Electronic Journal of Probability

Vol. 1 (1996) Paper no. 14, pages 1–84.

Journal URL
http://www.math.washington.edu/˜ejpecp/
Paper URL
http://www.math.washington.edu/˜ejpecp/EjpVol1/paper14.abs.html

MULTIPLE SPACE-TIME SCALE ANALYSIS FOR INTERACTING BRANCHING MODELS

Donald A. Dawson

The Fields Institute, 222 College St., Toronto, Ontario, Canada M5T 3J12
ddawson@fields.utoronto.ca

Andreas Greven

Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstraße 1 1/2,
D-91054 Erlangen, Germany
greven@mi-uni.erlangen.de

Abstact: We study a class of systems of countably many linearly interacting diffusions whose components take values in $[0, \infty)$ and which in particular includes the case of interacting (via migration) systems of Feller's continuous state branching diffusions. The components are labelled by a hierarchical group. The longterm behaviour of this system is analysed by considering space-time renormalised systems in a combination of slow and fast time scales and in the limit as an interaction parameter goes to infinity. This leads to a new perspective on the large scale behaviour (in space and time) of critical branching systems in both the persistent and non-persistent cases and including that of the associated historical process. Furthermore we obtain an example for a rigorous renormalization analysis.

Keywords: Branching processes, interacting diffusions, super random walk, renormalization, historical processes

AMS subject classification: 60K35, 60J80

Submitted to EJP on September 22, 1995. Final version accepted on February 28, 1996.

Multiple space-time scale analysis for interacting branching models

Donald A. Dawson   Andreas Greven

Abstract

We study a class of systems of countably many linearly interacting diffusions whose components take values in $[0, \infty)$ and which in particular includes the case of interacting (via migration) systems of Feller's continuous state branching diffusions. The components are labelled by a hierarchical group. The longterm behaviour of this system is analysed by considering space-time renormalised systems in a combination of slow and fast time scales and in the limit as an interaction parameter goes to infinity. This leads to a new perspective on the large scale behaviour (in space and time) of critical branching systems in both the persistent and non-persistent cases and including that of the associated historical process. Furthermore we obtain an example for a rigorous renormalization analysis.

The qualitative behaviour of the system is characterised by the so-called interaction chain, a discrete time Markov chain on $[0, \infty)$ which we construct. The transition mechanism of this chain is given in terms of the orbit of a certain nonlinear integral operator. Universality classes of the longterm behaviour of these interacting systems correspond to the structure of the entrance laws of the interaction chain which in turn correspond to domains of attraction of special orbits of the nonlinear operator. There are two possible regimes depending on the interaction strength. We therefore continue in two steps with a finer analysis of the longtime behaviour. The first step focuses on the analysis of the growth of regions of extinction and the complementary regions of growth in the case of weak interaction and as time tends to infinity. Here we exhibit a rich structure for the spatial shape of the regions of growth which depend on the finer structure of the interaction but are universal in a large class of diffusion coefficients. This sheds new light on branching processes on the lattice. In a second step we study the family structure of branching systems in equilibrium in the case of strong interaction and construct the historical process associated with the interaction chain explicitly. In particular we obtain results on the number of different families per unit volume (as the volume tends to infinity). In addition we relate branching systems and their family structure (i.e. historical process) to the genealogical structure arising in Fleming-Viot systems. This allows us to draw conclusions on the large scale spatial distribution of different families in the limit of large times for both systems.

Key words: Branching processes, interacting diffusions, super random walk, renormalization, historical processes

AMS-Subject classification: 60 K 35, 60 J 80

D.A. Dawson : The Fields Institute, 222 College St.,
Toronto, Ontario, Canada M5T 3J12
A. Greven : Mathematisches Institut, Universität Erlangen-Nürnberg,
Bismarckstraße 1 1/2, D-91054 Erlangen, Germany

Table of Content

Part A Introduction and main results

0 Basic Multiple Space Time Scale Analysis

a) Motivation and background

In a series of papers Dawson and Greven (1993 a)b)c)); Dawson, Greven and Vaillancourt (1995), we developed a scheme, the *multiple space-time scale analysis*, with which it is possible to study the longterm behaviour of locally and hierarchically interacting systems each with *infinitely* many components in the limit of mean-field interaction but in large time scales. This technique will be used here to investigate some *long time - large scale phenomena* occurring in branching systems, including their associated *historical processes* and to study the relation between branching systems and systems with evolution by resampling (Fleming-Viot systems). The branching models we consider are the diffusion limits of branching random walks, often called super random walk.

The value of this type of analysis lies in the fact that it yields good predictions for the long time large scale behaviour of systems without passing to the mean field limit and the fact that it is possible to relate the *space-time renormalised system* in the mean field limit to a Markov chain on the state space of a single component the so-called *interaction chain*. This chain describes the large scale space-time dependence structure of the model. The transition kernels of the chain are given in terms of the *orbit* of a *nonlinear operator* in function space. The latter relation allows us to describe *universality classes* of the longterm behaviour of the original interacting systems by determining the domain of attraction of a special orbit of this nonlinear operator. The analytic properties of this operator will be studied in a forthcoming paper Baillon et al (1995) part 2, which in particular results in identifying the domains of attraction of the orbit given by a fixed point. In the case of branching models we are able to execute the whole analysis as well on the level of the historical processes which we construct for both the interacting system and the interaction chain.

Other models, such as interacting Fisher Wright diffusions resp. interacting Fleming-Viot processes, have been analysed along these lines in: Dawson and Greven (1993 c), Baillon et al. (1995) part 1, respectively Dawson, Greven and Vaillancourt (1995). In all these cases the analysis proceeds in a two step program, of which the first probabilistic part is devoted to the multiple space-time scale analysis and then a second analytic part in which the study of the nonlinear map is carried out.

Related renormalisation ideas occur in the study of certain population growth models (involving interaction of different families), (Durrett (1993), Durrett and Neuhauser (1993)). Here however renormalised systems are associated with deterministic systems namely partial differential equations of the reaction-diffusion type and the mean field limit of the system is obtained by superimposing fast stiring of the components.

In the papers listed above the multiple space-time scale analysis was mainly applied to systems with interacting components in which the components take values in a *compact* set. The models of interest in this context were related to ideas from population genetics. Interacting Fisher-Wright diffusions, and interacting (measure-valued) Fleming-Viot processes were the most important examples. The purpose of the present article is to extend the analysis to systems where the components take values in a noncompact set (here the interval $[0,\infty)$) and to include the *historical process* in the analysis.

The point here is twofold. First we identify some *new phenomena* due to the noncompactness, described later on, and second we obtain some insight into the behaviour of a class of interacting diffusions containing the important case of *interacting systems of Feller's continuous state branching diffusions*. Concerning the second point we note that although much is known about various spatial branching models on Z^d or $\mathbb{R}^d$, nevertheless some basic questions related to the clustering behaviour of "lowdimensional" models and of the equilibrium family structure (historical process) of the "highdimensional" models are still open (compare Shiga (1992), Gorostiza and Wakolbinger (1992)) since many papers on the subject deal with establishing the dichotomy clustering versus stability rather than giving a finer analysis of the two regimes. Some analogues of these open questions will be addressed here in the hierarchical context.

We want to go further and are interested in determining whether the so-called branching property in population growth models is essential or whether there are wider *universality classes* of similar longterm behaviour, this is where the multiple space-time scale analysis is particularly useful. Recall that the branching property says that every subfamily of the population evolves independently from any other subfamily.

Here are the three main features of our analysis for the particular class of processes studied here:

(i) It turns out that the longterm behaviour of an interacting diffusion system on $[0,\infty)^{\Omega_N}$, where Ω_N is a specific countable group, depends to 'first order' only on the properties of the interaction term, which is induced by migration, and not on the diffusive term which is induced by the population growth. In particular for strong interaction we have *stability* and for weak interaction the systems *cluster*, that is in the first case each possible value of the spatial population density corresponds to an equilibrium state, while in the second case as time goes on the mass of bigger and bigger groups of components of the system goes to zero and on a thin spatial set the components become very large. It is the second regime which exhibits new features compared to the compact case. The first case shows interesting similarities and even relations to the compact case. However, both regimes are studied further using the concept of the interaction chain mentioned earlier.

(ii) First it is the *formation of clusters* which we analyse further, that is we determine the rate of growth of growing components and the rate of spatial expansion of such a cluster of growing components. Depending on the form of the interaction, the spatial shape of the regions where growth occurs displays a complex behaviour, which we classify and analyse in great detail. The analysis allows us to give a complete description of the population growth on the thin set of components where the system is not yet extinct: This growth can be described by passing to a mass, space and time transformed process with values in $[0,\infty)$, where the deterministic time transformation depends on the interaction and the mass transformation on the universality classes of the diffusive term while the space transformation is just building a block average which involves only the structure of the group indexing the components. Then taking the limit of large times and mean field migration results in a limit process the so-called *cluster process*, which we determine explicitly. Related questions about clustering in systems with compact components are found in Cox and Griffeath (1986), Fleischmann and Greven (1994) and Klenke (1995) and in noncompact cases in Fleischman (1978), Durrett (1979).

We determine classes of diffusive terms, which show the same pattern in the formation of clusters. These *universality classes* are described in terms of the properties of the orbit of a certain nonlinear operation on a cone in a subspace of $C([0,\infty))$. The determination of the corresponding universality classes can be viewed an purely analytic problem and will be treated in Baillon et al. (1995).

(iii) In the stable case the analysis provides a simpler *approximate* description for the *equilibrium measures* for the original system (that is, before passing to the mean field limit). In the case of branching systems we show that the multiple space time scale analysis can be applied even on the level of the *historical process*, that is, the process encoding the whole genealogical structure of the system in equilibrium (see Dawson (1993) for a survey on the historical process). In other words we construct the historical process associated with the interaction chain. In particular we obtain the hierarchical mean field approximation of *Kallenberg's backward tree* and hence the *Palm measure*. This allows us to study, the density of a single family and the number of different families per volume which make up equilibrium states. The most important point is the connection and quantitative relations we find between the family structure of the *historical process of branching systems*, respectively, the *genealogical structure* in *Fleming-Viot systems*.

The rest of section 0 contains the description of the models in 0b) and of our basic results (see (i) above) formulated in Theorems 1 - 4 in Subsections 0c) - 0f). Section 1 in Part A contains a finer analysis of the long time behaviour of the system presented. It is here the phenomena which are new to the noncompact component space are presented in Theorems 5-12. The proofs are in Part B Sections 2-6.

b) The model

We shall consider Markov processes $X^N(t)$ with state space in $[0,\infty)^{\Omega_N}$, where Ω_N is the following countable abelian group defined for every $N \in I\!N$ with $N \geq 2$:

$$(0.1)\qquad \begin{aligned} \Omega_N &= \{\xi_1, \xi_2, \cdots) \mid \xi_i \in \{0,1,\cdots,N-1\}, \mid \{i|\xi_i \neq 0\} \mid < \infty\} \\ \xi + \xi' &= (\xi_1 + \xi_1' (\text{mod}\,(N)), \cdots) \qquad \xi, \xi' \in \Omega_N \\ d(\xi,\xi') &= \min(K-1 \mid \xi_j = \xi_j' \quad \forall j \geq K). \end{aligned}$$

The group Ω_N is called the *hierarchical group* of order N. It is natural to use this group to describe the spatial dependence in the context of genetic models where degrees of relationship between different colonies $\xi^{(j)}$ are important rather than some euclidian notion of distance. However, interacting systems indexed by $Z\!\!\!Z^2$ rather than Ω_N can in their longterm behaviour be well approximated by Ω_N models with large N. (See DG 1993 b) for a discussion of the hierarchical group as index set for interacting systems in population genetics.)

The Markov process $X^N(t) = (x_\xi^N(t))_{\xi\in\Omega_N} \in [0,\infty)^{\Omega_N}$ will be defined by the following countable system of stochastic differential equations (we supress N in the components!):

$$(0.2)\qquad \begin{aligned} dx_\xi(t) &= \sum_{k=1}^{\infty} c_{k-1}/N^{k-1}(x_{\xi,k}(t) - x_\xi(t))dt + \sqrt{2g(x_\xi(t))}dw_\xi(t) \\ x_{\xi,k} &= \frac{1}{N^k} \sum_{\xi' : d(\xi',\xi)\leq k} x_\xi \\ X(0) &= X_0 \in E. \end{aligned}$$

The ingredients $(c_k)_{k\in I\!N}, g, (w_\xi(t))_{t\geq 0}$ and E in (0.2) are as follows:

$$(0.3)\qquad (c_k)_{k\in I\!N} \quad \text{satisfies: } c_k > 0 \quad \forall k \in I\!N, \quad \sum_k c_k N^{-k} < \infty \quad \forall N \geq 2$$

$$(0.4)\qquad g : [0,\infty) \to I\!R^+ \ \text{ satisfies: } \quad \begin{aligned} &g(0) = 0, g(x) > 0 \quad \forall x \in (0,\infty) \\ &g \ \text{ is locally Lipschitz-continuous} \\ &g(x) \leq Cx + Dx^2 \quad C \in I\!R^+, D \in (0,1). \end{aligned}$$

The collection of all functions g with the properties required in (0.4) will be denoted by $\mathcal{G}$.

$$(0.5)\qquad \{(w_\xi(t))_{t\geq 0}\}_{\xi\in\Omega_N}$$

is an independent collection of standard Brownian motions.

$$(0.6)\qquad E = \{(x_\xi)_{\xi\in\Omega_N} \mid \|(x_\xi)_{\xi\in\Omega_N}\| < \infty\},$$

where the norm $\|\cdot\|$ is defined by $\|(x_\xi)_{\xi\in\Omega_N}\| = \sum_{\xi\in\Omega_N} \alpha(\xi) x_\xi$, and where $\alpha(\cdot)$ is a fixed function on Ω_N with values in $(0,\infty)$ satisfying the following relations

$$(0.7)\qquad \sum_\xi \alpha(x_\xi) a(\xi,\xi') \leq M\alpha(x_{\xi'}), \qquad \sum_\xi \alpha(x_\xi) < \infty,$$

and $a(\cdot,\cdot)$ is given by:

$$(0.8)\qquad a(\xi,\xi') = a(0,\xi'-\xi), \qquad a(0,\xi) = \sum_{k=d(0,\xi)}^{\infty} c_{k-1}/N^{2k-1}.$$

The function α can be constructed from a strictly positive summable function β on Ω_N by putting $\alpha(\xi) = \sum_{n=0}^{\infty} \sum_{\eta \in \Omega_N} M^{-n} a^n(\xi, \eta)\beta(\eta)$, with $M > 1$.

Remark The system (0.2) has a *unique strong solution in* E. The process $X(t)$ has continuous paths and the strong Markov-property. See Shiga and Shimizu [SS]. (For the existence of a strong solution in $L^2(\alpha)$ see theorem 2.1. Extending this statement to E is done by standard techniques. For the uniqueness one has to remove the restriction to bounded intervals using approximation by finite systems together with coupling techniques.)

In the sequel we shall choose as *initial state* a product measure on $[0, \infty)^{\Omega_N}$, which is homogeneous and concentrated on E. (This restriction to independent components allows us to simplify notation, however, homogeneous ergodic laws could be used as well). We define for such a product measure μ a parameter Θ and impose a moment condition:

$$(0.9) \qquad \theta = E^\mu(x_\xi(0)), \qquad E^\mu(x_\xi(0))^2 < \infty.$$

(Note that we can this way define an initial state on $[0, \infty)^{\Omega_\infty}$ such that the initial states on $[0, \infty)^{\Omega_N}$ are simply the restrictions.

Example An example of a system as defined in (0.2) is a system of interacting Feller's branching diffusions ("super random walk"), where

$$g(x) = \text{const. } x.$$

This model arises as the *diffusion limit* of a continuous time particle branching model. Namely, every particle can migrate on Ω_N according to the transition rates $a(\cdot, \cdot)$ on $\Omega_N \times \Omega_N$ and split at rate 1 into M particles with probability q_M such that $\sum_M q_M M = 1$ and $\sum_M q_M M^2 < \infty$. Here the migration kernel is interpreted as follows: pick a block $\{\xi | d(0, \xi \leq k\}$ at rate (c_{k-1}/N^{k-1}) and a position therein according to the uniform distribution. Writing c_k/N^k as a parameter turns out to be suitable later on. Recall in that context that N^k is the volume of the k-ball $\{\xi | d(0, \xi) \leq k\}$. In order to pass to the diffusion limit give each particle mass ε, increase the number of initial particles like ε^{-1} and the branching rate by ε^{-1}. Let $\varepsilon \to 0$ to obtain a solution of (0.2) satisfying $g(x) = \text{const } x$.

The process with $g(x) = dx^2, d \in (0, \infty)$ plays a special role and exhibits some new phenomena which we cannot treat at this point. Aspects of that model are studied in Gauthier (1994). We shall therefore assume throughout the rest of this paper the following hypotheses:

$$(H) \qquad \overline{\lim}_{x \to \infty} g(x)/x^2 = 0.$$

c) The multiple space-time scale analysis

Our aim in this section is to analyse the *infinite* system given in (0.2) in various *renormalised forms* for large times and for large N. We first rescale *space* and pass from $(x_\xi(t))_{\xi \in \Omega_N}$ to $(x_{\xi,k}(t))_{\xi \in \Omega_N}$, i.e. we consider the field of blockaverages over blocks of size k (volume N^k). Second we rescale *time* as well and pass from $x_{\xi,k}(t)$ to $x_{\xi,k}(t\beta_j(N))$ with $\beta_j(N) = N^j$. Hence we obtain for each pair $(k, j) \in \mathbb{N}^2$ a renormalised system

$$(0.10a) \qquad (x_{\xi,k}(t\beta_j(N)))_{\xi \in \Omega_N}.$$

The goal is to determine the limiting dynamics for this system $(x_{\xi,k}(t\beta_j(N)))_{\xi \in \Omega_N}$ as $N \to \infty$ (limit of mean field interaction). The idea behind this limit is that it is expected that letting first $t \to \infty$, then $N \to \infty$ will result in the same picture and second that the approximation in N is extremely rapid. This idea has been rigorously verified in some cases (See Theorems 4, 11 of this paper, Klenke (1995) and Fleischmann, Greven (1994)).

In order to also get some insight in the behaviour during entire large time intervals, this picture is refined by looking at a renormalisation involving two time scales simultaneously, a *fast and a slow time scale*. For each $(k,j) \in \mathbb{N}^2$ with $k < j$ consider

$$(0.10b) \qquad (x_{\xi,k}(s\beta_j(N) + t\beta_k(N)))_{t\geq 0} \qquad k < j$$

in the limit $N \to \infty$ or more convenient $s, N \to \infty$, in such a way that we set $s = s(N)$ with $s(N) \uparrow \infty$ but $s(N)/N \to 0$ as $N \to \infty$.

In order to describe the limiting dynamics as $N \to \infty$ of (0.10) we need:

Definition 1 (Ingredients of multiple space-time scale analysis)

(i) *time scales* $\quad \beta_j(N) = N^j$

(ii) *block averages* $\quad x_{\xi,k} = N^{-k} \sum\limits_{\xi:d(\xi,\xi')\leq k} x_{\xi'} \quad \forall \xi \in \Omega_N.$

(iii) the *quasi-equilibrium* $\Gamma_\theta^k(\cdot)$ on the k-th level, the k-th level *associated diffusion* $Z^{\theta,k}(t)$ on $[0,\infty)$ and k-th level *diffusion function* F_k defined on $[0,\infty)$:
- – $\Gamma_\theta^k(\cdot)$ is the unique equilibrium of the diffusion $Y(t)$ defined via the SDE

$$dY(t) = c_k(\theta - Y(t))dt + \sqrt{2F_k(Y(t))}dw(t), \text{ with } w(t) \text{ brownian motion}$$

- – $Z^{\theta,k}(t)$ is the stationary solution of above SDE,
- – $\tilde{Z}^{\theta,k}(t)$ is the solution of above SDE with $\tilde{Z}^{\theta,k}(0) = \theta$.
- – $F_k(\theta) := \int\limits_0^\infty F_{k-1}(x)\Gamma_\theta^{k-1}(dx)$ and $F_0(\theta) = g(\theta) \quad \forall \theta \in [0,\infty)$.

