1[2*i]*(2*pi)); M1[i,2]=cos(N1[2*i]*(2*pi)); }
plot.igraph(as.undirected(g1,mode="each"),vertex.size=S[1:Y1]*4,layout=M1,
vertex.label=NA)
plot.igraph(as.undirected(g,mode="each"),vertex.size=S*4,layout=M, vertex.label=NA)
```

# Bibliography

- [AB02] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47, 2002.
- [All10] Linda J S Allen. *An introduction to stochastic processes with applications to biology*. CRC Press, Upper Saddle River, 2010.
- [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 T J Bailey. *The elements of stochastic processes with applications to the natural sciences*. Wiley, New York, 1964.
- [Bar11] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [Bar13] Albert-László Barabási. Network science. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 371(1987):20120375, 2013.
- [BB01] Ginestra Bianconi and Albert-László Barabási. Competition and multiscaling in evolving networks. *EPL (Europhysics Letters)*, 54(4):436, 2001.
- [BDML06] Tom Britton, Maria Deijfen, and Anders Martin-Löf. Generating Simple Random Graphs with Prescribed Degree Distribution. *Journal of Statistical Physics*, 124(6):1377–1397, 2006.
- [BHJ92] A D Barbour, Lars Holst, and Svante Janson. *Poisson approximation*, volume 2 of *Oxford Studies in Probability*. Oxford University Press, New York, 1992.
- [BL09] Tom Britton and Mathias Lindholm. The early stage behaviour of a stochastic SIR epidemic with term-time forcing. *Journal of Applied Probability*, 46(4):975–992, 2009.
- [BL10] Tom Britton and Mathias Lindholm. Dynamic random networks in dynamic populations. *Journal of Statistical Physics*, 139(3):518–535, 2010.
- [BLM<sup>+</sup>06] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.
- [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.
- [BPS12] Mamadou Ba, Etienne Pardoux, and Ahmadou B Sow. Binary trees, exploration processes, and an extended Ray-Knight theorem. *Journal of Applied Probability*, 49(1):210–225, 2012.

- [BRS<sup>+</sup>01] Béla Bollobás, Oliver Riordan, Joel Spencer, Gábor Tusnády, et al. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.
- [CCRM02] Guido Caldarelli, Andrea Capocci, Paolo Rios, and Miguel A Munoz. Scale-free networks without growth or preferential attachment: Good get richer. *arXiv preprint cond-mat/0207366*, 2002.
- [CFV04] Colin Cooper, Alan Frieze, and Juan Vera. Random deletion in a scale-free random graph process. *Internet Mathematics*, 1(4):463–483, 2004.
- [CHK<sup>+</sup>01] Duncan S Callaway, John E Hopcroft, Jon M Kleinberg, Mark E J Newman, and Steven H Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64(4):041902, 2001.
- [CJ10] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [CM77] David R Cox and Hilton D Miller. *The theory of stochastic processes*, volume 134. CRC Press, Boca Raton, Florida, 1977.
- [CS12] J Theodore Cox and Rinaldo B Schinazi. A branching process for virus survival. *Journal of Applied Probability*, 49(3):888–894, 2012.
- [CSN09] Aaron Clauset, Cosma Rohilla Shalizi, and Mark E J Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [DHH10] Sander Dommers, Remco van der Hofstad, and Gerard Hooghiemstra. Diameters in preferential attachment models. *Journal of Statistical Physics*, 139(1):72–107, 2010.
- [Die06] Reinhard Diestel. *Graph Theory*, volume 173 of *Graduate texts in mathematics*. Springer, Berlin, Heidelberg, 2006.
- [DM03] Sergey N Dorogovtsev and José F F Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW (Physics)*. Oxford University Press, Oxford, 2003.
- [DM10] Moez Draief and Laurent Massouli. *Epidemics and rumours in complex networks*. Cambridge University Press, New York, 2010.