(iv) The *level k-marginals* of the *interaction chain* of level j

$$\mu_\theta^{j,k}(\cdot) = \int\limits_{[0,\infty)} \cdots \int\limits_{[0,\infty)} \Gamma_\theta^j(d\theta_1)\Gamma_{\theta_1}^{j-1}(d\theta_2)\cdots\Gamma_{\theta_{j-k}}^k(\cdot).$$

Remark The fact that the SDE for $Y(t)$ in (iii) above has a unique weak solution, will follow from Lemma (2.2 a)) in Section 2.

We are now ready to formulate our result on the multiple space-time scale behaviour of $X^N(t)$. For that purpose we have to distinguish in (0.10 a) the three cases $k > j$, $k = j$ and $k < j$. Note that $\Omega_N \subseteq \Omega_M \subseteq \{\mathbb{N} \cup \{0\}\}^{\mathbb{N}}$ as sets if $N \leq M$.

We shall focus on the case where we consider $x_{\xi,k}(s\beta_j(N))$ for a fixed value of $\xi \in \Omega_N \subseteq \{\mathbb{N} \cup \{0\}\}^{\mathbb{N}}$. (It is possible to prove results about the whole field $\{x_{\xi,k}(s\beta_j(N))\}_{\xi\in\Omega_N}$ as was done in [DGV] for the case of interacting systems of Fleming-Viot processes and we refer the reader to the latter paper if he is interested in such an extension.) With $\mathcal{L}((Y_s^N)_{s\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Y_s)_{s\geq 0})$ we denote weak convergence on the path space $C([0,\infty), \mathbb{R})$.

Theorem 1 (Multiple space-time scale behaviour)
Consider the process $X^N(t)$ started in a homogeneous product measure satisfying (0.9) .
(a) $\quad k > j$

$$(0.11) \qquad \mathcal{L}((x_{\xi,k}(s\beta_j(N)))_{s\geq 0}) \underset{N\to\infty}{\Longrightarrow} \delta_{\{Y(s)\equiv\theta\}}.$$

(b) $\quad k = j$

$$(0.12) \qquad \mathcal{L}((x_{\xi,j}(s\beta_j(N)))_{s\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((\tilde{Z}^{\theta,j}(s))_{s\geq 0}).$$

(c) $k < j$,
for $s = s(N)$ *with* $s(N) \uparrow \infty$ *and* $s(N)/N \to 0$ *as* $N \to \infty$:

$$\mathcal{L}((x_{\xi,k}(s\beta_j(N) + t\beta_k(N)))_{t\geq 0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^{\theta_k^*,k}(t))_{t\geq 0}), \tag{0.13}$$

where θ_k^* *is independent of the evolution and satisfies*

$$\mathcal{L}(\theta_k^*) = \mu_\theta^{j,k}.$$

Furthermore the spatial correlation length built up in time $\beta_k(N)$ *is* k*, that is*

$$\mathcal{L}\Big((x_{\xi,k}(s\beta_j(N)+t\beta_k(N)))_{t\geq 0}|x_{\xi,k}(s\beta_j(N)) = \theta', x_{\xi'}(s\beta_j(N)) \text{ for } d(\xi,\xi') > k)\Big) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^{\theta',k}(t))_{t\geq 0}). \tag{0.14}$$

□

Remark Theorem 1 shows that blockaverages over blocks of size j need at least time $\beta_j(N)$ to start fluctuating. In larger time scales the evolution of such a block average is like a diffusion with a driftfield. The driftfield is a random variable depending on the averages over large blocks. The convenience in letting $s \to \infty$ is that all diffusions are then in equilibrium. Otherwise we would get $\mathcal{L}(\theta_k^*)$ is s dependent and equal to

$$\int \mu_\rho^{j-1,k} P_s(d\rho)$$

where $P_s = \mathcal{L}(\tilde{Z}^{\theta,j}(s))$. Similarly if we let $s \to \infty$ in (0.12) or replace $s\beta_j(N)$ by $s(N)\beta_j(N) + t\beta_j(N)$ with $s(N)$ satisfying the conditions in c) then we obtain $(Z_t^{\theta,j})$ on the r.h.s. of (0.12).

Remark The crucial object to describe the behaviour of the renormalised system are the distributions $\mu_\theta^{j,k}$, which are built using the diffusion coefficient functions on the various levels. These diffusion functions can be found by averaging successively the diffusion function on the previous level with respect to the equilibrium state on that level for frozen higher level averages. The basic mechanism here is the "coexistence" of two time scales: In short time the components relax into a quasiequilibrium, which is dictated by the higher level averages which themselves remain constant for times of that order and start fluctuating only after longer times. The limit $N \to \infty$ then actually *separates* slow and fast time scales. Taking $N \to \infty$ corresponds to taking the rapid stirring limit $\varepsilon \to \infty$ in Durrett (1993).

Remark Observe that in the case of interacting Feller's branching diffusions we have

$$F_k(\theta) = F_0(\theta) = \text{const } \theta \quad \forall\ k \in \mathbb{N}. \tag{0.15}$$

We shall see later on the implications of this *fixed point property*, which allows for explicit calculations.

d) A nonlinear integral operator and its orbit

In the last subsection we saw that the diffusion terms on the various levels in the limit $N \to \infty$, can be calculated recursively. We formalize and extend this point of view now a bit, in order to be able to later on discuss the question of the *universality properties*, of the behaviour of the interacting system as time tends to infinity.

Define for g satisfying (0.4) the *nonlinear* map:

$$\mathcal{F}_c(g)_{(\theta)} = E^{\Gamma_\theta^{c,g}}(g(X)) = \int_0^\infty g(x)\Gamma_\theta^{c,g}(dx) \tag{0.16}$$

where $\Gamma^{c,g}_\theta$ is the unique equilibrium of the diffusion $Y(t)$ given by

$$dY(t) = c(\theta - Y(t))dt + \sqrt{2g(Y(t))}dw(t). \tag{0.17}$$

The distribution $\Gamma^{c,g}_\theta$ can be calculated explicitly for $\theta \neq 0$ as follows:

$$\Gamma^{c,g}_\theta(A) = \frac{1}{Z(c,g,\theta)} \int_A \frac{1}{g(x)} \exp\left(c\int_\theta^x \frac{\theta-y}{g(y)}dy\right)dx, \quad \theta \in (0,\infty) \tag{0.18}$$

$$\text{with } Z(c,g,\theta) = \int_0^\infty \frac{1}{g(x)} \exp\left(c\int_\theta^x \frac{\theta-y}{g(y)}dy\right)dx.$$

For $\theta = 0$ one has $\Gamma^{c,g}_0 = \delta_0$.
Consequently $\mathcal{F}_c$ is given by an explicit formula as well:

$$\mathcal{F}_c(g)_{(\theta)} = \frac{1}{Z(c,g,\theta)} \int_0^\infty \exp(c\int_\theta^x \frac{\theta-y}{g(y)}dy)dx, \qquad \theta \in (0,\infty). \tag{0.19}$$

The case $\theta = 0$ is trivial, here $\Gamma^{c,g}_0 = \delta_0$ and hence $\mathcal{F}_c(g)_{(0)} = 0$.

For a given sequence $(c_k)_{k\in I\!N}$ we define:

$$\mathcal{F}^{(n)} = \mathcal{F}_{c_n} \circ \mathcal{F}_{c_{n-1}} \circ \cdots \circ \mathcal{F}_{c_0} \tag{0.20}$$

and obtain this way the *orbit of* g denoted

$$\{\mathcal{F}^{(n)}(g)\}_{n\in I\!N}. \tag{0.21}$$

The study of this orbit is now in view of (0.19) an analytical problem (see BCGH 1,2). It will turn out later on that for our purposes it is mainly the "endpieces" of the above sequence which are relevant. Furthermore much information about the orbit can be obtained from the special case $c_k \equiv c$, i.e. $\mathcal{F}^{(n)} = (\mathcal{F}_c)^n$, on which we focus now.

We shall define subclasses of $\mathcal{G} = \{g : [0,\infty) \to I\!R^+ | g \text{ satisfies (0.4) and (H) }\}$ of which we shall see later that they correspond to *universality classes* for the structure of the orbit (0.21) and at the same time for the longterm behaviour of the interacting systems.

Consider for the moment the sequence $c_k \equiv c$. We have to distinguish the three cases (here $\theta \in (0,\infty)$).

(A) $\mathcal{F}^{(n)}(g)_{(\theta)}$ diverges as $n \to \infty$,

(B) $\mathcal{F}^{(n)}(g)_{(\theta)}$ is bounded away from 0 respectively ∞ uniformly in n

(C) $\mathcal{F}^{(n)}(g)_{(\theta)}$ converges to 0 as $n \to \infty$.

It will be proved later on that these cases correspond to $g(x)/x \to \infty$, as $x \to \infty$, $g(x)/x$ remains bounded away from 0 and ∞, respectively $g(x)/x$ goes to 0 as $x \to \infty$. It turns out that for large $n, \mathcal{F}^{(n)}(g)_{(\theta)}$ is of the form $d_n\theta$. Therefore fix a sequence $(c_k)_{k\in I\!N}$ satisfying $\sum c_k^{-1} = \infty$ and let $(d_n)_{n\in I\!N}$ be the sequence of numbers with $d_n > 0$ for all $n \in I\!N$ which describe $\mathcal{F}^{(n)}(g)_{(\theta)}\theta^{-1}$ for $n \to \infty$. We view (d_n) as a representative of the equivalence class of sequences given by the relation $a_n \sim b_n$ iff $a_n/b_n \to 1$ as $n \to \infty$.

Let us focus on case (B) first. We set

$$\mathcal{G}((d_n)) = \{g \in \mathcal{G} \mid \|([\mathcal{F}^{(n)}(g)_{(\theta)}/\theta d_n] - 1)1((\varepsilon,\infty))\|_\infty \underset{n\to\infty}{\longrightarrow} 0,\ \forall \varepsilon > 0\}. \tag{0.22}$$

$$\hat{\mathcal{G}} = \bigcup_{d>0} \mathcal{G}((d)), \qquad \mathcal{G}^* = \bigcup \mathcal{G}((d_n)). \tag{0.23}$$

Of particular importance is the class $\mathcal{G}((1))$ resp. $\mathcal{G}((d))$ which contain the case of Feller's branching diffusions. Recall (0.15). We shall see in Baillon et al. (1995) part 2 that $\hat{\mathcal{G}}$ *contains all functions in* $\mathcal{G}$ *with* $g(x)/x$ *converging to the constant* $d \in (0,\infty)$ *as* $x \to \infty$.

In the case A and C things are more subtle, but in a first step one looks at

$$\mathcal{G}((d_n)) = \{g \in \mathcal{G} | (\mathcal{F}^{(n)}(g)_{(\theta)}/\theta d_n - 1) \underset{n\to\infty}{\longrightarrow} 0, \ \forall \theta > 0\}. \tag{0.22'}$$

See Baillon et al. (1995) for more details. It is still open to determine the right notion of uniformity in this convergence.

e) The interaction chain, its entrance laws and their qualitative behaviour

(i) Entrance Laws

The theorem of the subsection 0.c) showed that the behaviour of the interacting system (0.2) for large N and large times is regulated by Markov chains on $[0,\infty)$ defined as follows:

For every $j \in \mathbb{N}$ we define a time inhomogeneous Markov chain on $[0,\infty)$ with time index in $\{-j-1, -j, \dots, 0\}$ denoted

$$(Z^j_k)_{k=-j-1,\dots,0} \tag{0.24}$$

with initial state

$$Z^j_{-j-1} = \theta \tag{0.25}$$

and transition kernel $K_k(\cdot,\cdot)$ at time $-k$ $(k \in \mathbb{N})$ given by (recall Definition 1 below (0.10))

$$K_k(\theta, dy) = \Gamma^{k-1}_\theta(dy). \tag{0.26}$$

We call $(Z^j_k)_{k=-j-1,\dots,0}$ the *interaction chain at level* j.

We can now strengthen (0.13) in Theorem 1 as follows.

Corollary 1 (Large scale space-time dependence structure)

$$\mathcal{L}([x_{\xi,k}(s(N)\beta_j(N) + \sum_{i=k}^{j-1} t\beta_i(N))]_{k=j+1,j,\dots,0}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((Z^j_k)_{k=-j-1,\dots,0}). \qquad \square$$

This means we observe the system after a long time $s(N)\beta_j(N)$ for a number of time points spaced with distance $t\beta_i(N)$ with $i = 0, 1, \dots, j-1$ and space scaled accordingly. The dependence structure of this vector, reflecting the correllation in space and time, is described in the limit $N \to \infty$ by the level j *interaction chain.*

In order to control the behaviour of the interaction chain for large levels j, of particular interest are the *entrance laws* of that chain. Recall that an entrance law for a sequence of kernels $(P_k)_{k\in\mathbb{Z}^-}$ on $I \times I$ is a sequence of laws $(\alpha_k)_{k\in\mathbb{Z}^-}$ on I with $\alpha_{k+1} = \alpha_k P_k$ for all k. Often we shall use the word entrance law as well when we refer to the stochastic process starting at $-\infty$ which corresponds to $(\alpha_k)_{k\in\mathbb{Z}^-}$ and the transition kernels $(P_k)_{k\in\mathbb{Z}}$.

Theorem 2 (Entrance laws of interaction chain)

(a) For every $\theta \in [0,\infty)$ and $Z^j_{-j-1} = \theta$:

$$\mathcal{L}((Z^j_k)_{k=-j-1,\dots,0}) \underset{j\to\infty}{\Longrightarrow} \mathcal{L}((Z^\infty_k)_{k\in\mathbb{Z}^-})$$

where $(Z^\infty_k)_{k\in\mathbb{Z}^-}$ is a time inhomogeneous Markov chain with transition kernels K_k, $k \in \mathbb{N}$ at time $-k$, see (0.26), and the property $\lim_{k\to-\infty} Z^\infty_k$ exists, is finite and is denoted $Z^\infty_{-\infty}$.

(b) All extremal entrance laws of the Markov chain defined by the transition kernels (K_k), with $k \in \mathbb{N}$ are given by the processes $\mathcal{L}((Z^\infty_k)_{k\in\mathbb{Z}^-})$ with $Z^\infty_{-\infty}$ being deterministic i.e. a constant. $\square$

Remark Theorem 2(b) raises the question whether or not to each $\rho \in [0,\infty)$ corresponds an *extremal* entrance law with $Z_{-\infty}^{-\infty} = \rho$. The answer will depend on the sequence $(c_k)_{k \in \mathbb{N}}$ and will be discussed in the next paragraph.

(ii) Stability versus clustering

We shall now show that the qualitative behaviour of the interaction chains depends on the particular values of the coefficients $(c_k)_{k\in\mathbb{N}}$ but not on the choice of diffusion function g *(universality)*. There are two regimes

$$\begin{aligned} \text{strong interaction} \quad &: \quad \sum_{k=0}^{\infty} c_k^{-1} < \infty \\ \text{weak interaction} \quad &: \quad \sum_{k=0}^{\infty} c_k^{-1} = +\infty. \end{aligned}$$

In the first case the random walk with transition kernel $a(\cdot,\cdot)$ is transient and in the second recurrent for *every* $N \geq 2$ provided $\overline{\lim_{k\to\infty}} \sqrt[k]{c_k} < N$, see [DG, 1993b]. Indeed the system exhibits the following dichotomy. (Recall (0.9) for the definition of θ and (0.4) for $\mathcal{G}$):

Theorem 3 (Stability versus clustering)

For every $g \in \mathcal{G}$ the following property holds for the entrance laws defined in Theorem 2a):

$$\begin{aligned} (a) \quad & \sum c_k^{-1} < \infty \quad \textit{implies that} \quad : \begin{cases} Z_k^{\infty} \in (0,\infty)\ a.s. \quad \forall k \in \mathbb{Z}^-,\ if\ \theta > 0 \\ Z_{-\infty}^{\infty} = \theta \end{cases} \\ (b) \quad & \sum c_k^{-1} = +\infty \quad \textit{implies that} \quad : Z_k^{\infty} \equiv 0 \end{aligned}$$

□

Remark In the case of strong interaction to every $\theta \in [0,\infty)$ corresponds an *extremal entrance law*, ν_θ^∞, which is for $\theta \neq 0$ nondegenerate, i.e. it is not concentrated on constant path and the path stays with probability 1 away from the trap 0. In the case of weak interaction there is only the trivial entrance law concentrated on the path which is constant and equal to 0.

What does this dichotomy of Theorem 3 mean for the original system? In case (a) the density θ of the initial state is preserved by the system in the limit of large times and N, while in the second case (b) the original density is distributed more and more uneven as time goes on and large regions of values close to 0 respectively small regions with enormous values develop (recall the preservation of the mean during the evolution). Case (a) is called the *stable* case, case (b) the *clustering* case. We shall discuss in the subsection below the relevance of Theorem 3 for the original system in more detail.

Remark The condition for having stability or clustering in terms of the $(c_k)_{k\in\mathbb{N}}$ is the *same* as for the earlier studied case where the components take values in $[0,1]$ rather than $[0,\infty)$. See [DG,1993b]. Even though the finer analysis will reveal *different types of behaviour* in the two systems, in particular in the clustering case, the same conditions on the c_k play a role. The reason for this is the similar structure of $\mathcal{F}^{(n)}((0.20))$.

Example In the case where $g(x) = d \cdot x$, i.e. Feller's branching diffusions, we can calculate the transition kernels and the Laplace transforms of the marginal distributions for the entrance law explicitly. That is we can identify $\mathcal{L}(Z_{-k}^{\infty})$ for every $k \in \mathbb{Z}^+$ and we can give an explicit expression for every m-step transition probability of the entrance law that is $\mathcal{L}(Z_{-k+m}^{\infty} | Z_{-k}^{\infty} = \theta)$. Recall first the transition kernel of $(Z_k^\infty)_{k\in\mathbb{Z}^-}$, $K_{-k}(\theta,\cdot)$ is given by $\Gamma_\theta^{c_k,dx}(\cdot)$ and $\Gamma_\Theta^{c,dx}$ is a *Gamma distribution* with parameters $(\Theta c/d, d/c)$ (that is, with density $const \cdot x^{\frac{c\theta}{d}-1} e^{-\frac{cx}{d}}$). Define $\psi_k(\lambda) = (c_k/d)\log(1+(d/c_k)\lambda)$ and $\psi_{j,k} = \psi_j(\psi_{j-1}(\cdots(\psi_k(\cdot))\cdots)$. Then:

$$\mathcal{L}(Z_{-k}^{\infty}) \text{ has Laplace Transform } \exp(-\theta\psi_{\infty,k}(\lambda))$$
$$\mathcal{L}(Z_{-k+m}^{\infty} | Z_{-k}^{\infty} = \rho) \text{ has Laplace transform } \exp(-\rho\psi_{k,k-m}).\square$$

(iii) Comparison of Hierarchical and Hierarchical Mean Field Equilibrium Behaviours

Finally we consider the question as to what the results obtained in the $N \to \infty$ limit imply for the Ω_N-systems for fixed N but large times. Eventually we hope to carry out a version of the multiple space-time scale analysis for fixed N. For the moment let us note that the prediction of Theorem 3 gives the correct answer for fixed N systems, if the c_k do not fluctuate too much and furthermore provides good approximations of the equilibra and the spatial dependence structure of the equilibria.

Theorem 4 (Mean field approximation)
(a) *Assume that* $\overline{\lim} \sqrt[k]{c_k} < N$. *Then for every homogeneous ergodic initial measure* $\mu \in \mathcal{P}([0,\infty)^{\Omega_N})$ *with*

$$E^{\mu} x_{\xi} = \theta < \infty \tag{0.27}$$

the following holds ($\underline{0}$ *means the state* $x_\xi \equiv 0$)

(i) *if* $\sum_{k=0}^{\infty} c_k^{-1} = \infty$, *then*

$$\mathcal{L}(X^N(t)) \underset{t\to\infty}{\Longrightarrow} \delta_{\underline{0}}. \tag{0.28}$$

(ii) *if* $\sum_{k=0}^{\infty} c_k^{-1} < \infty$, *then*

$$\mathcal{L}(X^N(t)) \underset{t\to\infty}{\Longrightarrow} \nu_\theta^N, \tag{0.29}$$

where ν_θ^N *is a homogeneous ergodic measure in* $\mathcal{P}([0,\infty)^{\Omega_N})$ *with* $E^{\nu_\theta^{(N)}} x_\xi^N = \theta$.

(b) *Let* $\tilde{\nu}_\theta^N$ *denote the law of* $\{x_{\xi,-k}^N : k \in \mathbb{Z}^-\}$ *induced by* ν_θ^N. *Then under these laws*

$$0.76)\, Var(x_{\xi,-k}^N) \underset{k\to\infty}{\longrightarrow} 0 \tag{0.30}$$

uniformly in N, *and*

$$\tilde{\nu}_\theta^N \underset{N\to\infty}{\Longrightarrow} \nu_\theta^\infty,$$

where ν_θ^∞ *equals* $\mathcal{L}((Z_k^\infty)_{k\in\mathbb{Z}^-})$.

For an extensive discussion of the approximation properties as $N \to \infty$ in particular for the approximation properties of the equilibrium processes (with marginals $\tilde{\nu}_\theta^N$) by ν_θ^∞ we refer the reader to Dawson, Greven and Vaillancourt (1995), where these questions are treated for the compact case (specialize to the two type case there) i.e. the state space $[0,\infty)$ is replaced by $[0,1]$. Analogous results for these statements can be easily derived.

1 Finer analysis of the large scale behaviour

a) Motivation and description of the problem

In this chapter 1 we discuss the results which are of a different flavour compared to the analysis of systems with compact components in [DG, 93a)b)c)] and [DGV, 95]. The analysis is organized in two main parts, in 1b) we discuss the case of weak interaction and the cluster formation, while in 1c) we come to the case of strong interaction and the historical process.

In this section 1a) we start by giving the intuitive background for both situations since some concepts will involve considerable technicalities.

(i) The case of weak interaction

In the case of weak interaction the processes $(Z^j_k)_{k\in\mathbb{Z}^-}$ cluster in the following sense: Z^j_k has the property that for large j and as k increases the chain is with high probability close to 0 and with small probability at very large values. In the limit of $j \longrightarrow \infty$ one then obtains extinction, that is $Z^\infty_k \equiv 0$. In this case we would like to understand the process of cluster formation better by observing $Z^j_\bullet$ for every j on the event of nonextinction at time $k = 0$ normalizing it suitably and then sending $j \longrightarrow \infty$.

Our main aim is to exhibit the dependence of the above large scale behaviour on the parameter g on the one hand and $(c_k)_{k\in\mathbb{N}}$ on the other. We proceed for that purpose in two steps. First we do the rescaling reflecting the interaction given by the $(c_k)_{k\in\mathbb{N}}$ and then second we study the process conditioned on nonextinction at the "end". In contrast to the case of strong interaction the behaviour of the interaction chain for large j is *universal* if $g \in \hat{\mathcal{G}}$. However the dependence on $(c_k)_{k\in\mathbb{N}}$ is rather rich and we will see depending on these coefficients *different regimes* of clustering.

For the purpose of motivating the notions which we need to describe the interaction chain, we turn again to our original interacting systems parameterized by $(c_k)_{k\in\mathbb{N}}$ and g. There are two natural questions in view of (0.28) and $Ex_\xi(t) = Ex_\xi(0)$:

(i) At what rate do components grow conditioned they are not yet "extinct".

(ii) At what rate do clusters of components which are not extinct but grow, increase and remain correllated even as $t \longrightarrow \infty$. (The correllation can be viewed as resulting from common ancestry, compare the $g(x) = x$ case)

In systems with weak interaction the clusters (i.e. regions of correllated growth) differ first in their *spatial shape* and second in the *rate of growth* in the growing components. This will lead to a classification of the longterm behaviour in two main regimes for the case of weak interaction. The way to describe such phenomena is the following. Consider an observer sitting at the site $\tilde{\xi} = (0, 0, \dots)$ (for example) and who sees nontrivial mass at time t, that is, $x_{\tilde{\xi}}(t) \geq \varepsilon$. Then we ask first at what scale does $x_{\tilde{\xi}}(t)$ grow in t on the event $\{x_{\tilde{\xi}}(t) \geq \varepsilon\}$, which gives us the rate of growth. Second in order to capture the spatial extension of the cluster we consider balls $\{\xi \in \Omega_N | d(\xi, \tilde{\xi}) \leq k\}$ and the average density in this ball, i.e. $x_{\tilde{\xi},k}(t)$. In fact if we want to know at what rate the cluster expands, we should look at expanding balls $\{\xi \in \Omega_N | d(\xi, \tilde{\xi}) \leq f_\alpha(t)\}$, where $\alpha \longrightarrow f_\alpha(t)$ is nondecreasing and $f_\alpha(t) \uparrow \infty$ as $t \longrightarrow \infty$ and observe whether this density in these growing balls is still of the order of the value $x_{\tilde{\xi}}(t)$ at the center of the ball on the event $\{x_{\tilde{\xi}}(t) \geq \varepsilon\}$ and such that the correllation on that event between $x_\xi(t)$ and $x_{\tilde{\xi}}(t)$ is a nontrivial function of α if $d(\xi, \tilde{\xi}) = f_\alpha(t)$.

We can summarize this procedure as follows: Find $h(t) \uparrow \infty$ as $t \longrightarrow \infty$ and $f_\alpha(t)$ and determine the limit of

$$\mathcal{L}(x_{\tilde{\xi}, f_\alpha(t)}(t)/h(t) | x_{\tilde{\xi},0}(t) \geq \varepsilon)$$

as $t \longrightarrow \infty$. If the limit depends in a nontrivial way on α we will see clusters of random order of magnitude. Namely, consider all components where the value of the limiting field exceeds a certain fixed value. This will then define the spatial extension of the cluster. Note that this might be a geometrically quite irregular object.

The case of fixed N is more difficult to analyse and therefore here we pass to the limit $N \longrightarrow \infty$ in suitable time scales. The problem then translates into finding a function $j \longrightarrow h(j)$ and $f_\alpha(j)$ such that

$$\mathcal{L}(Z^j_{f_\alpha(j)}/h(j) | Z^j_0 \geq \varepsilon)$$

converges as $j \to \infty$ to a nontrivial limit independent of ε which has nontrivial fluctuations in α.

Both the spatial shape and the rate of growth depend on the strength of the recurrence of the underlying migration. The rate of growth depends also on the behaviour of $g(x)$ as $x \to \infty$. We first consider a function $g \in \hat{\mathcal{G}}$ defined in (0.23) and in this class the cluster formation is *universal.*

We now discuss the influence of the $(c_k)_{k \in \mathbb{N}}$ on the behaviour. Return to the interacting system again, i.e. N finite. There are *two basic regimes of clustering* with fundamentally different qualitative properties possible. They are characterized by both the expansion in space and the speed of growth in a component. We shall see later on that the distinction between the two regimes depends on whether c_k decays slower than exponential or decays exponentially fast. We label these regimes I and II, respectively. The dichotomy corresponds to the $d = 2$ versus $d = 1$ cases for systems indexed with $\mathbb{Z}^d$ instead of the hierarchical group and in fact the dichotomy can in general be described in terms of the random walk with transition probabilities $a_t(.,.)$ generated by $a(\cdot,\cdot)$. To get a rough idea fix $\delta \in (0,\infty)$ and look at the set of points $L_t = \{\xi | a_t(\tilde{\xi},\xi)/a_t(\tilde{\xi},\tilde{\xi}) \in (\delta,\delta^{-1})\}$. Then the question is whether $|L_t| \sim t$ or $|L_t| = o(t)$, the background for this distinction being that a branching system needs initial mass of order t in order to survive with positive probability until time t . Both on the hierarchical group and on the lattice the set L_t can be described in terms of balls around the point $\tilde{\xi}$.

In the first regime the heights within clusters grow only slowly and the clusters expand in space at *different* (random) *orders of magnitude* (diffusive clustering). This first regime displays a well defined growth rate for *all* components which have a value exceeding some fixed number $\varepsilon > 0$. The time scale is such that for fixed N the growth rate in j is a slowly varying function of the time.

The second regime displays a more irregular clustering behaviour in the sense that we will observe more spatial variability in the order of magnitude at which the growing components diverge as $t \to \infty$, there will be various growth rates possible and furthermore colonies with rapidly increasing mass occur. Rapidly growing means here that the growth rate for fixed N is not slowly varying as a function of time. In fact in the case $c_k = c^k,\ c < 1$ it is regularly varying. However these clusters of sites with fast growing components fluctuate hardly at all in space (concentrated clustering), in the sense that the order of magnitude of such clusters is deterministic and only a multiplicative factor is random. Furthermore almost all the mass of one cluster sits at sites, where growth at the "maximal" rate takes place, even though almost all components of the cluster show a growth at a slower than the maximal order.

In order to understand this phenomenon of two different regimes on a heuristic level, assume $g(x) = x$ (and hence obtain a branching system) and note that a unit mass starting at 0 and evolving according to (0.2) produces a system whose total mass is at time $s(N)N^j$ either 0 or of size $Zs(N)N^j$ for N large with $\mathcal{L}(Z) = \exp(1)$. The surviving mass has a possible range of migration at most up to distance j in the time considered. The density profile of the mass given nonextinction will then depend on $c_0, c_1, \dots, c_j$. Note furthermore that for a system started in a homogeneous distribution there will exist in the ball $\{\xi | d(0,\xi) \le j\}$ about $N/s(N)$ colonies such that a unit mass started there is not yet extinct at time $s(N)N^j$. Combining the facts given so far we expect that if the c_k do not decay too fast different surviving families will charge a given ball and hence the *localization of the ancestor* of the mass surviving until the observation time will not be relevant for the density profile as a function from the distance of the ancestor and measured by the mass with which a ball is charged. Otherwise since away from the ancestor we see a dramatic drop in population density, separated families produce a second regime of cluster formation.

We call these latter clusters *concentrated clusters*, since most of their mass is going to be located in a small part of the cluster; the growth of different parts of the cluster will be of different order of magnitude. The first regime is called *diffusive clustering* since the clusters even though they will not cover the whole box will charge balls at every distance from a reference point with a spatial extension of random order of magnitude and with components of a particular deterministic order of magnitude everywhere in space.

These distinctions discussed in the paragraphs above can be translated into a statement of scaling properties of the interaction chain, which reflects the two regimes. This will be described in section 1b) in Theorem 5 and Theorem 6 respectively.

(ii) The case of strong interaction
Since in the case $\sum c_k^{-1} < \infty$ we shall see later that the orbit $\{\mathcal{F}^{(n)}(g)\}_{n \in \mathbb{N}}$, recall (0.21), looks eventually like the constant sequence $\{g_\infty\}$ and this g_∞ does depend on the original g, we see that the entrance laws themselves are not universal objects. It is only the dichotomy between a one parameter set of extremal entrance laws versus a single one which is universal. This means in particular that the equilibrium process of the original system will locally reflect properties of g strongly. For example the local dependence structure of the equilibrium state will be different for different g as well as properties of the marginal distribution. This is different in the clustering case. In the sequel in point 1c) we will focus for a finer analysis of the stable case at least for $g(x) = \text{const. } x$ which has a special structure and is particularly important for applications since it is a branching system. In order to extract the universal properties of the equilibrium states on a large spatial scale (as opposed to the local dependence structure) some additional tools have to be developed which we defer to a future paper, but we construct here some fundamentals for this enterprise.

The system $X^N(t)$ for fixed N and $g(x) = dx$ can be analysed by embedding it in a richer structure, the so-called *historical process*, which serves as a tool to obtain finer information about the original system itself as well. This will be contained in Theorem 7, 8 below in section 1c). Furthermore this analysis reveals a relation between the branching system and the Fleming-Viot process which is the topic of Theorem 9 and allows us to study the spatial distribution of families in Theorem 10. In Theorem 11 we discuss the quality of the mean-field approximation. This is the analogue of Theorem 4 but now on the level of the historical process.

In order to give an intuitive idea of the historical process we first consider a technically simpler object, namely, a branching random walk on Ω_N. This is a system in which particles are located on Ω_N, and hence the state space is $(\mathbb{N} \cup \{0\})^{\Omega_N}$. (A state if the system is often viewed as a counting measure on Ω_N). The particles perform independent random walks with transition kernel $a_N(\cdot,\cdot)$ and they split into two particles or die at exponential rate one. Consider first the system starting with one particle. The resulting counting measure-valued process has the form $\sum_{j=1}^{N(t)} \delta_{\xi_j(t)}$ where $N(t)$ is the number of particles at time t and $\xi_j(t)$ is the location of the jth particle at time t. Due to the branching structure, the system starting in an initial distribution with infinitely many particles is just the superposition of the processes starting in a single particle.

In order to define the associated historical process at time t we enrich the state space to encode information on the trajectory which had been followed by each particle (and its ancestors). This is achieved by considering a counting measure-valued process $(H_t^{N,0})_{t \geq 0}$ given by $\sum_{j=1}^{\infty} \delta_{\xi_j(\cdot \wedge t)}$, where $\xi_j(\cdot \wedge t)$ is a shorthand for the stopped trajectory $\xi_j(s \wedge t)_{s \in \mathbb{R}}$ which had been followed by the jth particle alive at time t. The state of the historical process is a counting measure in $\tilde{E}_0 = \{\mu \in M(D([0,\infty),\Omega_N)) : \{\mu(\{y : y(t) = \xi\})\}_{\xi \in \Omega_N} \in E \ \forall t\}$.

Of special interest in the description of the evolution is often the Palm distribution, which is the measure on path space obtained by picking at time t at site ξ a random member of the population (size-biased sampling). The "family tree" of such randomly chosen particle is the *backward tree.*

If the random walk $a_N(\cdot,\cdot)$ is transient, then the branching random walk has for every $\theta \in \mathbb{R}^+$ an equilibrium state which is spatially homogeneous and ergodic and has the expected number of particles per site equal to θ. In order to describe the analogous equilibrium for the historical process it is useful to view the process with time parameter set $(-\infty, \infty)$ and to describe the equilibrium at time 0. Furthermore we introduce an equivalence relation on $D((-\infty,\infty),\Omega_N)$ by setting $y_1 \equiv y_2$ iff there exists a time s such that $y_1(u) = y_2(u)\ \forall u \leq s$. A measure on $D((-\infty,\infty),\Omega_N)$ concentrated on a single such equivalence class is called a *clan (or family) measure.* It can be shown that with probability one, the equilbrium historical measure can be decomposed into a countable sum of clan measures.

There is a close relationship between the historical process and the genealogical structure of the branching particle system. However since there is a positive probability that two independent random walks can stay at the same location for a finite time interval, the history of a particle does not uniquely determine the genealogy in finite time intervals. However over infinite time intervals two distinct particles cannot have a common trajectory. Therefore the historical decomposition is equivalent to a *genealogical* decomposition that is, decomposition of the population into families of particles in which any two particles have a common ancestor.

Some natural questions about the equilibrium process are:

(i) What is the density of a single family in space?

(ii) What is the number of different families per unit volume?

(iii) How does the frequency of different families in large volumes behave?

(iv) What can we say about the large scale properties of the clan measure of a randomly chosen particle (backward tree)?

It is these questions which we shall address for the system of interacting Feller branching diffusions using multiple space-time scale analysis.

In order to analyse the questions (i)-(iv) for our processes $X^n(t)$, which are diffusion limits of branching random walks, we need to decompose $x_\xi(t)$ and $x_{\xi,k}(t)$ into the contributions coming from different families. (We shall make this precise in Theorem 7 in the sequel). Suppose we have decomposed $x_{\xi,k}(t) = \sum_i M^{*N}_{t,k}(\xi,i)$ in a size ordered way (at $\xi = 0$). If we consider as initial mass an equilibrium state and let $t \to \infty$ we have the decomposition of the equilibrium state into different families. Then we want to know whether we can find functions $k \to h(k), \quad h(k) \uparrow \infty$ such that

$$\mathcal{L}(h(k) \cdot M^{*N}_{\infty,k}(\xi,i))$$

converges as $k \to \infty$. This gives us the asymptotic spatial density profile $h(\cdot)$ of a single family. Furthermore if we define

$$T_k = \inf(j | \sum_{i=0}^{j} M^{*N}_{\infty,k}(\xi,i) / \sum_{i=0}^{\infty} M^{*N}_{\infty,k}(\xi,i) \geq 1-\delta)$$

then T_k describes the number of different families within a ball of size k. The question is whether we can find a scaling function $k \to h'(k)$ with $h'(k) \uparrow +\infty$ such that $\mathcal{L}(T_k/h'(k))$ converges as $k \to \infty$.

The next basic question is to focus on the whole spatial family distribution at a site and in an increasing sequence of blocks. Since the influence travels from higher levels to lower level we run the index from bottom to top. In other words look at the $\mathcal{P}(\mathbb{N})$-valued process $(Q^*_k)_{k\in\mathbb{Z}^-}$ where for $k \in \mathbb{N}$:

$$Q^{*N}_{-k}(\{i\}) = \frac{M^{*N}_{\infty,k}(\xi,i)}{x_{\xi,k}(0)}.$$

Finally we are interested in the decomposition of a family of a randomly chosen individual according to the degree of relationship and the spatial distribution of such subfamilies.

Since these questions are difficult to analyse we pass to the corresponding quantities for the entrance law $(Z^\infty_k)_{k\in\mathbb{Z}^-}$ and identify h and h' in this case as functions of $(c_k)_{k\in\mathbb{N}}$.

Furthermore we identify the law of $(Q_k^*)_{k\in\mathbb{Z}^-}$ in terms of the entrance law of a well known system driven by resampling (Fleming-Viot process). Finally we describe the structure of the limiting object of the backward tree in an explicit way.

b) Results on Cluster formation

In this subsection we study the cluster formation in the case of weak interaction, i.e. $\sum c_k^{-1} = \infty$ and on the level of the interaction chain i.e. in the limit $N = \infty$. First we introduce the regimes of clustering discussed for finite N in subsection (a), now on the level of $N = \infty$ and start by introducing the necessary ingredients in (1.1) - (1.2) below. We shall consider the process $(Z_{-k}^j)_{k=j+1,j,\ldots,0}$ on the event $Z_0^j \geq \varepsilon$. In order to obtain a description of this process via a limit theorem we need to define suitable scales for mass and time.

Consider for a $g \in \mathcal{G}((d))$ the *mass transformation*

$$(1.1) \qquad \tilde{Z}_k^j = Z_k^j / \Big(\sum_{\ell=0}^{j} c_\ell^{-1}\Big) d \qquad , j \in I\!N,\ k \in \mathbb{Z}^-$$

and the *time transformation*

$$(1.2) \qquad \overset{\approx}{Z}{}_\alpha^j := \tilde{Z}_{f_\alpha(j)}^j \quad \alpha \in [0,1].$$

Here the scaling functions $[0,1] \to \mathbb{Z}^-$ are the following:
Let $\{(f_\alpha(j))_{\alpha\in[0,1]}\}_{j\in I\!N}$ be a set of functions $\alpha \longrightarrow f_\alpha(j)$ which satisfies for every j:

$$(1.3) \qquad f_0(j) = -j-1, \quad f_1(j) = 0$$

$$(1.4) \qquad f_\alpha(j) \text{ is non-decreasing in } \alpha,\ f_\alpha(j) - (-j-1) \to \infty \text{ as } j \to \infty \text{ for } \alpha \in (0,1).$$

We say that $(c_k)_{k\in I\!N}$ belongs to regime I, the so-called *diffusive clustering*, if we can find scaling functions as above such that the corresponding process $(\overset{\approx}{Z}{}_\alpha^j)_{\alpha\in[0,1]}$ conditioned on the event $Z_0^j \geq \varepsilon$ converges to a nontrivial limit process which is independent of ε. If this limit is a process with path constant and equal to 0 then $(c_k)_{k\in I\!N}$ belongs to regime II, which we refer to as *concentrated clustering*. The limit process in the first regime will be denoted by $(\hat{Z}_\alpha)_{\alpha\in[0,1]}$ and is called the *cluster process*. We shall now discuss in (i) and (ii) below the two regimes separately.