- [Eis08] Bennett Eisenberg. On the expectation of the maximum of iid geometric random variables. *Statistics & Probability Letters*, 78(2):135–143, 2008.
- [ER59] Paul Erdős and Alfréd Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.
- [ER02] Güler Ergün and Geoff J Rodgers. Growing random networks with fitness. *Physica A: Statistical Mechanics and its Applications*, 303(1–2):261–272, 2002.
- [Fel43] William Feller. On a general class of "contagious" distributions. *The Annals of mathematical statistics*, 14(4):389–400, 1943.

- [Fel67] William Feller. *An introduction to probability theory and its applications*. Wiley, New York, 1967.
- [FKZ11] Sergey Foss, Dmitry Korshunov, and Stan Zachary. *An introduction to heavy-tailed and subexponential distributions*. Springer, 2011.
- [Fre74] David Freedman. The Poisson approximation for dependent events. *The Annals of Probability*, pages 256–269, 1974.
- [Ger08] Tanja Gernhard. The conditioned reconstructed process. *Journal of theoretical biology*, 253(4):769–778, 2008.
- [Gil59] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 12 1959.
- [GKST10] Shilpa Ghadge, Timothy Killingback, Bala Sundaram, and Duc A Tran. A statistical construction of power-law networks. *International Journal of Parallel, Emergent and Distributed Systems*, 25(3):223–235, 2010.
- [Grz06] Peter Grzybek. *Contributions to the science of text and language: word length studies and related issues*, volume 31. Springer Science & Business Media, 2006.
- [Har50] Theodore Edward Harris. Some mathematical models for branching processes. Technical report, No. P-152, RAND Corporation, Santa Monica, 1950.
- [Har02] Theodore Edward Harris. *The Theory of Branching Processes*. Dover Publications, Mineola, New York, 2002.
- [HJV05] Patsy Haccou, Peter Jagers, and Vladimir A Vatutin. *Branching processes: variation, growth, and extinction of populations*. Cambridge University Press, Cambridge, 2005.
- [HW13] Edward J S Hearnshaw and Mark M J Wilson. A complex network approach to supply chain network theory. *International Journal of Operations & Production Management*, 33(4):442–469, 2013.
- [Jag75] Peter Jagers. *Branching processes with biological applications*. Wiley, New York, 1975.
- [JKK05] Norman L Johnson, Adrienne W Kemp, and Samuel Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. Wiley, 2005.
- [Kei75] Niels Keiding. Maximum likelihood estimation in the birth-and-death process. *The Annals of Statistics*, pages 363–372, 1975.
- [Ken49] David G Kendall. Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):230–282, 1949.
- [KSR08] Joseph S Kong, Nima Sarshar, and Vwani P Roychowdhury. Experience versus talent shapes the structure of the web. *Proceedings of the National Academy of Sciences*, 105(37):13724–13729, 2008.
- [Küc12] Fabian Kück. Dynamik zufälliger Netzwerke. Master’s thesis, University of Hamburg, 2012.

- [KX05] Dimitris Karlis and Evdokia Xekalaki. Mixed poisson distributions. *International Statistical Review*, 73(1):35–58, 2005.
- [Lam08] Amaury Lambert. Population dynamics and random genealogies. *Stochastic Models*, 24(S1):45–163, 2008.
- [LPW09] David Asher Levin, Yuval Peres, and Elizabeth Lee Wilmer. *Markov chains and mixing times*. American Mathematical Soc., 2009.
- [MGN06] Cristopher Moore, Gourab Ghoshal, and Mark E J Newman. Exact solutions for models of evolving networks with addition and deletion of nodes. *Physical Review E*, 74(3):036121, 2006.
- [Mil67] Stanley Milgram. The small-world problem. *Psychology Today*, 1(1), 1967.
- [New10] Mark E J Newman. *Networks: An Introduction*. Oxford University Press, 2010.
- [New13] Mark E J Newman. Power laws, Pareto distributions and Zipf’s law. *Cities*, 30:59–67, 2013.
- [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.
- [Pen03] Mathew Penrose. *Random Geometric Graphs*. Oxford studies in probability. Oxford University Press, Oxford, 2003.
- [Pen15] Mathew Penrose. Inhomogeneous random graphs, isolated vertices, and Poisson approximation. *arXiv preprint arXiv:1507.07132*, 2015.
- [Per98] Richard Perline. Mixed Poisson Distributions Tail Equivalent to their Mixing Distributions. *Statistics & Probability Letters*, 38(3):229–233, 1998.
- [PPS15] Angelica Pachon, Federico Polito, and Laura Sacerdote. Random graphs associated to some discrete and continuous time preferential attachment models. *arXiv preprint arXiv:1503.06150*, 2015.
- [PRR13] Erol A Peköz, Adrian Röllin, and Nathan Ross. Total variation error bounds for geometric approximation. *Bernoulli*, 19(2):610–632, 2013.
- [Rei09] Gesine Reinert. *Barabási-Albert random graphs, scale-free distributions and bounds for approximation through Stein’s method*. dissertation, University of Oxford, 2009.
- [Rei12] Rolf-Dieter Reiss. *Approximate distributions of order statistics: with applications to non-parametric statistics*. Springer Science & Business Media, 2012.
- [Rig16] Michel Rigo. *Advanced Graph Theory and Combinatorics*. Iste Series. Wiley, 2016.
- [RKSF13] Svetlozar T Rachev, Lev Klebanov, Stoyan V Stoyanov, and Frank Fabozzi. *The methods of distances in the theory of probability and statistics*. Springer Science & Business Media, 2013.
- [Roo03] Bero Roos. Improvements in the Poisson approximation of mixed Poisson distributions. *Journal of Statistical Planning and Inference*, 113(2):467–483, 2003.

- [SHR13] Igor E Smolyarenko, Konrad Hoppe, and Geoff J Rodgers. Network growth model with intrinsic vertex fitness. *Physical Review E*, 88(1):012805, 2013.
- [SKBD13] Tanja Stadler, Denise Kühnert, Sebastian Bonhoeffer, and Alexei J Drummond. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences*, 110(1):228–233, 2013.
- [Sni01] Tom A B Snijders. The statistical evaluation of social network dynamics. *Sociological Methodology*, 31(1):361–395, 2001.
- [SR04] Nima Sarshar and Vwani Roychowdhury. Scale-free and stable structures in complex ad hoc networks. *Physical Review E*, 69(2):026101, 2004.
- [SvdBS10] Tom A B Snijders, Gerhard G van de Bunt, and Christian E G Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, 2010.
- [TP11] My T Thai and Panos M Pardalos. *Handbook of optimization in complex networks: theory and applications*, volume 57. Springer Science & Business Media, 2011.
- [vdH15] Remco van der Hofstad. *Random Graphs and Complex Networks, Vol. II*. Lecture Notes, <http://www.win.tue.nl/~rhofstad/NotesRGCNII.pdf>, 2015.
- [vdH16] Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 43 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2016.
- [VR07] Fernando Vega-Redondo. *Complex social networks*. Number 44. Cambridge University Press, 2007.
- [Wau58] WA O’N Waugh. Conditioned Markov processes. *Biometrika*, 45(1-2):241–249, 1958.
- [Wil90] Gordon E Willmot. Asymptotic Tail Behaviour of Poisson Mixtures with Applications. *Advances in Applied Probability*, 22(1):147, 1990.
- [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*, volume 1. World Scientific, Singapore, New Jersey, London, Hong Kong, 1995.
- [Yul25] G Udney Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 213(402):21–87, 1925.
- [Zol84] Vladimir M Zolotarev. Probability metrics. *Theory of Probability & Its Applications*, 28(2):278–302, 1984.