(i) Regime I (diffusive clustering)

The most important features of the regime 1 for the interacting system is that the spatial size of clusters have a random order of magnitude after passing to the limit $N \to \infty$ which is reflected in the fact that Z_k^j can be scaled and conditioned such that we obtain a limiting diffusion in α. We start with classifying, within regime I, the spatial size of clusters and defining the candidate for the cluster process.

The spatial distribution of the cluster within the j-block of a component with value at least size $\varepsilon(\varepsilon > 0)$ also depends on the particular form of the $(c_k)_{k\in I\!N}$ and three main cases have to be distinguished according to the quantitatively and qualitatively quite different rates of spatial expansion, which is captured in the distinction of $|f_\alpha(j)|/j$ converging to 1, a nontrivial function of α or 0 as $j \to \infty$ giving the distinction between *large, moderate,* respectively *small clusters.* This classification depends on the geometry of the hierarchical group. We collect the classification for further reference:

$$(1.5) \qquad \text{For every } \alpha \in (0,1):\ f_\alpha(j)/j \underset{j\to\infty}{\longrightarrow} 0 \text{ (small clusters)}$$

$$(1.6) \qquad f_\alpha(j)/j \underset{j\to\infty}{\longrightarrow} f_\alpha(\infty) \text{ with } f_\bullet(\infty) \text{ non constant} \quad \text{(moderate clusters)}$$

$$(1.7) \qquad \text{for every } \alpha \in (0,1): f_\alpha(j)/j \underset{j\to\infty}{\longrightarrow} -1 \text{ (large clusters)}$$

The most important case in (1.6) is $f_\alpha(\infty) = -(1-\alpha)$.

To describe $\hat{Z}$ the *limit process* as $j \to \infty$ of $\approx\!\!\!\!\!Z$, we need the following time inhomogeneous diffusion on $[0,\infty]$. Let $(Z(\alpha))_{\alpha\in[0,1]}$ be the diffusion with generator G_α at time α and initial value 0. Here

$$G_\alpha = a_\alpha(x)\frac{\partial}{\partial x} + b_\alpha(x)\Big(\frac{\partial}{\partial x}\Big)^2 \tag{1.8}$$

with

$$\begin{aligned} a_\alpha(x) &= 2x\Big(\frac{1-\alpha}{e^{x/1-\alpha}-1}\Big) \\ b_\alpha(x) &= (2x + a_\alpha(x)). \end{aligned} \tag{1.9}$$

This process has paths which start at 0, never return to 0 and end at time 1 at a random point which has distribution exp (1). The process $Z(\cdot)$ and others obtained by deterministic time transformation from it will occur as limit process of $\overset{\approx}{Z}{}^j_\alpha$ as $j \to \infty$.

Candidates for regime I are $(c_k)_{k\in\mathbb{N}}$ which do not decay exponentially fast. Within the regime I we shall now see that how the cluster process looks like depends on whether c_k diverges, converges to 0 or converges to a constant $\in (0,\infty)$, the latter being the critical case for which we get an explicitly determined clusterprocess. Namely (recall (1.1) – (1.4))

Theorem 5a) (diffusive clusters, moderate cluster size)
For $g \in \mathcal{G}((d)), \varepsilon > 0$ and with $f_\alpha(j) = [\alpha(j+1)] - j - 1$ the following holds:

$$c_k \to c \in (0,\infty) \textit{ as } k \to \infty \textit{ implies diffusive clustering and} \tag{1.10}$$

$$\mathcal{L}((\overset{\approx}{Z}{}^j_\alpha)_{\alpha\in[0,1]} | Z^j_0 \geq \varepsilon) \Longrightarrow \mathcal{L}((Z(\alpha))_{\alpha\in[0,1]}), \textit{ as } j \to \infty$$

$$c_k \sim \text{const.}k^\beta,\ \beta \in (-\infty, 1) \textit{ implies diffusive clustering and} \tag{1.11}$$

$$\mathcal{L}((\overset{\approx}{Z}{}^j_\alpha)_{\alpha\in[0,1]} | Z^j_0 \geq \varepsilon) \Longrightarrow \mathcal{L}((Z(\alpha^{-\beta+1}))_{\alpha\in[0,1]}). \qquad \square \tag{1.12}$$

If we ask for the implications of above theorem for the $N < \infty$ case, we see here that the height of a cluster (ratio of mass and volume) grows slowly in time, namely it is of order $(\sum_1^j c_k^{-1})$ at time N^j with $\sum_1^j c_k^{-1}$ *slowly varying* as function of N^j (= time scale) for fixed N. The profile of the normalized density in space is *random* and the law of the profile is given via an explicitly given diffusion. This means in particular for the original interacting system that the spatial extension of the clusters in space are of different and random order of magnitude. This is analogous to the compact case of (0.2): $g > 0$ on $[0,1]$, where spatial extensions of clusters of 0's and 1's have a random order of magnitude whose law is given via the Fisher-Wright diffusion (See [DG3], [FG], [K]).

Remark Note that the case $c_k \equiv c$ is the analogue of the lattice model with dimension $d = 2$ and symmetric random walk kernel with finite variance. We obtain for $\alpha = 1$ as marginal of the cluster process the exponential distribution that is we have *no atom at* 0. This is an unresolved problem in lattice models, see Fleischman (1978), Durrett (1979) for related results.

In the cases of moderate cluster sizes dealt with in Theorem 5a)

$$\sum_{[\alpha n]}^{n} c_k^{-1} \Big/ \sum_{0}^{n} c_k^{-1} \underset{n\to\infty}{\longrightarrow} 1-\alpha \text{ or } 1-\alpha^{1-\beta},\ \forall \alpha \in (0,1]. \tag{1.13}$$

By contrast, the cases of diffusive clustering with small or large cluster size corresponds to the following two situations:

$$\begin{aligned} &c_k \to \infty \textit{ as } k \to \infty, \quad \sum_0^\infty c_k^{-1} = \infty \\ &\sum_{[\alpha n]}^{n} c_k^{-1} \Big/ \sum_0^n c_k^{-1} \underset{n\to\infty}{\longrightarrow} 0,\ \forall \alpha \in (0,1] \end{aligned} \tag{1.14}$$

$$(1.15)\qquad \begin{array}{l} \sum_{[\alpha n]}^{n} c_k^{-1} / \sum_0^n c_k^{-1} \underset{n\to\infty}{\longrightarrow} 1 \\ \sum_0^{n-m} c_k^{-1} \quad / \sum_0^n c_k^{-1} \underset{n\to\infty}{\longrightarrow} 1 \quad \forall m \in I\!N. \end{array}$$

We analyse first the case (1.14) and then discuss (1.15). Define a sequence h_n of functions,

$$(1.16)\qquad h_n : [0,1] \to \{0,1,\dots,n\}$$

with $h_n(0) = n$, $h_n(\cdot)$ is nonincreasing and satisfies

$$(1.17)\qquad \Big(\sum_{h_n(\alpha)}^{n} c_k^{-1}\Big) / \Big(\sum_{k=0}^{n} c_k^{-1} h_n(0) = n, \Big) \underset{n\to\infty}{\longrightarrow} 1-\alpha.$$

Define in the case where the $(c_k)_{k\in I\!N}$ satisfy (1.14) the following scaling functions:

$$(1.18)\qquad f_\alpha(j) = h_j(1-\alpha) - j - 1.$$

Note that $h_n(\alpha)/n$ converges to 0 as $n \to \infty$, if we are in the regime given in (1.14). In the case (1.15) we then define the scaling function $f_\alpha(j)$ as follows:

$$(1.19)\qquad \sum_{\ell=0}^{|f_\alpha(n)|} c_\ell^{-1} / \sum_{\ell=0}^{n} c_\ell^{-1} \underset{n\to\infty}{\longrightarrow} \alpha.$$

We are now ready to complete the picture from Theorem 5a) as follows:

Theorem 5b) (diffusive clustering, large and small cluster size)
Suppose that $g \in \mathcal{G}((d))$, $\varepsilon > 0$ and that either relation (1.14) or relation (1.15) are satisfied and furthermore use the scaling function $f_\alpha(j)$ from (1.18) respectively (1.19). Then

$$(1.20)\qquad \mathcal{L}((\overset{\approx j}{Z}_\alpha)_{\alpha\in[0,1]} | Z_0^j \geq \varepsilon) \Longrightarrow \mathcal{L}((Z_\alpha)_{\alpha\in[0,1]}).$$

□

Example An example for large cluster size is $c_k = \exp(-bk^\beta)$, $b \in (0,\infty)$, $\beta \in (0,1)$ and for small cluster size $c_k = k$.

Remark Note the different structure of the interaction chain in the following cases:

$$K_j(\theta,\cdot) \underset{j\to\infty}{\longrightarrow} \begin{cases} \delta_\theta \quad \text{(small cluster size, } c_k \to \infty) \\ \text{nontrivial law (moderate cluster size, } c_k \to c) \\ \delta_0 \quad \text{(large size, } c_k \to 0). \end{cases}$$

(ii) Regime II (concentrated clustering)

Consider first the interacting system (i.e. $N < \infty$) again. If the $c_k \to 0$ but do not satisfy (1.14) or (1.15), for example $c_k = e^{-k}$, we are now in a situation where clusters develop with enormous peaks so that the structure of the boundary of the cluster of growing and correlated components becomes very rich. Furthermore different "families" are separated (this is sit in different balls!). Hence conditioning the interaction chain at 0 to be at least ε and rescaling the mass as in (1.1) does not yield the right description because we will get qualitatively different behaviour if we are in a situation where 0 is at the boundary or the center of the cluster.

The Theorem 6 below will formally state the $N = \infty$ analogue of the fact that a randomly chosen component in a cluster lies with overwhelming probability on the boundary of the cluster, that is it shows much slower growth of mass than the peaks, but on the other hand we give a description of the spatial profile of the density occuring in the center of the cluster, which also describes a typical mass in a very large block (Palm-distribution). This reveals in particular that the points where growth occurs at "maximal" rate make up only a assymptotically vanishing fraction of the components of the entire cluster but makes up asymptotically all the mass. (This explains the name).

In order to describe the spatial structure of the "peaks" we need a new object. The structure we want to use here is that for $g(x) = x$ the equilibria $\Gamma_\theta^{c,g}$ are infinitely divisible and can be represented using the *canonical measure.* Define a time inhomogeneous Markov chain $(\tilde{Z}_m)_{m \in I\!N}$. The chain starts at θ and the transition kernel $\tilde{K}_m(\theta, d\rho)$ at time m is given by (recall $c < 1$)

$$\tilde{K}_m(\theta, d\rho) = \Gamma_\theta^{c^{-m},x}(d\rho), \quad m \in I\!N.$$

This Markov chain is a martingale and $\tilde{Z}_m \to \tilde{Z}_\infty$ as $m \to \infty$ with $E\tilde{Z}_\infty = \theta$. The Laplace transforms $\tilde{L}_m(\lambda)$ resp. $\tilde{L}_\infty(\lambda)$ of $\mathcal{L}(\tilde{Z}_m)$ respectively $\mathcal{L}(\tilde{Z}_\infty)$ have the representation

$$\tilde{L}_m(\lambda) = \exp(-\theta \int_0^\infty (1 - e^{-\lambda u}) R_m(du)),$$

defining a sequence of unique σ-finite positive measures (R_m).

Finally let X be a random variable with distribution G, then we denote by

$$d\widehat{G} = (\int yG(dy))^{-1} x dG$$

the *Palm distribution.* (For $X = Z_0^j$ this distribution corresponds in the original system to the following: Take a large block of components at time $\beta_j(N)$ and sample a "typical particle". Then look at the distribution of mass at the in this way sampled component, this is distributed according to the limit $\widehat{G}$ for $N \to \infty$ of the sequence of Palm distributions.) This definition can be extended to define the Palm measure $\widehat{P}$ of the object $(Z^j_{-j-1+m})_{m \in \{0,1,\dots,j+1\}}$ with respect to Z_0^j by simply putting

$$\widehat{P}(Z_1 \in A_1, \dots, Z_m \in A_m) = E(Z_0^j 1(Z^j_{-j-1+1} \in A_1, \dots, Z^j_{-j-1+m} \in A_m))\theta^{-1}.$$

Now we are ready to formulate our result:

Theorem 6 (Clusterformation in regime II, concentrated clusters)
Assume $g \in \mathcal{G}((d))$ and $\varepsilon > 0$ and furthermore $(c_k)_{k \in I\!N}$ satisfies

$$(1.21) \qquad \sum_{k=0}^{n-m} c_k^{-1} / \sum_{k=0}^{n} c_k^{-1} \underset{n \to \infty}{\longrightarrow} \beta(m) \in (0,1) \quad \forall m \in I\!N, \quad \beta(m) \to 0 \text{ as } m \to \infty.$$

(a) Then for every nondecreasing $\alpha \to f_a(j) \in \{-j-1, \dots, 0\}$ with $f_0(j) = -j-1$, $f_1(j) = -1$ and $f_\alpha(j) - (-j-1) \to -\infty$ if $\alpha \in (0,1)$ the following holds:

$$(1.22) \qquad \mathcal{L}((Z^j_{f_\alpha(j)})_{\alpha \in [0,1]} / \sum_0^j c_k^{-1} d) | Z_0^j > \varepsilon) \underset{j \to \infty}{\Longrightarrow} \delta_{\{Z_\alpha \equiv 0\}}.$$

Furthermore

$$(1.23) \qquad E((Z_{-j-1+m})^2 | Z^j_{-j-1} = \theta) \sim (1 - \beta(m)) \sum_{k=0}^{j} c_k^{-1}.$$

(b) Assume that $c_k = c^k$, $c < 1$ and define the rescaled interaction chain at level j by $\tilde{Z}^j_{-k} = c^j Z^j_{-k}$. Then

$$(1.24) \qquad E[1 - \exp(-\lambda \tilde{Z}_0^j)] \sim c^j \theta \int_0^\infty (1 - e^{-\lambda u}) R_\infty(du),$$

$$(1.25) \qquad \int_0^\infty R_\infty(du) = +\infty, \quad \int_0^\infty u R_\infty(du) = 1,$$

which implies that $c^{-j} Prob(Z_0^j \geq \varepsilon) \to \infty$ *as* $j \to \infty$.

Furthermore for $m \in \mathbb{N}$

$$(1.26) \qquad E[1-\exp(-\lambda \tilde{Z}^j_{-j-1+m})] \sim c^j \theta \int_0^\infty (1-e^{-\lambda u}) R_m(du)$$

$$(1.27) \qquad \text{and } R_m \Rightarrow R_\infty \text{ as } m \to \infty.$$

c) Assume $c_k = c^k, \quad c < 1$. *Then there exists a distribution function* $\widehat{F}_\infty$ *such that*

$$(1.28) \qquad \widehat{P}(\tilde{Z}_0^j \geq \delta) \to \widehat{F}_\infty(\delta), \text{ as } j \to \infty \qquad \forall \delta \in (0,\infty),$$

where $\widehat{F}_\infty$ *is the Palm measure of the Levy-measure of* $\tilde{Z}_\infty$. *The corresponding Laplace transform is given by:*

$$\int_0^\infty e^{-\lambda u} \widehat{F}_\infty(du) = \sum_0^\infty (-1)^k \frac{\lambda^k}{k!} D_{k+1}$$

with D_k *given through the recursion formula in (5.116).*

Consider now the Palm distribution of $(\tilde{Z}^j_{-j-1+m})_{m=0,1,\ldots,k}$ *(with respect to* $\tilde{Z}_0^j$*). Then*

$$(1.29) \qquad \widehat{P}((\tilde{Z}^j_{-j-1+m})_{m=0,1,\ldots,k}) \underset{j\to\infty}{\Longrightarrow} P((\tilde{Z}_m)_{m=0,1,\ldots,k}) \quad \forall k \in \mathbb{N}. \qquad \square$$

Remark The parts a) and b) mean for the $N < \infty$ situation, that the clusters at a particular site have again a height, which creates at a small number of points a density described by the quantity $\sum_{k=0}^j c_k^{-1} d$ at time $s(N)N^j$. On the other hand if we study the system conditioned on a particular component being at least ε (see (1.22)) then we see at this site a positive mass density of smaller order of size than $(\sum_1^j c_k^{-1})$ due to the fact that the spatial extension of the region where the maximal growth $\sum_0^j c_k^{-1} d$ is attained, is small compared to the total extension of the cluster.

Remark Part (c) implies that in regime II the conditional laws $\mathcal{L}(\tilde{Z}_0^j | \tilde{Z}_0^j \geq \varepsilon)$ converge as $j \to \infty$ (to an ε-dependent limit). Recall that by (b) we know $\mathcal{L}(\tilde{Z}_0^j | Z_0^j \geq \varepsilon)$ does <u>not</u> converge to a nontrivial limit.

Remark The above result confirms that also in the regime II the growth of order c^{-j} as $j \to \infty$ plays an important role if we view things from a different perspective. Namely suppose we go back to a branching particle system and after a long time we choose from a large block an individual at random. Then look at the component at which it sits and around this site. What growth do we see in the case where the random walk is of the form $c_k = c^k$ with $c < 1$? The particle should sit within a cloud of maximal growing height, since even though the boundaries of the cluster are large, they do not contain much mass. On the mean-field limit level this is (1.28) and (1.29).

(iii) Extension and questions.

¿From Theorems 5 and 6 we see that the spatial structure of clusters is determined by the random walk kernel, i.e. the sequence $(c_k)_{k\in\mathbb{N}}$. It is the absolute speed of growth which reflects the influence of the diffusive term as given by the function g. Hence in order to understand the finer structure of the longterm behaviour of the system we need to determine $\mathcal{G}((d_n))$. This question has been resolved in a forthcoming paper by Baillon, Clément, Greven and den Hollander [BCGH 2].

The considerations so far raise the question as to what happens, if we have g not in some $\mathcal{G}((d))$ for $d \in (0,\infty)$. We conjecture that two changes should occur. Assume that $g \in \mathcal{G}((d_m))$ where d_m is a sequence which converges to a number in $[0,\infty]$. (In fact, we are interested in the cases 0 and ∞ which correspond to $g(x)/x \to 0$ resp. $g(x)/x \to \infty$ as $x \to \infty$). The first change is now that the rescaling of the mass which was previously obtained by dividing by $(\sum_0^n c_k^{-1})\cdot$const, see (1.5), is now replaced by dividing by $d_n \cdot \sum_0^n c_k^{-1}$ which is asymptotically different if $d_n \to 0$ or $d_n \to \infty$ as $n \to \infty$. The second change is in the spatial distribution of the clusters. In the case $g(x)/x \to \infty$ large clusters can occur easier (if $g \in \mathcal{G}((d_n))$ while in the case $g(x)/x \to 0$ small clusters can occur easier. In particular we see that the dichotomy of Theorem 3 holds for all $g \in \mathcal{G}$ and hence is universal but some of the finer properties of the clusterformation (as clusterheight) should depend in a crucial way on the shape of $g(x)$ for x very large.

Conjecture It is an open problem to prove the following conjecture about the behaviour of the interacting system $X^N(t)$ for N fixed and $t \to \infty$.
(a) In regime I

$$\mathcal{L}((x_{\xi,f_\alpha(t)}(N^t)_{\alpha\in[0,1]}|x_{\xi,0}(N^t) \geq \varepsilon) \underset{t\to\infty}{\Longrightarrow} \mathcal{L}((Z_\alpha)_{\alpha\in[0,1]})$$

(b) In regime II ($c_k = c^k, c < 1$)

$$\widehat{P}(x_{\xi,0}((cN)^t) \geq \delta c^j) \underset{j\to\infty}{\longrightarrow} \widehat{F}^N_\infty(\delta)$$

$$\widehat{F}^N_\infty(\delta) \underset{N\to\infty}{\longrightarrow} \widehat{F}_\infty$$

c) Results on the historical process and family structure

This subsection deals with the *stable case* and in contrast to the last section consideration is given to both $X^N(t)$ and the $N \to \infty$ limiting object, that is, $(Z_k^\infty)_{k\in\mathbb{Z}^-}$. The subsection consists of five parts. In parts (i) and (ii) we construct the historical processes of the interacting process $X^N(t)$ resp. the interaction chain (Z_k^∞). Part (iii) relates the family structure of the branching system with the genealogical structure of the Fleming-Viot systems and (iv) studies some consequences of (iii) namely the spatial distribution of families. Part (v) relates the fixed N equilibrium behaviour of the historical process to the corresponding mean-field limit behaviour associated with $(Z_k^\infty)_{k\in\mathbb{Z}^-}$.

(i) Historical Representation and Family Decomposition I: Interacting Feller Branching Diffusions

In section 1a) we introduced the intuitive idea of a historical representation for a branching particle system and the related *backward tree* constituted by a randomly chosen family (Palm distribution). The *historical representation* then easily allows us to decompose the state of a component at a finite time t into different clans(families). The objective of this section is to first obtain an analogous historical representation for our system $\{x_\xi^N\}_{\xi\in\Omega_N}$ of interacting Feller branching diffusions and then in a second step the family decomposition.

The main tool is the historical process which is an enriched version of our infinite interacting system of Feller branching processes. Intuitively it can be obtained by passing to the *diffusion limit of the historical process* of the system of branching random walks described in section 1a) (see ([D], Theorem 12.3.1.1)). We begin by introducing the basic ingredients needed for the formal description of the *historical process* and the *equilibrium historical process.*

Consider the random walk, $(y_t)_{t\geq s}$, starting at time s on Ω_N with jump rate 1 and transition kernel $a(\cdot,\cdot)$ given in (0.8). Denote by $\Pi^t_{s,\xi} \in \mathcal{P}(D([s,\infty),\Omega_N)$ the law of the random walk started at ξ at time s and stopped at time $t > s$. For convenience we extend the time domain to $(-\infty, \infty)$ by setting $y_u \equiv y_s \ \forall u \leq s$. If $y := (y(s))_{s \in I\!R}$ denotes a path in Ω_N then we denote by $\widehat{y}_t := (y(s \wedge t))_{s \in I\!R}$, that is, the path stopped at time t. The *path process* $(\widehat{y}_t)_{t \in I\!R}$ associated to the random walk is a time inhomogeneous Markov process with state space $D((-\infty, \infty), \Omega_N)$ and time-inhomogeneous semigroup $S_{s,t}$ acting on $C_b(D((-\infty, \infty), \Omega_N))$ by

$$(S_{s,t}F)(\widehat{y}_s) := \Pi^t_{s,y_s} F(\widehat{y}_t).$$

Let $\tilde{E}$ be the set of measures on path space such that the time 0 projection gives an element in E, the state space of our interacting system:

$$\tilde{E} = \{\mu \in M(D((-\infty, \infty), \Omega_N) : \mu \circ y(0)^{-1} \in E\}.$$

The subspace of $\tilde{E}$ consisting of measures carried by the set of paths stopped at time t is denoted by:

$$\tilde{E}^t = \{\mu \in \tilde{E} : \mu(\{y : y(\cdot) \neq y(\cdot \wedge t)\}^c) = 0$$

We define now the historical process as follows:

Definition (Historical process)
The *historical process* $\{H^N(t)\}_{t \in I\!R+}$ is an $\tilde{E}$-valued Markov process (and $H^N(t) \in \tilde{E}^t$). The law $\tilde{P}_{s,\mu}$ of H^N starting at time s in a point in $\tilde{E}$, uniquely defined by $H^N(s) = \mu \in \tilde{E}^s$, is given in terms of its Laplace functional

$$\tilde{P}_{s,\mu} \exp(- < H^N(t), \Phi >) = \exp(- < \mu, V_{s,t}\Phi >), \quad \Phi \in C_+(D((-\infty, \infty), \Omega_N)),$$

where $V_{s,t}\Phi$ is the unique solution to

$$V_{s,t}\Phi = S_{s,t}\Phi - d \int_s^t (V_{s,u}\Phi)^2 du. \Box$$

Remark The probability laws of the random measures $H^N(t)$ as well as the equilibrium measures to be described below are infinitely divisible. This allows us to use as important tools the canonical representation and the associated Palm distributions. A brief review of these notions are given in Appendix 2.

Next we come to the equilibrium historical process. In the transient case there exists an *equilibrium measure* for this historical process by letting $s \to -\infty$ for suitable initial laws. The representation for the equilibrium historical measure for super-Brownian motion in dimensions $d \geq 3$ has been constructed in detail in Dawson and Perkins (1991). We now reformulate this result for our system of Feller branching diffusions on Ω_N (super random walk) in the next theorem. First we construct the needed ingredients.

A system of independent random walks on a countable group has a unique extremal equilibrium with density θ (namely a Poisson system with intensity measure $= \theta \cdot$ counting measure) and hence a unique entrance law. That is, there exists a unique collection of locally finite measures $\{\tilde{\lambda}_{\theta,s}\}_{s \in I\!R}$ on $D((-\infty, \infty), \Omega_N)$ such that (i) $\tilde{\lambda}_{\theta,s}$ is concentrated on $\{y : y(\cdot) = y(\cdot \wedge s)\}$, (ii) $\tilde{\lambda}_{\theta,s}(\{y : y(s) \in A\}) = \theta|A|$ where $|A|$ denotes the volume of A, that is, the number of sites in A, and (iii) for $t > s$ and $B \in \mathcal{B}(D([s, \infty), \Omega_N))$

$$\tilde{\lambda}_{\theta,t}(B) = \sum_\xi \tilde{\lambda}_{\theta,s}(\{y : y(s) = \xi\}) \Pi^t_{s,\xi}(B).$$

In other words $(\tilde{\lambda}_{\theta,s})_{s \in (-\infty,\infty)}$ is an entrance law for the path process associated with the random walk system.

In order to construct the *equilibrium historical decomposition* we first consider the historical process $(H^{N,s}(t))_{t \geq s}$ with initial state given by the measure $\tilde{\lambda}_{\theta,s}$ at time s and law

$\tilde{P}_{s,\tilde{\lambda}_{\theta,s}}$. The equilibrium historical structure turns out to be identical at all times (up to a time translation) and therefore for convenience we consider it at time 0. The law of $H^{N,s}(0)$ under the law just mentioned is denoted by $\tilde{P}^0_{s,\tilde{\lambda}_{\theta,s}}$ (this is a probability measure on $M(D((-\infty,\infty),\Omega_N)))$.

Since we want to construct a decomposition of the configuration into families, which are related we need to recall that $\mu \in \tilde{E}$ is a *clan measure* if there exists a $y \in D((-\infty,\infty),\Omega_N)$ such that $\mu(((\mathcal{C}(y))^c) = 0$ where

$$\mathcal{C}(y) := \{y' : \exists u \text{ such that } y'(s) = y(s) \quad \forall s \leq u\}.$$

A basic tool which will be used below is the Palm measure associated to the random measure $H_0^{N,-\infty}$ at a point $y \in D((-\infty,\infty),\Omega_N)))$. The reader not familiar with canonical measures and Palm measures may want to look at Appendix 2 before reading parts (c) and (d) of the next theorem. In the context of the historical process the Palm measure describes the subpopulation mass which shares a common ancestor with a given individual who has followed a given path y up to the present.

Theorem 7 (Family Decomposition of the Equilibrium Historical Process)
Assume that $\sum_{k=0}^{\infty} c_k^{-1} < \infty$. Then for $N \geq 2$ the random walk on Ω_N is transient:
(a) $\tilde{P}^0_{s,\tilde{\lambda}_{\theta,s}}$ converges weakly as $s \downarrow -\infty$ to the law, $\tilde{P}^0_{\theta,equil}$ of an infinitely divisible random measure, $H_0^{N,-\infty}$, on $D((-\infty,\infty),\Omega_N)$ with intensity

$$E(H_0^{N,-\infty}(A)) = \tilde{\lambda}_{\theta,0}(A)$$

and Laplace functional

$$E(\exp(- < H_0^{N,-\infty}, \phi >)) = \exp(-\bar{V}_0^N(\phi)),\ \phi \in C_+(D((-\infty,\infty),\Omega_N))$$

where

$$\bar{V}_0^N(\phi) = \lim_{s\to-\infty} \int (V_{s,0}^N\phi)(y)\tilde{\lambda}_{\theta,s}(dy)$$

(b) $X_0^{N,-\infty} := \Pi_0 H_0^{N,-\infty}$ is a version of the equilibrium random measure for the system of interacting Feller branching diffusions with density θ where $(\Pi_0\mu)(A) := \mu(\{y : y(0) \in A\})$.
(c) The canonical measure $R^N_{-\infty,0}$ of $H_0^{N,-\infty}$ is supported by the set of clan measures and satisfies

$$\int (1 - e^{-<m,\phi>}) R^N_{-\infty,0}(dm) = \bar{V}_0^N \phi.$$

(d) The Palm distribution, $(\hat{P}^N)_y$, of the canonical measure $R^N_{-\infty,0}$ of $H_0^{N,-\infty}$ at y is supported by the the set of y clan measures, that is, $\{\mu : \mu((\{\mathcal{C}(y)\}^c) = 0\}$ and has the Laplace functional (here $y^r(s) = y(s \wedge r)$)

$$\int e^{-\langle m,\phi\rangle} (\hat{P}^N)_y(dm) = \exp(-2d \int_{-\infty}^0 (V_{r,0}^N\phi)(y^r)dr). \tag{1.30}$$

Moreover the Palm distribution $(P^N)_y$ of $H_0^{N,-\infty}$ is defined by the following Radon Nikodym derivative:

$$(P^N)_y(B) := \frac{\int_B \tilde{P}^0_{\theta,equil}(d\mu)\mu(dy)}{\tilde{\lambda}_{\theta,0}(dy)}, \quad \tilde{\lambda}_{\theta,0} - a.s., \quad B \in \mathcal{B}(\tilde{E})$$

and this Palm distribution satisfies:

$$(P^N)_y = \tilde{P}^0_{\theta,equil} * (\hat{P}^N)_y. \tag{1.31}$$

(where $*$ denotes convolution). □

We conclude now this paragraph by constructing the *decomposition of $x_\xi(t)$ into clans* (families) using Theorem 7. Indeed Part (c) of Theorem 7 tells us that the equilibrium historical $H_0^{N,-\infty}$ can be decomposed into a countable superposition of clan measures (families). We shall label the families by the index $i \in \mathcal{I}$ and write the family decomposition as follows:

$$H_0^{N,-\infty}(\cdot) = \sum_{i \in \mathcal{I}} H_0^{N,-\infty}(i, \cdot)$$

where for each $i \in \mathcal{I}$, $H_0^{N,-\infty}(i, \cdot)$ is a clan measure. Then

$$M_0^N(i, \xi) := H_0^{N,-\infty}(i, \{w : w(0) = \xi\})$$

denotes the mass of the family indexed by i at the site $\xi \in \Omega_N$ at time 0. The corresponding normalized measure of the family in a k-block is given by

$$M_{-k}^N(i, \xi) := \frac{1}{N^k} \sum_{d(\xi', \xi) \leq k} M_0^N(i, \xi') \quad \xi \in \Omega_N, \quad i \in \mathcal{I}.$$

We have noted above that in the transient case the interacting system of Feller branching diffusions has a unique equilibrium with density θ. We next describe the associated equilibrium measure for the historical process H^N on the time interval $(-\infty, 0]$. To do this we first set as initial condition at time $-S(N)\beta_j(N)$, $\quad H_{-S(N)\beta_j(N)}^{N,-S(N)\beta_j(N)} = \tilde{\lambda}_{\theta,-S(N)\beta_j(N)}$. Then put

$$M_0^{N,j}(i, \xi) := H_0^{N,-S(N)\beta_j(N)}(i, \{w : w(0) = \xi\}) \qquad \xi \in \Omega_N, \ j \in \mathbb{N}, \ i \in \mathcal{I}, \tag{1.32}$$

$$M_{-k}^{N,j}(i, \xi) := \frac{1}{N^k} \sum_{d(\xi', \xi) \leq k} M_0^{N,j}(i, \xi') \qquad k \in \{0, 1, \ldots, j\}. \tag{1.33}$$

Then $M_{-k}^{N,j}(i, \xi)$ is the density of mass of the i-th family in the block around ξ of size k where j specifies the time scale and the corresponding equilibrium quantity is (recall $\beta_j(N) \to \infty$ as $j \to \infty$)

$$\lim_{j \to \infty} M_{-k}^{N,j}(i, \xi) = M_{-k}^N(i, \xi).$$

Under the Palm distribution of Theorem 7(d), the spatial distribution of mass at time 0 of the family to which a randomly chosen individual at ξ belongs (and which is contained in a block of size k), has Laplace functional

$$\int \exp(-2d \int_0^\infty (V_{-r,0}^N \phi)(y^r) dr) \, \Pi_{0,\xi}^\infty(dy).$$

(This follows as in (6.2) of [DP] using the fact that the entrance law $\tilde{\lambda}_{\theta,0}$ satisfies $\tilde{\lambda}_{\theta,0}(\{y(-\cdot) \in B, y(0) = \xi\}) = \theta \Pi_{0,\xi}^\infty(B)$. The latter fact is a simple consequence of the time reversibility of the random walk on Ω_N.) In particular the density of mass at time 0 of the family of a randomly chosen individual at ξ in a block of size k around ξ, denoted by $M_{-k}^N(\xi)$, satisfies

$$E(\exp^{-\lambda M_{-k}^N(\xi)}) = \int \exp(-2d \int_0^\infty (V_{-r,0}^N(\lambda \phi_k))(y^r) dr) \, \Pi_{0,\xi}^\infty(dy) \tag{1.34}$$

where $\phi_k := \frac{1}{N^k} 1_{d(\xi, \xi') \leq k}$.

By homogeneity we can let $\xi = (0, 0, \ldots)$ and we will henceforth suppress ξ in the notation.

We next introduce a *random relabelling* of the families i according to the size of the mass at level 0, that is, so that $M_0^{*N,j}(i)$ is nonincreasing in i. This way we derive from (1.31) new objects the *size-ordered family decomposition* of the interacting systems, which will be called

$$\{M_k^{*N,j}(i),\ i \in \mathcal{I}\}; \quad k \in \{-j, \dots, 0\}, \tag{1.35}$$

$$\{M_k^{*N}(i),\ i \in \mathcal{I}\}; \quad k \in \mathbb{Z}^-. \tag{1.36}$$

These objects are difficult to analyse directly. However we shall show in Theorem 11 that the density of a single family in blocks of size k converges in distribution as $N \to \infty$ to the size of the single family at time $-k$ in the historical representation of the interaction chain. In the latter process we can then calculate and estimate the quantities of interest. The historical enrichment of the interaction chain is constructed in the next subsection below.

(ii) Historical Representation and Family Decomposition II: Interaction Chain

As in the previous subsection for the interacting system we construct also in the case of the interaction chain first a historical representation and then we are able to give in a second step the desired family decomposition. In order to understand the construction the reader should recall (0.17) and that the equation

$$dx_t = c(\theta - x_t)dt + \sqrt{2x_t}dw_t$$

can be viewed as the diffusion limit of a *subcritical branching particle system with immigration* at rate $c\theta$. In a subcritical branching system with immigration the equilibrium state can be decomposed into clans each consisting of the set of descendents of the same immigrant. Hence in the diffusion limit x_t, the equilibrium state can be written as the sum of jumps of a process with independent increments, which is called Moran gamma process and will be constructed explicitely below.

The *historical representation* of the interaction chain will be obtained by constructing a particular version of the interaction chain with the help of a collection of the Moran gamma processes.

First recall that the transition kernels of the entrance law $\mathcal{L}((Z_k^\infty)_{k \in \mathbb{Z}^-})$ are given for $k \in \mathbb{N}$ by $K_{-k}(\theta, \cdot) = \Gamma_\theta^{c_{k-1}, dx}$, the latter being an infinitely divisible law. Since we shall fix the function $g(x) = dx$, we put

$$\Gamma_\theta^{k-1} := \Gamma_\theta^{c_{k-1}, dx}(\cdot).$$

Γ_θ^k has a Laplace transform of the form $\exp(-\theta\psi_k(\lambda))$ with $\psi_k(\lambda) = c_k/d \log(1 + (d/c_k)\lambda)$. Such infinitely divisible laws on $(0, \infty)$ have Laplace transforms which can be represented using the Lévy measure $R_k(du)$ (see e.g. Dawson 1993, Section 3.3) as follows:

$$\exp(-\theta\psi_k(\lambda)) = \exp\Big(-\theta \int_0^\infty (1 - e^{-\lambda u}) R_k(du)\Big).$$

In fact we can explicitly give the Lévy measure corresponding to the ψ_k:

$$\psi_k(\lambda) = \frac{c_k}{d} \int_0^\infty (1 - \exp(-u\lambda d/c_k)) \frac{e^{-u}}{u} du \tag{1.37}$$

$$= \frac{c_k}{d} \int_0^\infty (1 - e^{-u\lambda}) \frac{e^{-\frac{c_k u}{d}}}{u} du. \tag{1.38}$$

Next we consider the subordinator known as the *standard Moran gamma process* $\gamma(s)$, that is, the nondecreasing process with independent increments and with Lévy measure

$R(du) = \frac{e^{-u}}{u}du$ (cf. A2.3). Then for $s_2 > s_1$, the increment $\gamma(s_2) - \gamma(s_1)$ has distribution with Laplace transform

$$\exp\Big\{ -(s_2 - s_1) \int (1 - e^{-u\lambda}) \frac{e^{-u}}{u} du \Big\}.$$

For each $k \in \mathbb{Z}$ we then define the process

$$\gamma_k(u) := \frac{d}{c_k} \gamma(\frac{c_k}{d} u). \tag{1.39}$$

Then it is easy to verify that $\Gamma^k_\theta = \mathcal{L}(\gamma_k(\theta))$ and $\gamma_k(\theta)$ is the sum of a countable set of jumps of the Moran process in the time interval $[0, \theta)$.

We are now ready to construct the *historical process* associated with the interaction chain as an enriched version of our interaction chain $(Z^\infty_k)_{k \in \mathbb{Z}^-}$. Namely the first step is to define the enriched version of $\{Z^\infty_k\}_{k \in \{-j, \ldots, 0\}}$ conditioned on Z^∞_{-j-1} and the second step is to let $j \to \infty$. We first build on one probability space a sequence of independent Moran gamma processes $\{\gamma_k(u) : 0 \le u < \infty\}_{k \in \mathbb{Z}^+}$ defined in terms of the standard Moran gamma process by (1.39). Next define for $k \in 0, 1, \ldots, j$

$$z^j_{-k}(u) := \gamma_k(\gamma_{k+1}(\cdots \gamma_j(u))), \quad u \ge 0 \tag{1.40}$$

One can prove that with the above objects we get a version of $(Z^\infty_k)_{k \in \mathbb{Z}^-}$ as follows:
Lemma 1.1 *Assume that* $\sum c_k^{-1} < \infty$. *Then for* $k \in \mathbb{N}$
(a)

$$z^\infty_{-k}(u) := \lim_{j \to \infty} z^j_{-k}(u)$$

exists a.s. and has mean $E[z^\infty_{-k}(u)] = u$.
(b) Define for $k \in \mathbb{N}$

$$Z^\infty_{-k} := z^\infty_{-k}(\theta). \tag{1.41}$$

Then $\{Z^\infty_k\}_{k \in \mathbb{Z}^-}$ *is a version of the entrance law for the Markov chain with transition kernels* $K_{|k|}$. □

We shall next in Theorem 8 below construct the *historical representation* $H = (\mathcal{H}_n)_{n \in \mathbb{Z}^-}$ associated with the entrance law $(Z^\infty_k)_{k \in \mathbb{Z}^-}$. This will be a process H with values in the functions $\mathbb{N} \to \mathbb{N} \times \mathbb{R}^+$ and with time index set $\mathbb{Z}^-$. The interpretation is the following. For the "individual" contributing at time 0 the i-th biggest mass we specify the rank of the ancestral mass at time $-k$ and it's mass contribution at this time.

The process H will be constructed on the probability space $(\Omega, \mathcal{F}, P)$ generated by the collection of independent Moran Gamma processes given by (1.39). Each jump of γ_k is interpreted to be an *individual* and the set of jumps of γ_k occurring in a time interval representing a single jump of γ_{k+1} is interpreted as the set of descendents at level k of single individual at level $k+1$. We will see later in Theorem 11 that we can reinterpret the latter in terms of the interacting system as the set of descendents of an individual at distance $k+1$ from ξ which immigrate into in a k-block around ξ in the limit $N \to \infty$.