# List of Symbols

|                             |   |
|-----------------------------|---|
| $\wedge$                    | minimum operator  |
| $\vee$                      | maximum operator  |
| $f \sim g$                  | tail equivalence  |
| $\mathbb{1}$                | indicator function  |
| $\alpha$                    | a positive constant related to the rates at which edges are created                       |
| $A$                         | $\text{Exp}(\lambda)$ distributed random variable   |
| $A_i(T)$                    | age of node $i$ at time $T$ in the linear birth and process $(Y_t)_{t \geq 0}$            |
| $\beta$                     | death rate of the edges   |
| $B$                         | beta function   |
| $\mathfrak{B}_T$            | number of births up to time $T$ in the linear birth and process $(Y_t)_{t \geq 0}$        |
| $\text{Be}$                 | Bernoulli distribution  |
| $\mathcal{C}$               | a $m$ -dimensional unit hypercube   |
| $\mathfrak{D}_T$            | number of deaths up to time $T$ in the linear birth and process $(Y_t)_{t \geq 0}$        |
| $\stackrel{\mathcal{D}}{=}$ | equal in distribution   |
| $D_T$                       | degree of a node picked uniformly at random at time $T$                                   |
| $d_{TV}$                    | total variation distance  |
| $d_W$                       | Wasserstein distance  |
| $\mathcal{E}$               | edge set/multiset   |
| $\mathbb{E}^*$              | conditional expected value defined by $\mathbb{E}^*(\cdot) = \mathbb{E}(\cdot   Y_T > 0)$ |
| $\Gamma$                    | gamma function  |



---

|                                     |  |
|-------------------------------------|--|
| Geo                                 | geometric distribution on $\mathbb{N}$   |
| $\mathcal{H}$                       | a set of piecewise linear functions  |
| $(H_x)_{x \geq 0}$                  | the (deflected) contour process  |
| $J_T$                               | node/individual picked uniformly at random at time $T$   |
| $\kappa(T)$                         | defined by $\kappa(T) = \lfloor e^{\frac{3}{2}(\lambda+\mu)T} \rfloor$ .                               |
| $\mathcal{K}(T)$                    | defined by $\mathcal{K}(T) = \max\{k : T_k < T/2\}$  |
| $\lambda$                           | per-capita birth rate of the linear birth and process $(Y_t)_{t \geq 0}$                               |
| $\Lambda_T$                         | random variable such that $\mathcal{L}(\Lambda_T   Y_T > 0) = \mathcal{L}(\Lambda_T^*)$                |
| $\Lambda_T^*$                       | defined such that $\text{MixPo}(\Lambda_T^*)$ is the degree distribution in the Britton Lindholm model |
| $\mathcal{L}(X)$                    | probability law of a random variable $X$   |
| $L$                                 | lifetime of an individual  |
| $\mathcal{M}_T$                     | number of events up to time $T$ in the linear birth and process $(Y_t)_{t \geq 0}$                     |
| $\mu$                               | per-capita death rate of the linear birth and process $(Y_t)_{t \geq 0}$                               |
| $M$                                 | random variable such that $\mathcal{L}(M   Y_T > 0) = \mathcal{L}(M^*)$                                |
| $M^*$                               | defined such that $\text{MixPo}(M^*)$ is the asymptotic degree distribution in Chapter 6 and 7         |
| MixPo                               | mixed Poisson distribution   |
| $\mathbb{N}$                        | the set of positive integers   |
| $\mathbb{N}_0$                      | the set of non-negative integers   |
| $(\Omega, \mathcal{A}, \mathbb{P})$ | a probability space  |
| $O, o$                              | Landau symbols   |
| $\phi$                              | bijection between $\mathcal{H}$ and $\mathcal{T}$ .  |
| $\varphi$                           | a slowly varying function  |