Remark In this section we will complement the methods of infinitely random measures with relations between Gamma and Beta distributions and also the Poisson Dirichlet distribution, size-biased sampling and the GEM representation. A brief review of the facts that we will need is given in Appendix 3.

Theorem 8 (Historical Representation of the Entrance Law)
Assume that $\sum c_k^{-1} < \infty$. *Then there exists a* $(\mathbb{N} \times \mathbb{R}^+)^{\mathbb{N}}$*-valued process* $\{\mathcal{H}_n\}_{n \in \mathbb{Z}^-}$, $\mathcal{H}_{-k}(i) = (G_{-k}(i), m_{-k}(G_{-k}(i)))$, $k \in \mathbb{N}$, *defined on the same probability space such that for* $k \in \mathbb{N}$

(i) $m_{-k}(1), m_{-k}(2), \ldots$ *is a non-increasing sequence*

(ii)
$$z^{\infty}_{-k}(\theta) = \sum_{\ell \in I\!N} m_{-k}(\ell)$$

(iii) Conditioned on $\{z^{\infty}_{-k}(\theta)\}_{k\in I\!N^-}$ *the* $\{\frac{m_{-k}(\cdot)}{Z^{\infty}_{-k}}\}_{k\in I\!N^-}$ *are independent and for each* k, $\{\frac{m_{-k}(i)}{Z^{\infty}_{-k}}\}_{i\in\mathcal{I}}$ *has the Poisson Dirichlet distribution with parameter* $\frac{Z^{\infty}_{-k-1}c_k}{d}$.

(iv) $G_0(i) = i$ *and*

$$G_{-k}(i_1) = G_{-k}(i_2) \Rightarrow G_{-k'}(i_1) = G_{-k'}(i_2) \quad \forall \quad k' > k$$

(v) conditioned on $\{\mathcal{H}_{-k-1}(\cdot), \mathcal{H}_{-k-2}(\cdot), \ldots\}$, $\{G_{-k}(j)\}_{j\in I\!N}$ *are independent and*

$$P(G_{-k}(j) = i | m_{-k}(\cdot), m_{-k-1}(\cdot), \ldots) = \frac{m_{-k}(i)}{z^{\infty}_{-k}(\theta)}$$

(vi) conditioned on $\{m_{-k-1}(\cdot), m_{-k-2}(\cdot), \ldots\}$, $\{\sum_i m_{-k}(i)1_{G_{-k-1}(i)=j}\}_{j\in I\!N}$ *are independent and*

$$\begin{aligned}(1.42) \qquad & \mathcal{L}(\sum_i m_{-k}(i)1_{G_{-k-1}(i)=j} | m_{-k-1}, m_{-k-2}, \ldots) \\ & = Gamma(\frac{c_{k+1}m_{-k-1}(j)}{d}, \frac{d}{c_{k+1}}).\square\end{aligned}$$

Definition (Historical representation of the entrance law (Z^{∞}_k)).
The collection $\{\{G_{-k}(i), m_{-k}(G_{-k}(i))\}_{i\in I\!N}\}_{k\in I\!N}$ is called the *historical representation* of the equilibrium system associated with the entrance law $(Z^{\infty}_k)_{k\in\mathbb{Z}^-}$. □

Corollary 1 *If* $i_1, \ldots, i_j$ *belong to different families, then*

$$(1.43) \qquad \mathcal{L}(\{(\frac{c_k}{d}m_{-k}(G_{-k}(i_\ell)))_{\ell=1,\ldots,j,k\geq K}) \underset{K\to\infty}{\Longrightarrow} \text{independent exponential(1) laws}$$

and for $\ell \in I\!N$,

$$(1.44) \qquad E(\frac{c_k}{d}m_{-k}(G_{-k}(i_\ell))) \underset{k\to\infty}{\longrightarrow} 1. \qquad \square$$

We now have the tools to proceed with the second step in (ii) and to give the *family decomposition* of $(Z^{\infty}_k)_{k\in\mathbb{Z}^-}$. We introduce some equivalence relations on $I\!N$ such that for $k \in I\!N$, $i_1 \equiv_k i_2$, (resp. $i_1 \equiv i_2$) iff $G_{-\ell}(i_1) = G_{-\ell}(i_2)$ for some $\ell \in \{-k, \ldots, 0\}$, (resp. $\ell \in \mathbb{Z}^-$), (i.e. i_1 and i_2 have a *common ancestor of order k, (resp. of some order))*. We will show that under the hypothesis $\sum c_k^{-1} < \infty$, with probability one there are countably many distinct equivalence classes and we will label the equivalence classes (families or clans) by the index $n \in \mathcal{I}$. Given $i \in I\!N$ the mass of the *family* of i, that is the equivalence class to which it belongs, at level 0 is defined by

$$(1.45) \qquad M_0(i) := \sum_{\ell\equiv i} m_0(\ell).$$

We will now size order the collection $\{M_0(i)\}_{i\in I\!N}$ and label the equivalence class by the integer giving the rank, $\ell(i)$, of $M_0(i)$ in this order. This way the labels of the families can be identified with $I\!N$. We then define

$$(1.46) \qquad M^*_0(n) := M_0(\ell^{-1}(n)), \quad n \in \mathcal{I}.$$

Thus families are labelled so that the family masses at level 0 are ordered. We now introduce the corresponding family masses at level k, namely

$$M_k^*(n) := \sum_{i \in \ell^{-1}(n)} m_{-k}(G_{-k}(i)). \tag{1.47}$$

Thus we have now also decomposed the entrance law into *clans* (or *families*)

$$Z_k^\infty = \sum_{n \in \mathcal{I}} M_k^*(n), \qquad k \in \mathbb{Z}^-. \tag{1.48}$$

Given $j \in \mathbb{N}$ we can carry out the same decomposition of the law of $\{Z_k^j\}_{k \in \{-j,\dots,0\}}$ using the equivalence relation $\equiv_j$ instead of $\equiv$. This yields the collection $\{M_k^{*j}(i)\}_{k \in \{-j,\dots,0\}}$. However for this object we can pass to the limit $j \to \infty$ to get the object discussed above: for each $i \in \mathcal{I}$,

$$\mathcal{L}((M_k^{*j}(i))_{k \in \{-j,-j+1,\dots,0\}}) \underset{j \to \infty}{\Longrightarrow} \mathcal{L}((M_k^*(i))_{k \in \mathbb{Z}^-}).$$

In the proof of the results a refinement of the above considerations is useful. Note that $M_{-k}^*(i)$, the family mass at level k, corresponds to the total mass at level k of a subpopulation that has a common ancestor at some level $\ell > k$. In order to study the behaviour of $M_k^*(i)$ as $k \to -\infty$ we now introduce a finer decomposition based on the the notion of last common ancestor where we say that $i_1,\ i_2$ have a *last common ancestor* at level $|k|$ if

$$k = max\{-\ell : \ell \in \mathbb{N},\ G_{-\ell}(i_1) = G_{-\ell}(i_2)\}.$$

Then a family mass at level $k \in \mathbb{N}$, can be decomposed into subfamilies. Namely if we are given the family of $i \in \mathbb{N}$ and its mass at level k we can decompose this mass into the mass contributed by relatives of i or order exactly ℓ which are those masses with index j such that $G_{-k-\ell}(i) = G_{-k-\ell}(j)$ and $G_{-k-\ell-1}(i) \neq G_{-k-\ell-1}(j)$.

Then the quantity $M_k^*(i)$ can for every $i \in \tilde{\mathcal{I}}$ be decomposed:

$$M_{-k}^*(i) = \sum_{\ell=k+1}^{\infty} M_{-k,-\ell}^*(i), \quad k \in \mathbb{N} \tag{1.49}$$

and we can identify the law of this decomposition explicitly:

Corollary 2 (Family Decomposition)
(a) Conditioned on $\{M_{-j}^*(i)\}_{i \in \mathcal{I}}$, $\{\{M_k^*(i)\}_{k \in (-j,\dots,0)}\}_{i \in \mathcal{I}}$ is a collection of independent Markov chains with time index $\mathbb{Z}^-$ each with the same transition kernels K_k ($k \in \mathbb{N}$) at time $-k$ and which are also martingales.
(b) For each $i \in \mathcal{I}$ and $k \in \mathbb{N}$,

$$M_{-k,-\ell}^*(i) \quad = \quad \hat{Z}_k^\ell(m_{-\ell}(G_{-\ell}(i))), \quad \ell \geq k+1$$

where $\{\hat{Z}_k^j(m_{-j}(G_{-j}(i)))\}_{j \in \mathbb{N}}$ are independent copies of the interaction chain starting at level j in $m_{-j}(G_{-j}(i))$.

(iii) Relation to the Interaction Chain associated to a System of Interacting Fleming-Viot Processes

Using the representation in terms of the Moran gamma processes we will now obtain rather detailed information about the "entrance law" $\{M_{-k}^*(i); i \in \mathcal{I}\}_{k \in \mathbb{Z}}$ describing the different family masses in the entrance law. In particular the main objective of this section is to characterize the law of the sequence of normalized $\mathcal{P}(\mathbb{N})$-valued random variables

$$\{Q_{-k}^*(i)\}_{i \in \mathcal{I}} := \{\frac{M_{-k}^*(i)}{Z_{-k}^\infty}\}_{i \in \mathcal{I}}, \quad k \in \mathbb{N}$$

where $\mathcal{P}(\mathbb{N})$ denotes the set of probability measures on $\mathbb{N}$. Those random variables give the relative proportions the various families contribute to the total mass present at time $-k$ (which corresponds to the contribution various families give to the mass in a k block at large times and in equilibrium in the original interacting system).

The first question is to whether these laws $\{\mathcal{L}(Q_k^*)\}_{k\in\mathbb{Z}^-}$ are again the entrance law of some Markov chain, at least if we condition on $(Z_k^\infty)_{k\in\mathbb{Z}^-}$? The answer will be yes and we will characterize its law in terms of the interaction chain associated with a system of interacting Fleming-Viot processes. In addition we will compare the *family structure* of our *branching systems* with the *type structure* of the corresponding systems of *interacting Fleming-Viot processes* studied in Dawson, Greven and Vaillancourt (1995).

We now define the ingredients of the interacting Fleming-Viot system. Let Θ denote the uniform distribution on $[0,1]$.

(1.50) $\qquad \Lambda_\theta^{c,d} \in \mathcal{P}(\mathcal{P}([0,1])) \qquad$ is defined as:

$$\Lambda_\theta^{c,d} = \mathcal{L}\Big(\sum_{l=1}^{\infty}\Big(\prod_{i=1}^{l-1}(1-V_i)V_l\Big)\delta_{U_l}\Big)$$

where $(U_l)_{l\in\mathbb{N}}$ are i.i.d. Θ-distributed $[0,1]$-valued random variables and the $(V_i)_{i\in\mathbb{N}}$ are independent Beta(1,c/d) distributed $[0,1]$-valued random variables.

Next define a random sequence $(c_k')_{k\in\mathbb{N}}$ by requiring that for given $(c_k)_{k\in\mathbb{N}}$, d and $(Z_k^\infty)_{k\in\mathbb{Z}^-}$:

$$\frac{c_k'}{d_k'} = \frac{c_k}{d} Z_{-k-1}^\infty, \qquad d_k' = \frac{d_0'}{d_0' + \sum_0^{k-1}(c_j')^{-1}}. \tag{1.51}$$

For every $k\in\mathbb{N}$ define a transition kernel $L_{-k}(\cdot,\cdot)$ on $\mathcal{P}([0,1])\times\mathcal{P}([0,1])$ by setting

$$L_{-k}(\rho, d\varphi) = \Lambda_\rho^{c_k', d_k'}(d\varphi). \tag{1.52}$$

If $\sum \frac{d_k'}{c_k'} < \infty$, then according to [DGV], Theorem 0.5 there exists a unique *entrance law* for the sequence $(L_k)_{k\in\mathbb{Z}^-}$ starting in $\Theta\in\mathcal{P}([0,1])$. Choose Θ equal to the uniform distribution on $[0,1]$. Then this entrance law is $\mathcal{P}([0,1])$-valued stochastic process and is denoted by $(\eta_k^\infty)_{k\in\mathbb{Z}^-}$. The states of this process are in fact atomic probability measures on $[0,1]$. Write (x_i is the position of the i-th atom)

$$\eta_k^\infty = \sum_{i=0}^{\infty} m_k^\infty(i)\delta_{x_i} \quad k\in\mathbb{Z}^-. \tag{1.53}$$

We size-order the collection $\{m_0^\infty(\cdot)\}$ and let $\ell(i)$ denote the rank of $m_0^\infty(i)$ in this order. Then let

$$m_k^{*\infty}(j) := m_k^\infty(\ell^{-1}(j)), \quad j\in\mathbb{N}, k\in\mathbb{Z}^-. \tag{1.54}$$

Now define a probability measure q_k^* on $\mathbb{N}$ by setting:

$$q_k^*(\{j\}) := m_k^{*\infty}(j) \qquad j\in\mathbb{N}, k\in\mathbb{Z}^-. \tag{1.55}$$

With the last object we have the main ingredient for:

Theorem 9 (Relation with Fleming-Viot, family dynamics)
Assume that $\sum c_k^{-1} < \infty$.

(a) Then $\sum \frac{d_k'}{c_k'} < \infty$ *and conditioned on* $(Z_k^\infty)_{k\in\mathbb{Z}^-}$, $(Q_k^*)_{k\in\mathbb{Z}^-}$ *is characterized by*

$$\mathcal{L}((Q_k^*)_{k\in\mathbb{Z}^-} | (Z_k^\infty)_{k\in\mathbb{Z}^-}) = \mathcal{L}((q_k^*)_{k\in\mathbb{Z}^-}). \tag{1.56}$$

(b) The chain $(q_k^*)_{k\in\mathbb{Z}^-}$ can (in addition to (1.54) and (1.55)) also be described as follows: $(q_k^*)_{k\in\mathbb{Z}^-}$ is a $\mathcal{P}(\mathbb{N})$-valued time inhomogeneous Markov chain with transition function

$$K_k(\{p_\ell\}_{\ell\in\mathbb{N}},\cdot) := \mathcal{L}\Big(\sum_{i=1}^{\infty}\Big[V_i^k\prod_{\ell=1}^{i-1}(1-V_\ell^k)\Big]\delta_{\bar U_i}\Big) \in \mathcal{P}(\mathcal{P}(\mathbb{N})) \tag{1.57}$$

where

(i) $(\bar U_i)_{i\in\mathbb{N}}$ is an i.i.d. sequence of $\mathbb{N}$-valued random variables with marginal distribution $\{p_\ell\}_{\ell\in\mathbb{N}}$

(ii) $(V_i^k)_{i\in\mathbb{N}}$ is an i.i.d. sequence with marginal Beta$(1,\frac{c'_k}{d'_k})$.

(iii) $(\bar U_i)_{i\in\mathbb{N}}$ and $(V_i^k)_{i\in\mathbb{N}}$ are independent processes. □

Remark Thus the normalized process $(q_k^*)_{k\in\mathbb{Z}^-}$ conditioned on the total mass process Z_{-k}^∞ is the entrance law of the interaction chain associated to the system of interacting Fleming-Viot processes with migration coefficients $(c'_k)_{k\in\mathbb{N}}$ and sampling coefficients $(d'_k)_{k\in\mathbb{N}}$. Note that this is an analog of the result of Perkins (1991) which states that a measure-valued branching process conditioned on the total mass process and normalized is a Fleming-Viot process with time inhomogeneous sampling rate. Also note that even though $c'_k\neq c_k$, c'_k/c_k converges as $k\to\infty$ almost surely to a constant in $(0,\infty)$ if $\sum c_k^{-1}<\infty$. The background for this phenomenon is as follows. In a stable branching system, resp. Fleming-Viot system, the weights of different families at a site are small. Since the evolution of branching and Fleming-Viot differ only in that the latter is restricted to total sum 1, this difference "disappears" when considering a fixed number of small subpopulations.

(iv) Spatial Distribution of Families

Let $\widehat{M}_k^*(\cdot)$ denote for every k the size-reordered values of $M_k^*(\cdot)$ (recall the * indicated reordering according to the rank for $k=0$). We can then introduce the notion of the number of types making up a proportion $(1-\delta)$ of the total mass in a k-block:

$$T_k(\delta) = \inf(j|\sum_{i=1}^{j}\widehat{M}_{-k}^*(i)/Z_{-k}^\infty \geq 1-\delta), \quad \delta\in(0,1). \tag{1.58}$$

Theorem 10 (Density of families, number of families)
Assume that $\sum c_k^{-1}<\infty$.

a) $$\max_i M_{-k}^*(i) \underset{k\to\infty}{\longrightarrow} 0. \tag{1.59}$$

b) Assume that the $(c_k)_{k\in\mathbb{N}}$ are of the form $c_k=bc^k$ with $c>1$. Then the sequences

$$\mathcal{L}\Big(T_k(\delta)\sum_{\ell\geq k}c_\ell^{-1}\Big), \qquad \mathcal{L}(T_k(\delta)c_k^{-1}) \tag{1.60}$$

are tight and limit laws are nontrivial.

c) For fixed $i\in\mathbb{N}$,

$$\frac{E(M_{-k}^*(i))}{\sum_{\ell\geq k}\frac{d}{c_\ell}} \underset{k\to\infty}{\longrightarrow} 1. \tag{1.61}$$

Remark The result a) says that a single family surviving indefinitely will have spatial density 0. Note that (1.60) implies that for any fixed pair $i_1, i_2 \in I\!N$,

$$\frac{E(M^*_{-k}(i_1))}{E(M^*_{-k}(i_2))} \underset{k\to\infty}{\longrightarrow} 1.$$

Therefore in large enough blocks different families have approximately the same mean density. If $c_k = bc^k, c > 1$, then the number of different such families in a k-block grows exponentially in k .

(v) Convergence of the historical representations and family decompositions

In this section we pursue two (related) goals. First we construct and identify the analogue to the "*Kallenberg backward tree*" for the interaction chain and then we study the convergence of the historical representation of X^N_ξ to that of the interaction chain as $N \to \infty$.

Part (b) of Theorem 4 states that the vector of block averages under the equilibrium measure has laws $\tilde{\nu}^N_\theta$ which converge as $N \to \infty$ to the entrance law ν^∞_θ of the interaction chain. In this section we strengthen this to include the convergence of the corresponding family decompositions and in fact also describe the sense in which the historical processes converge. This can be made precise using the framework of *Palm distributions* which describe the family decomposition of a "randomly chosen family" also known as the *backward tree* (refer to Appendix 2 for some basic facts on Palm distributions).

In order to get an intuitive idea of this consider again the Branching Random Walk on Ω_N, look at the population in a (k+1)-block $\{\xi' \in \Omega_N | d(\xi, \xi') \leq k+1\}$, then choose an individual at random from the (k+1)-block and finally determine which k-subblock of this (k+1)-block this individual belongs to. This involves *size-biased sampling* which means that k-subblocks which contain more mass will have a higher probability of being chosen. Size-biased sampling also occurs, if we choose an "individual" at $\xi \in \Omega_N$ at random and consider the family (clan) associated with this randomly selected individual and the masses contributed by this family to blocks $\{\xi | d(\xi, \xi') \leq k\}$. This latter random object is described by the *Palm distribution of the family represention* which we will now introduce rigorously Again the Palm measure reflects the fact that it is more likely to choose an individual from a larger family..

In order to consider the Palm distribution of the equilibrium historical process recall that $\tilde{P}^0_{\theta,equil}$ is the law of the equilibrium historical measure $H^{N,-\infty}_0$ (which is a locally finite random measure on $D((-\infty,\infty),\Omega_N))$. For $y \in D((-\infty,\infty),\Omega_N))$ and $y(s) = \xi\ \forall s \geq 0$, we denoted by $(P^N)_y$ the historical Palm distribution at y (see A2 for a definition of the Palm distribution of a random measure). The law $(P^N)_y$ is the law of a random measure on $D((-\infty,\infty),\Omega_N))$ which intuitively corresponds to the family mass of an individual whose spatial (random walk) trajectory is given by y. What we now want to consider is however a typical family tree y which is at ξ at time 0. This we construct next.

Recall that $\Pi^\infty_{0,\xi}$ denotes the law of the random walk started at time 0 at site ξ. Then , based on $(P^N)_y$, we define the *Palm distribution at ξ of the family representation* of $H^{N,-\infty}_0$, denoted by $(P^N)_{\xi,0}$:

Definition (Palm measure of family representation)
Let $A \in \sigma(\{M^N_{-k}(\cdot,\cdot)\}_{k\in Z\!\!Z^-})$ and set

$$(1.62) \qquad (P^N)_{\xi,0}(A) := \int \Pi^\infty_{0,\xi}(dy)(P^N)_y(A). \qquad \square$$

This means that we are considering the law of the family of an individual chosen at site ξ and in the above we have averaged the historical Palm distribution over all possible trajectories that could have been followed by the ancestral individual.

Recall that the *backward tree* is the family tree of a "particle" chosen at random at ξ at time 0 in the equilibrium process. In order to discuss the convergence of the x^N_ξ backward tree to the one of Z_k^∞ we have to deal with the following problem: The family decomposition in the case of the infinite system $(x_\xi^N)_{\xi\in\Omega_N}$ is based on the idea that any two individuals in the same family have a common ancestor at some time in the past and this information is contained in the historical process. On the other hand in the case of the interaction chain the time parameter corresponds to a spatial characterization of the interacting system (in the limit $N \to \infty$). Therefore in order to discuss the convergence of the historical processes we need to relate the time parameter to the spatial parameter in the historical representation of the interaction chain. Furthermore we need to describe an analogue of the Palm distribution representation obtained in Theorem 8. The family decomposition of a given individual for the infinite system is based on a "backward" (or "bottom-up") decomposition of the members of the family according to the times at which they have a *last common ancestor* with the given individual. On the other hand the interaction chain is a "forward description" (or "top-down"). In order to unify the two views discussed above we consider the decomposition of the infinite family under the Palm distribution of the family representation for X^N (denoted $(P^N)_{\xi,0}$) as follows; $\bar{M}^N_{0,-k}(\cdot)$ denotes the spatial distribution at time 0 of the subfamily for which the time of the last common ancestor lies in $(\tau^N_{k-1}, \tau^N_k]$ where τ^N_k is the exit time from the ball of radius k for the "backbone" of the family (or path starting at ξ). We then have the *historical Palm representation* of the family of a randomly selected individual at site ξ

$$\bar{M}^N_0 = \sum_{\ell=1}^\infty \bar{M}^N_{0,-\ell}(\{\xi\}).$$

Note that also for the interaction chain we will also use a bar to denote objects such as M_k^*, m_k under the Palm distribution, which we introduce below in (1.63). Also, without loss of generality we can consider the site $\xi = 0$.

We now turn to the interaction chain. We will first introduce the analogue of the Palm distribution $(P^N)_y$ used above. To motivate this definition note that picking a path of descent at random from a site in the $N < \infty$ case introduces a size biasing according to the family size at this site. The appropriate reweighting in the $N = \infty$ limit turns out to be:

Definition (Palm distribution of the historical representation $\mathcal{H}$ with respect to Z_0^∞)
Let P^∞ be the law of the process $(\mathcal{H}_n)_{n\in\mathbb{Z}^-}$ constructed in our historical representation of the entrance law (Compare Theorem 8). Then define $(P^\infty)_0$ by

$$(P^\infty)_0(B) := \theta^{-1}\int Z_0^\infty \cdot 1_B(\mathcal{H})P^\infty(d\mathcal{H}) \quad \text{for} \quad B \in \sigma(\{\mathcal{H}_n\}_{n\in\mathbb{Z}^-}).\square \tag{1.63}$$

In particular the Palm measure $(P^\infty)_0$ is the law of a $(\mathbb{N}\times\mathbb{R}^+)^{\mathbb{N}}$-valued process $(\bar{\mathcal{H}}_n)_{n\in\mathbb{Z}^-}$ with $\bar{\mathcal{H}}_k(i) = (\bar{G}_{-k}(i), \bar{m}_{-k}(\bar{G}_{-k}(i)))$. Then we can define (analogously to (1.49)):

$$\bar{M}^*_{-k} = \sum_{\ell=k}^\infty \bar{M}^*_{-k,-\ell}.$$

Recall the interpretation of the historical decomposition: $\bar{M}^*_{-k,-\ell}(i)$ is the mass at level k contributed by the relatives of i of order exactly ℓ but now (bar) under the Palm measure $(P^\infty)_0$. Define: $(\widehat{P}^\infty)_0 := \mathcal{L}((\bar{M}^*_{0,\ell})_{\ell\in\mathbb{Z}^-})$

The law of the decomposition of $\bar{M}^*_{-k}$ can explicitly be identified and is now exactly what corresponds in the limit $N \to \infty$ to the modified historical Palm distribution in the interacting system $X^N(t)$ introduced above. We start by describing $(P^\infty)_0$ explicitly.

Theorem 11 (Historical Palm representation and the family representation)
(a) Under the Palm distribution, $(P^\infty)_0$ the sequence $\{\bar{m}_{-k}\}_{k\in\mathbb{Z}}$ consists of independent

exponentially distributed random variables with means $\frac{d}{c_k}$.
(b) The restriction of $(P^\infty)_0$ *and* P^∞ *to* $\sigma\{M^*_{0,-\ell}\}$ *satisfy*

$$(1.64) \qquad (P^\infty)_0 = (\widehat{P}^\infty)_0 * P^\infty$$

and $\bar{M}^*_{-k}$ *has the representation*

$$(1.65) \qquad \bar{M}^*_{-k} = \sum_{\ell=k}^{\infty} \bar{M}^*_{-k,-\ell} = \sum_{j=k}^{\infty} \hat{Z}^j_k(\bar{m}_{-j})$$

and the $\hat{Z}^j_k(\bar{m}_{-j})$ *are independent and have the same distribution as the interaction chain at level* k *started at* $\bar{m}_{-j}$ *at level* j. □

The next statement shows that indeed $(P^\infty)_0$ is the right object to describe $(P^N)_{\xi,0}$ in the limit $N \to \infty$.

Theorem 12 (Convergence of Palm representation and the family representation)
(a) (Convergence of the modified historical Palm representations)
The historical decompositions $\{\bar{M}^N_{0,j}(\xi)\}_{j\in \mathbb{Z}^-}$ *converge weakly as* $N \to \infty$ *to the process* $\{\bar{M}^*_{0,j}\}_{j\in\mathbb{Z}^-}$. *where* $\bar{M}^*_{0,j} = \hat{Z}^j_0(\bar{m}_j)$.
(b) (Convergence of the family representations)

$$(1.66) \qquad \mathcal{L}((M^{*N,j}_k(i))_{k\in\{-j-1,-j,\dots,0\}}, i \in \mathcal{I}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}((M^{*j}_k(i))_{k\in\{-j-1,-j,\dots,0\}}, i \in \mathcal{I}),$$

$$(1.67) \qquad \mathcal{L}((M^{*N}_k(i))_{k\in\mathbb{Z}^-,\ i\in\mathcal{I}}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}(M^*_k(i))_{k\in\mathbb{Z}^-,\ i\in\mathcal{I}}) \qquad \square$$

Remark In (a) the term $\hat{Z}^j_0(\bar{m}_{-j})$ denotes the mass of the subpopulation at site 0 for which the "closest common ancestor" is at distance $k+1$. Recalling that $\frac{\tau^N_k}{N^k}$ is an exponential random variable this corresponds to the subpopulation whose last common ancestor appears in the N^k time scale.
Remark If $c_k = k^n$ with $n > 1$, then $\frac{(c_k)^{-1}}{\sum_{j=0}^k (c_j)^{-1}} \to 0$ and $k \to \infty$ and the $\frac{M^*_{-k}}{E(M^*_{-k})} \to 1$ in probability as $k \to \infty$. On the other hand if $c_k = c^k$ with $c > 1$ then $\{\frac{M^*_{-k}}{E(M^*_{-k})}\}_{k\geq j}$ is asymptotically self-similar as $j \to \infty$. Note that this limiting behaviour is shared by both the interacting system of Feller branching diffusions and the system of interacting Fleming-Viot processes suggesting a *new type of universality result* for the *transient regime*.

Part B Proofs

2 Preparations

In this section we prepare some basic tools needed frequently for the proofs of our theorems. There are first of all formulas and equations for the moments (first and second) of systems of interacting diffusions, of certain one dimensional diffusions and of the interaction chain. Second there are properties of the map $\mathcal{F}_c, \mathcal{F}^{(n)}$ and the map $G_{c,g}: \theta \to \Gamma_\theta^{c,g}$ which are essential in many arguments later on. See (0.16), (0.20), (0.17) and (0.18) for definitions. Finally we establish uniform continuity properties for the kernels appearing in the interaction chain.

a) Moment calculations

Since we shall need during the proofs not only the original system defined in the introduction we proceed in a sufficiently general setting for our calculations of moments. Consider a finite or countable group I and a homogeneous transition kernel $b(\cdot,\cdot)$ on $I \times I$. Let g be as in (0.4). Define a process $(Y(t))$ through the following system of stochastic differential equations.

$$Y(t) = (y_i(t))_{i \in I}$$

$$dy_i(t) = \sum_{j \in I} b(j,i)(y_j(t) - y_i(t))dt + \sqrt{2g(y_i(t))}dw_i(t) \tag{2.1}$$

$$Y_{(0)} \in E \qquad (E \text{ as in } (0.6)).$$

Lemma 2.1 *Define* $b_t(\cdot,\cdot) = \sum\limits_{n=0}^{\infty} e^{-t} \cdot \frac{t^n}{n!} b^n$ *and* $\hat{b}(i,j) = (b(i,j) + b(j,i))/2$. *Then*

$$Ey_i(t) = \sum_{j \in I} b_t(i,j) y_j(0) \tag{2.2}$$

$$E(y_i(t)y_j(t)) = \sum_{k,l} b_t(i,k) b_t(j,l) y_k(0) y_l(0) + \int\limits_0^t \sum_k b_{(t-s)}(i,k) b_{t-s}(j,k) Eg(y_k(s))ds. \tag{2.3}$$

If $Y(0)$ is random and this initial law is homogeneous then

$$E(y_0(t)y_i(t)) = \sum_j \hat{b}_{2t}(i,j) Ey_0(0)y_j(0) + \int_0^t \hat{b}_{2(t-s)}(0,i) Eg(y_0(s))ds \qquad \square$$

Proof The proof proceeds via the Ito formula (see Shiga and Shimizu [SS]).

Furthermore consider the diffusion

$$dz(t) = c(\theta - z(t))dt + \sqrt{2g(z(t))}dw(t) \quad \theta \in \mathbb{R}^+, c \in \mathbb{R}^+. \tag{2.4}$$

The unique equilibrium measure $\Gamma_\theta^{c,g}(\cdot)$ of this process has the property

$$\int z^2 \Gamma_\theta^{c,g}(dz) = \theta^2 + \frac{1}{c}\mathcal{F}_c(g)(\theta). \tag{2.5}$$

This follows by explicit calculation using Ito calculus (see [DG2] for details.)

Finally we present a formula for the second moments of the interaction-chain at level n (compare 0.d) for definitions and [DG3], Equation 3.1. for a proof).

Lemma 2.2 *For every $\theta \in [0,\infty), j \in \{0,1,\dots,n\}, n \in \mathbb{N}$*

$$E((Z^j_{-k})^2 | Z^j_{-j-1} = \theta) = (\sum_{\ell=k}^{j} c_\ell^{-1}) \mathcal{F}^{(j)}(g)_{(\theta)} + \theta^2. \tag{2.6}$$
□

b) Smoothness properties of the mapping $\mathcal{F}_c$, $G_{c,g}$

We start by properties of the mapings $\mathcal{F}_c$, $\mathcal{F}^{(n)}$.

Lemma 2.3

a) For every $c \in (0,\infty)$, $\theta \to \mathcal{F}_c(g)_{(\theta)}$ is continuous on $[0,\infty)$ and Lipschitz continuous at 0.

b) for every $c \in (0,\infty)$ and $n \in \mathbb{N}$:

$$\lim_{\theta\to\infty} (\mathcal{F}_c(g)_{(\theta)}/\theta^2) = 0, \tag{2.7}$$

$$\lim_{\theta\to\infty} (\mathcal{F}^{(n)}(g)_{(\theta)}/\theta^2) = 0. \tag{2.8}$$
□

Proof We prove first b) and then we use some of the estimates of that proof to obtain part a) of the lemma.

Proof of b), (2.7), (2.8) Next observe that (2.8) follows immediately from iterating the argument leading to (2.7). The measure $\Gamma_\theta^{c,g}$ is defined as the marginal of the stationary solution of the SDE

$$dx(t) = c(\theta - x(t))dt + \sqrt{2g(x(t))}dw(t). \tag{2.9}$$

The solution of this equation can be represented as follows.
Let $u_t(x)$ be the solution of the initial value problem $y'(t) = -cy(t)$, $y(0) = x - \theta$. Then we decompose $x(t)$ as follows

$$x(t) = \theta + u_t(x(0)) + \int_o^t u_{t-s}(\sqrt{2g(x(s))})dw(s) = [\theta + u_t(x(0))] + M_t, \tag{2.10}$$

into a deterministic part (given $x(0)$) and a stochastic integral. Next use that for square integrable M_t we have (note that $u_t(\cdot)$ is linear and $Ey(0) = 0$)

$$EM_t^2 = \int_0^t u_{t-s}(Eg(x_s))ds \tag{2.11}$$

so that we can calculate

$$Ex^2(t) = \theta^2 + E(u_t(x(0)))^2 + \int_0^t u_{(t-s)}(Eg(x(s)))ds. \tag{2.12}$$

Define

$$f(t) = \sup_{s\le t} E(x^2(s)) \tag{2.13}$$

and choose $b(D), D$ such that $b(D) < 2c$,

$$g(x) \le Dx + b(D)x^2 \quad \text{and} \quad b(D) \to 0 \quad \text{as} \quad D \to \infty. \tag{2.14}$$

We get with these definitions from (2.12) for all D sufficently large the relation:

$$(2.15)\qquad f(t) \le (1 - \frac{b(D)}{2c})^{-1}((\theta^2 + E(x(0)-\theta)^2 e^{-2ct} + \frac{1}{2c} D\theta).$$

Letting $t \to \infty$ and using (2.5) there exists $A_1 \in (0,\infty)$ such that we get the estimate

$$(2.16)\qquad \mathcal{F}_c(g)_{(\theta)} \le (\frac{4cb(D)}{2c - b(D)})\theta^2 + A_1\theta \qquad \forall\, \theta \in (0,\infty)$$

and hence

$$(2.17)\qquad \overline{\lim_{\theta\to\infty}} \mathcal{F}_c(g)(\theta)/\theta^2 \le \frac{4cb(D)}{2c-b(D)} \underset{D\to\infty}{\longrightarrow} 0,$$

which proves (2.7).

Proof of a) First we prove the continuity of $\theta \to \mathcal{F}_c(g)(\theta)$ for $\theta \in [0,\infty)$. Since $g(x)/x^2$ converges to 0 as $x \to \infty$ and since by combining (2.5) and (2.16):

$$(2.18)\qquad \sup_{\theta \le M} \int_0^\infty x^2 \Gamma_\theta^{c,g}(dx) < \infty,$$

the function g is uniformly integrable under $\Gamma_\theta^{c,g}(\cdot)$ for $\theta \le M$. Hence it suffices to show:

$$(2.19)\qquad \Gamma_{\theta_n}^{c,g} \Longrightarrow \Gamma_\theta^{c,g} \quad \text{if} \quad \theta_n \to \theta \text{ for } n \to \infty.$$

The latter follows from the estimate

$$(2.20)\qquad |\langle \Gamma_\theta^{c,g}, f\rangle - \langle \Gamma_{\theta'}^{c,g}, f\rangle| \le L(f)|\theta - \theta'|$$

for every f with $|f(x) - f(y)| \le L(f)|x-y| \;\; \forall\, x,y \in [0,\infty)$ and with $L(f) < \infty$, since the latter functions are dense in $C_b([0,\infty))$. The estimate (2.20) follows from Lemma 2.4 below.

The relation (2.20) says that application of the kernel $K(x,dy) = \Gamma_x^{c,g}(dy)$ preserves the Lipschitz constant. This implies in particular that application of compositions of the kernels preserve the Lipschitz constant. Then we obtain with (2.7) and (2.5) as in the argument above for $\mathcal{F}(g)$ that also the map $\theta \to \mathcal{F}^{(n)}(g)(\theta)$ is continuous.

Finally we prove that $\mathcal{F}_c(g)_{(\theta)}$ is Lipschitz at $\theta = 0$. Since $g(0) = 0$ it suffices to prove that for all $\theta \in (0,\infty)$:

$$(2.21)\qquad \mathcal{F}_c(g)_{(\theta)} \le C\theta^2 + D\theta \quad \text{for} \quad C, D \in [0,\infty).$$

However (2.21) follows via (2.16).

(c) Contraction properties of the kernels $K_{c,g}$.

Next we formulate the most significant estimates derived from the coupling of dynamics with two different initial points by using the same driving brownian motion. That is we consider the bivariate system $(x_1(t), x_2(t))$

$$(2.22)\qquad dx_i(t) = c(\theta - x_i(t))dt + \sqrt{2g(x_i(t))}dw(t) \quad i = 1,2.$$

(See [DG2] for details.)
Introduce the kernel $K_{c,g}(\cdot,\cdot)$ on $[0,\infty) \times [0,\infty)$, defined by

$$(2.23)\qquad K_{c,g}(\theta, d\rho) = \Gamma_\theta^{c,g}(d\rho).$$

Lemma 2.4 *Let $f : [0,\infty) \to \mathbb{R}$ be a function satisfying*

$$|f(x_1) - f(x_2)| \le C|x_1 - x_2| \quad \forall x_1, x_2 \in [0,\infty). \tag{2.24}$$

Then the function f^* defined by

$$\theta \to \langle \Gamma_\theta^{c,g}, f\rangle \tag{2.25}$$

has the property

$$|f^*(x_2) - f^*(x_2)| \le C|x_2 - x_1| \quad \forall x_1, x_2 \in [0,\infty). \tag{2.26}$$

□

The functions satisfying (2.24) with some $C < \infty$ are called Lipschitz functions.

Corollary *Suppose f is as in Lemma 2.4 and f_k^* is defined by*

$$f_k^* : \theta \to \langle \delta_\theta K_{c_k, \mathcal{F}^{(k)}_{(g)}} \circ \cdots \circ K_{c_0,g}, f\rangle. \tag{2.27}$$

Then

$$|f_k^*(x_2) - f_k^*(x_1)| \le C|(x_2 - x_1)| \quad \forall x_1, x_2 \in [0,\infty). \tag{2.28}$$

□

3 Proof of Theorem 1

In [DG3] (Theorem 1) (in turn based on results of [DG2]) the analogue of the present Theorem 1 has been proved for processes with state space $[0,1]^{\Omega_N}$. That is the state space is *compact* and the components of the system as well as the function g are bounded. These latter facts were used in the proof at several points. Our task here is to show how to replace those arguments based on the boundedness of the state space of the components. It turns out that this can be done using L_2-techniques.

A screening of the proof (in [DG3]) shows that the boundedness of the components was used essentially at two points were we needed *uniform integrability in N* of the components respectively of $\mathcal{F}^{(j)}$ applied to the components of the system $X^{N,j}(s\beta_j(N))$. Here we denote by $(X^{N,j}(t))$ the system obtained by replacing the sequence $(c_k)_{k\in\mathbb{N}}$ by $(c_0, c_1, \ldots, c_j, 0, 0, \ldots)$. Altogether this means (recall that the assumption $g(x)/x^2$ converges to 0 as $x \to \infty$) that in order to get Theorem 1 it suffices to prove the Proposition formulated below.