| $\psi$                              | function defining the spatial component of the edge process in the spatial model                       |
| $p_n(t)$                            | probability mass function of $Y_t$   |

|                                    |  |
|------------------------------------|--|
| $\mathfrak{P}(\mathcal{X})$        | set of probability measures on a measurable space $(\mathcal{X}, \mathcal{B})$   |
| $\mathcal{P}(i)$                   | position of node $i$ in the spatial model  |
| $\mathbb{P}^*$                     | probability measure defined by $\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot   Y_T > 0)$   |
| $\mathbb{P}^X$                     | distribution of the random variable $X$  |
| $P_{\lambda, \mu, T}$              | probability law of the contour process with parameters $\lambda$ and $\mu$ deflected at $T$  |
| $Q_{\lambda, \mu, T}$              | probability law of the tree corresponding to $(Y_t)_{t \geq 0}$ up to time $T$   |
| $\text{Pareto}(\tilde{\theta}, b)$ | Pareto distribution with parameters $\tilde{\theta}$ and $b$ having density $\frac{\tilde{\theta} b^{\tilde{\theta}}}{x^{\tilde{\theta}+1}} \mathbb{1}_{\{x \geq b\}}$ |
| $r$                                | function that maps the birth number to the event number  |
| $R_{T_l, T}$                       | number of nodes that are alive at time $T_l$ and survive up to time $T$ in $(Y_t)_{t \geq 0}$  |
| $\sigma(\cdot)$                    | generated $\sigma$ -algebra  |
| $\sigma_S$                         | standard deviation of the social index distribution  |
| $S$                                | random variable distributed according to the social index distribution   |
| $S_i$                              | social index of node $i$   |
| $\leq_{st}$                        | ordinary stochastic order  |
| $\mathcal{T}$                      | set of binary trees  |
| $T_l$                              | time of the $l$ th event in the linear birth and process $(Y_t)_{t \geq 0}$  |
| $T_i^+, T_i^-$                     | birth and death time of node $i$ , respectively, in the linear birth and process $(Y_t)_{t \geq 0}$  |
| $\mathcal{V}$                      | node set   |
| $\xrightarrow{w}$                  | weak convergence   |
| $(Y_t)_{t \geq 0}$                 | a linear birth and death process   |
| $Y_\infty$                         | defined by $Y_\infty = \lim_{t \rightarrow \infty} Y_t$  |

### Personal Information

Date of birth 26 June 1987  
Place of birth Hamburg  
Nationality German

### Education

Since **Doctoral Student in Mathematical Science**, *University of Göttingen*.  
04/2013 Supervisor: Prof. Dr. Dominic Schuhmacher  
09/2010– **Master in Mathematics**, *University of Hamburg*.  
03/2013  
08/2010– **Exchange student**, *Lund University*.  
01/2011  
10/2007– **Bachelor in Mathematics**, *University of Hamburg*.  
09/2010  
08/1998– **Abitur**, *Immanuel-Kant-Gymnasium in Hamburg*.  
07/2007

### Presentations at Conferences

03/2016 **The Age Distribution in a Linear Birth and Death Process**, *12th German Probability and Statistics Days*, Ruhr University Bochum.  
08/2014 **Rates of Convergence for Degree Distributions in a Dynamic Network Model**, *PhD Student Conference in Stochastics*, Martin Luther University of Halle-Wittenberg.  
03/2014 **Rates of Convergence for the Degree Distribution in a Dynamic Network Model**, *11th German Probability and Statistics Days*, Ulm University.  
08/2013 **Asymptotic Properties of the Dynamic Network Model by Britton and Lindholm**, *PhD Student Conference in Stochastics*, University of Göttingen.

### Workshops, Summer Schools and further Conferences Attended

06/2017 **NetSci**, *International School and Conference on Network Science*, Indianapolis.  
10/2016 **Workshop “What’s new in networks?”**, *University of Munich*.  
05/2015 **Summer School “Complex networks: theory, methods and applications”**, *Lake Como School of Advanced Studies*.