Proposition 3.1 *The two sequences*

$$\{x_\xi^{N,j}(s\beta_j(N))\}_{N\in\mathbb{N}}, \quad \{\mathcal{F}^{(j)}(g)(x_\xi^{N,j}(s\beta_j(N)))\}_{N\in\mathbb{N}} \tag{3.1}$$

are for every $\xi \in \Omega_\infty$, $s \in [0,\infty)$ and $j \in \mathbb{N}$ uniformly integrable in N. □

Proof Using Lemma 2.3b) we see that both statements of the Proposition are implied by

$$\sup_N [E(x_\xi^{N,j}(s\beta_j(N)))^2] < \infty \quad \forall j \in \mathbb{N}, s \in [0,\infty),\ \xi \in \Omega_\infty. \tag{3.2}$$

To prove this, we turn to Lemma 2.1. First note that since $\sum\limits_\xi a(0,\xi)$ (recall (0.8)) converges to c_0 as $N \to \infty$, we can neglect the fact that the kernel $a(\cdot,\cdot)$ is not normalized and use Lemma 2.1. Note that our kernels are symmetric, which is useful in the subsequent calculation. Integration over the initial distribution μ of the equation for second moments and fixed initial point yields for the kernel $b^{N,j}(\cdot,\cdot)$ on $\{\xi | d(0,\xi) \le j) \times \{\xi | d(0,\xi) \le j\}$ given by the restriction of $a(\cdot,\cdot)$ to these subsets of Ω_N i.e. $b^{N,j}(0,\xi) = a(0,\xi)/\sum\limits_{\|\eta\|\le j} a(0,\eta)$:

$$E(x_\xi^{N,j}(s\beta_j(N)))^2 = \sum_{\eta,\eta'} b^{N,j}_{s\beta_j(N)}(\xi,\eta)\, b^{N,j}_{s\beta_j(N)}(\xi,\eta')\, E(x_\eta^{N,j}(0) x_{\eta'}^{N,j}(0)) \tag{3.3}$$

$$
\begin{aligned}
&+ \int_0^{s\beta_j(N)} \sum_\eta b^{N,j}_{(s\beta_j(N)-u)}(\xi,\eta) b^{N,j}_{(s\beta_j(N)-u)}(\xi,\eta) Eg(x^{N,j}_\eta(u))du \\
&\leq E^\mu(x^{N,j}_\xi(0))^2 \\
&+ \int_0^{s\beta_j(N)} \sum_\eta b^{N,j}_{(s\beta_j(N)-u)}(\xi,\eta) b^{N,j}_{(s\beta_j(N)-u)}(\xi,\eta) E(g(x^{N,j}_\eta(u)))du.
\end{aligned}
$$

¿From our assumptions on g we know that, using (3.3) we can bound as follows (using the spatial homogeneity of μ! and the Kolmogorov equation for $b^{N,j}_{s\beta_j(N)-u}(\xi,\eta)$):

$$
\begin{aligned}
E(x^{N,j}_\xi(s\beta_j(N)))^2 \quad &\leq \quad \text{Const} + D_1 \int_0^{s\beta_j(N)} \hat{b}^{N,j}_{2(s\beta_j(N)-u)}(\xi,\xi)du \\
&+ \quad D_2 \int_0^{s\beta_j(N)} \hat{b}^{N,j}_{2(s\beta_j(N)-u)}(\xi,\xi) \cdot E(x^{N,j}_\xi(u))^2 du.
\end{aligned} \tag{3.4}
$$

Here D_2 can be made as small as we want (on the cost of D_1, compare (2.14)).

The relation (3.4) has (for fixed ξ, N and j) the form of a renewal inequality in the macroskopic time variables for the function $s \to Ex^{N,j}_\xi(s\beta(N))^2$ (see [F]). Choosing D_2 sufficiently small it suffices for (3.2) to hold to show that

$$
\sup_N \Big(\int_0^{s\beta_j(N)} b^{N,j}_u(\xi,\xi)du \Big) < \infty. \tag{3.5}
$$

For verifying (3.5) we need the behaviour of the transition kernel $b^{N,j}_u = \hat{b}^{N,j}_u$ in particular for $u = t\beta_j(N)$ with $t > 0$. We prove in fact the stronger result

$$
\int_0^{s\beta_j(N)} b^{N,j}_u(\xi,\xi)du \underset{N\to\infty}{\longrightarrow} j + 1 - e^{-s}, \tag{3.6}
$$

which follows from the two relations:

$$
\int_0^{K_N} b^{N,j}_u(\xi,\xi)du \underset{N\to\infty}{\longrightarrow} j, \tag{3.7}
$$

$$
\int_{K_N}^{s\beta_j(N)} b^{N,j}_u(\xi,\xi)du \underset{N\to\infty}{\longrightarrow} 1 - e^{-s}, \tag{3.8}
$$

where $K_N/\beta_j(N) \underset{N\to\infty}{\longrightarrow} 0$, $K_N/[\beta_{j-1}(N)\log N] \to \infty$, as $N \to \infty$.

These relations in turn are based on the following facts about the behaviour of random walks on Ω_N as $N \to \infty$:

$$
b^{N,j}_u(\xi,\eta) \underset{u\to\infty}{\longrightarrow} N^{-j} \qquad \forall\, \xi,\eta \in \{\zeta \in \Omega_N \mid \|\zeta\| \leq j\} \tag{3.9}
$$

$$
b^{N,j}_{u\beta_j(N)}(\xi,\xi)N^j \underset{N\to\infty}{\longrightarrow} e^{-u} \quad \text{uniformly in } u \in [a,b] \text{ for } 0 < a < b < \infty. \tag{3.10}
$$

$$(3.11)\qquad \sup_N \Big(\sup_{K_N \le t \le s\beta_j(N)} (b_t^{N,j}(\xi,\xi)N^j)\Big) \downarrow 0 \text{ as } s \to 0$$

$$(3.12)\qquad b_u^{N,j}(\xi,\xi) \underset{N\to\infty}{\longrightarrow} e^{-u}, \text{ for all } u > 0.$$

These relations are proved using characters below. First we finish the proof. The relation (3.10) and (3.11) implies (3.8), with (3.9), (3.10) and (3.12) one proves with induction over j relation (3.7).

The relations (3.9) - (3.12) follow via Fourier transformation as follows:
Let $\hat{\Omega}_N$ denote

$$(3.13)\qquad \{(\lambda_k)_{k\in I\!N} : \lambda_k \in \{0,1,\dots,N-1\}\}$$

and abbreviate

$$(3.14)\qquad <\lambda,\xi>= \exp\Big(\frac{2\pi i}{N}\sum_{k=1}^{\infty}\lambda_k\xi_k\Big).$$

Then we can consider the transform of the distribution $b_t^{N,j}(0,\cdot) : \lambda \to \sum_\xi b_t^{N,j}(0,\xi)\langle\lambda,\xi\rangle$.
Define $K_N = \Big(\sum\limits_{k=1}^{\infty} \frac{c_{k-1}}{N^{k-1}}\cdot\frac{1}{N^k}\Big)^{-1}$ and $r_{N,k} = K_N\frac{c_{k-1}(N-1)}{N^k}\Big(1-K_N\sum\limits_{j+1}^{\infty}\frac{c_{l-1}(N-1)^{-1}}{N^l}\Big)$ for $k \le j$ and $r_{N,k} = 0$ for $k > j$.
Next introduce the value of the transform on $\lambda_1 = 0,\dots,\lambda_{k-1} = 0,\ \lambda_k \ne 0$:

$$(3.15)\qquad f_k^{N,j} = r_{N,1}^j + \dots + r_{N,k}^j - \frac{1}{N-1}r_{N,k+1}^j.$$

Then the transition probabilities can be represented (see Fleischmann and Greven 1994, (2.11)), by the "inverse Fourier transform" as:

$$(3.16)\qquad b_t^{N,j}(0,\xi) = (N-1)\sum_{k>\|\xi\|}^{\infty} N^{-k}e^{-t(1-f_k^{N,j})} + (1\!\mathrm{I}_{\{0\}}(\xi)-1)N^{-\|\xi\|}e^{-t(1-f_{\|\xi\|}^{N,j})}.$$

Relations (3.9) - (3.12) are shown by explicit calculation as follows. Observe that $f_k^{N,j} < 1$ for $k < j$ and $f_k^{N,j} = 1$ for $k \ge j$. This implies (3.9), by letting tend $t \to \infty$ in (3.16). The relations (3.10) respectively (3.12) are implied by (3.16) and the fact that for $k < j$:

$$(3.17)\qquad N^j(1-f^{N,j}) \underset{N\to\infty}{\longrightarrow} 1 \text{ and } N^j(1-f^{N,k}) \underset{N\to\infty}{\longrightarrow} N^{j-k}.$$

The latter is found by inspection.

It remains to verify relation (3.11). Here return to (3.16) and use (3.17) to bound the term $k = j$ in the sum. The terms $k > j$ go to zero since $N \to \infty$. The terms $k < j$ are handled using that $K_N(1-f^{N,k}) \sim K_N N^{-k}$ and hence $N^{j-k}\exp(-K_N N^{-k})$ converges to 0 by the assumption on K_N.

This completes the proof of Proposition 3.1 and hence in connection with [DG3] the proof of Theorem 1.

4 Proof of Theorem 2, Theorem 3 and Theorem 4

We shall prove theorems 2 and 3 together, but we treat the two cases $\Sigma c_k^{-1} < \infty$ and $\Sigma c_k^{-1} = \infty$ separately.

a) Theorem 2 and 3, Case $\Sigma c_k^{-1} < \infty$

In step 1 and 2 we prove Theorem 2 and in step 3 the Theorem 3. The key here are *second moment estimates* in connection with *coupling techniques* (compare (2.6) and (2.28)).

Step 1 (Proof of Theorem 2).

Start by observing that for $m \geq n \geq k$, $\mathcal{L}((Z_j^n)_{-k,\ldots,0}|Z_{-n-1}^n = \theta)$ and $\mathcal{L}((Z_j^m)_{-k,-k+1,\ldots,0}|Z_{-m-1}^m = \theta)$ are two time inhomogeneous Markov chains with transition kernel $K_{-\ell}$ at time $-\ell$ given by

$$K_{-\ell}(\varphi, d\theta) = \Gamma_\varphi^{\ell-1}(d\theta) \quad \ell \in \mathbb{Z}^-, \ell \leq k \tag{4.1}$$

which differ only in their initial distributions. The latter are

$$\delta_\theta K_{-n-1} \circ \cdots \circ K_{-k-1} \quad \text{resp.} \quad \delta_\theta K_{-m-1} \circ \cdots \circ K_{-k-1}, \tag{4.2}$$

where $\circ$ denotes the composition of kernels. Note furthermore that

$$\int_{[0,\infty)} \theta K_{-\ell}(\varphi, d\theta) = \varphi \quad \forall \ell \in \mathbb{N},\ \varphi \in [0,\infty). \tag{4.3}$$

Hence we can rewrite the moment formula (2.6) as

$$\text{Var}\,((Z_{-k-1}^n)|Z_{-n-1}^n = \theta) = (\sum_{k+1}^{n} c_k^{-1})\mathcal{F}^{(n)}(g)_{(\theta)}. \tag{4.4}$$

We shall prove below in step 2:

Lemma 4.1

$$\sum_k c_k^{-1} < \infty \Longrightarrow \sup_n \mathcal{F}^{(n)}(g)(\theta) < \infty. \tag{4.5}$$

$\square$

As a consequence of (4.5) and (4.4) we get

$$\sup_{n \geq k} \text{Var}\,(Z_{-k-1}^n|Z_{-n-1}^n = \theta) \underset{k\to\infty}{\longrightarrow} 0 \tag{4.6}$$

which implies in particular that for every $n = n(k) \geq k$

$$\mathcal{L}(Z_{-k-1}^n|Z_{-n-1}^n = \theta) \underset{k\to\infty}{\Longrightarrow} \delta_\theta.$$

Use this with the Corollary to Lemma 2.4 in Section 2a) and the fact that the Lipschitz functions are dense in $C_b([0,\infty))$ (take $x \to e^{\lambda x}$ with $\lambda \in (-\infty, 0]!$) to conclude that ($m \geq n!$)

$$(\delta_\theta K_{-n-1} \circ \cdots \circ K_{-k-1} - \delta_\theta K_{-m-1} \circ \cdots K_{-k-1}) \underset{n\to\infty}{\Longrightarrow} 0. \tag{4.7}$$

Finally note that (use (2.26) again) the kernels K_{-l} satisfy: if ν_n converges weakly to a measure ν such that $\int x^2 \nu_n(dx)$ is bounded in n (guaranteeing linear functions are uniformly integrable) then $\nu_n K_{-l}$ converges weakly to νK_{-l}. Hence with (4.7) and again with the Corollary to Lemma 2.4 we get then that the process:

$$\mathcal{L}((Z_{-j}^n)_{-k,-k+1,\ldots,0}|Z_{-n-1}^n = \theta) \tag{4.8}$$

converges to a Markov chain $(Z_j^\infty)_{-k,-k+1,\ldots,0}$ with transition kernel $K_{-\ell}$ at time $-\ell$.

Due to (4.3), (4.4), and (4.5) (the latter two implying $\sup_n E((Z_j^n)^2|Z_{-n-1}^n = \theta) < \infty$) we have that the Markov chain

$$(Z_j^\infty)_{=-k,-k+1,\ldots,0} \tag{4.9}$$

is a uniformly integrable martingale with $E(Z_j^\infty) = \theta$, hence by the backward martingale convergence theorem

$$(4.10) \qquad \lim_{j\to\infty} Z_{-j}^\infty = Z_{-\infty}^\infty \quad \text{exists a.s.}$$

and satisfies

$$(4.11) \qquad E(Z_{-\infty}^\infty) = EZ_0^\infty = \theta.$$

The relations (4.8) - (4.11) prove Theorem 2a) for the case $\Sigma c_k^{-1} < \infty$. Note that (4.11) and (4.6) combined imply (via the lemma of Fatou) that $Z_{-\infty}^\infty = \theta$ a.s and hence we obtain an extremal entrance law. Then theorem 2b) is immediate with (2.26), (2.27) via (4.8). It remains to prove Lemma 4.1.

Step 2 (Proof of Lemma 4.1)

The proof will be based on Lemma 2.2 and the following elementary fact:
Let f and h be functions $[0,\infty) \to [0,\infty)$ which are continuous and satisfy for some $a, b \in \mathbb{R}^+$

$$(4.12) \qquad \begin{aligned} & a + bh \geq f \\ & f(x) > 0 \quad \text{for} \quad x > 0 \\ & h(x)/f(x) \underset{x\to\infty}{\longrightarrow} \infty. \end{aligned}$$

Furthermore let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{R}^+$-valued random variables. Then

$$(4.13) \qquad \sup_{n\in\mathbb{N}} [Eh(X_n)/Ef(X_n)] < \infty \qquad \text{implies that}$$

$$\{f(X_n)\}_{n\in\mathbb{N}} \quad \text{is uniformly integrable.}$$

Suppose now we establish

$$(4.14) \qquad \liminf_{n\to\infty} \mathcal{F}^n(g)_{(\theta)} > 0, \quad \text{for } \theta > 0.$$

Then by Lemma 2.2 and $\sum c_k^{-1} < \infty$ we conclude:

$$(4.15) \qquad \sup_n \Big[\frac{E((Z_o^n)^2 | Z_{-n-1}^n = \theta)}{\mathcal{F}_{(g)}^{(n)}(\theta)}\Big] < \infty.$$

Hence with setting $h(x) = x^2$, $f(x) = g(x)$ and $\mathcal{L}(X_n) = \mathcal{L}(Z_0^n | Z_{-n-1}^n = \theta)$ we get from (4.13) that (recall $\mathcal{F}^{(n)}(g)_{(\theta)} = E(g(Z_0^n)|Z_{-n-1}^n = \theta))$):

$$(4.16) \qquad \{g(X_n)\}_{n\in\mathbb{N}} \quad \text{are uniformly integrable}$$

and hence in particular (recall $\mathcal{F}^{(n)}(g)_{(\theta)} = Eg(X_n)$) the following estimate holds:

$$(4.17) \qquad \sup_n \mathcal{F}^{(n)}(g)_{(\theta)} < \infty \quad \theta \in [0,\infty).$$

which is the desired bound for proving Lemma 4.1.

It remains to verify (4.14). If (4.14) does not hold we get from (4.4) that along a subsequence n_k where $\mathcal{F}^{(n)}(g)(\theta)$ converges to 0, we must have due to (2.28) that

$$(4.18) \qquad \mathcal{L}((Z_j^{n_k})_{j=-l-1,-l,\dots,0} | Z_{-n_k-1}^{n_k} = \theta) \Longrightarrow \delta_{(\theta,\dots,\theta)}, \text{ for all } l \in \mathbb{N}.$$

By the Lemma of Fatou this would imply using $(\mathcal{F}_g^n)_{(\theta)} = E(g(Z_0^n)|Z_{-n-1}^n = \theta)$ that

$$(4.19) \qquad \liminf_{k\to\infty} \mathcal{F}^{n_k}(g)(\theta) \geq g(\theta) > 0,$$

which is a contradiction. Hence (4.14) holds and (4.5) is proved.

Step 3 (Theorem 3 in the case where $\sum c_k^{-1} < \infty$.)

Recall that (4.11) says that $E(Z^\infty_{-\infty}) = \theta$.
Since formula (4.6) implies that for $\sum c_k^{-1} < \infty$:

$$\text{(4.20)} \qquad \text{Var}\,(Z^\infty_{-\infty}) = 0$$

we get

$$\text{(4.21)} \qquad Z^\infty_{-\infty} = \theta \quad \text{a.s.}$$

It remains to show that for $\theta > 0$, $Z^\infty_j > 0$ for $j \in \mathbb{Z}^-$. Since we know from the explicit form of $\Gamma^k_\theta(\cdot)$ (see (0.18)) that for every $\ell \in \mathbb{Z}^-$:

$$\text{(4.22)} \qquad K_\ell(\varphi, (0,\infty)) = 1 \quad \text{for} \quad \varphi > 0,$$

it suffices to prove that

$$\text{(4.23)} \qquad \text{Prob}\,(Z^\infty_{-j} = 0) \le b_j \quad b_j \to 0 \quad \text{as} \quad j \to \infty.$$

This latter relation follows however immediately from the inequality (use (4.4)):

$$\text{(4.24)} \qquad \begin{aligned} \text{Prob}\,(|Z^n_{-j} - \theta| \ge \varepsilon | Z^n_{-n-1} = \theta) &\le \frac{1}{\varepsilon^2}\text{Var}\,(Z^n_j | Z^n_{-n-1} = \theta) \\ &\le \frac{1}{\varepsilon^2}(\sum_{k=j}^\infty c_k^{-1})\mathcal{F}^{(n)}(g)_{(\theta)}, \end{aligned}$$

via (4.5) and the assumption $\sum c_k^{-1} < \infty$.
This completes the proof of Theorem 3 in the case $\sum c_k^{-1} < \infty$.

b) Theorem 2 and 3, Case $\sum c_k^{-1} = \infty$

The key to the argument are *consistency considerations* based again on (2.6) and (2.28), which we present now in three steps.
Step 1. We first note that the weak limit points as $j \to \infty$ of $\mathcal{L}((Z^j_k)_{k=-j-1,-j,\dots,0})$ are entrance laws. (Recall that since $E(Z^j_k) = \theta$, this sequence of measures is weakly relativ compact). Namely, suppose we have a transition kernel $K(x, dy)$ such that $x \to \int K(x,dy)f(y)$ is continuous for every $f \in C_b([0,\infty))$. Then $<\nu_n, f> \to <\nu, f>$ for all $f \in C_b([0,\infty))$ as $n \to \infty$ implies that $<\nu_n K, f> \to <\nu K, f>$ for all $f \in C_b([0,\infty))$. This relation is then by iteration extended to $\nu K_{-k} \circ \cdots \circ K_0$ for every $k \in \mathbb{N}$. Since the above continuity assumption on the kernels is satisfied for $K(\theta, d\rho) = \Gamma^{c,g}_\theta(d\rho)$, we know

(4.25) Every weak limit point of $\mathcal{L}((Z^j_k)_{k=-j-1,\dots,0})$ is
an entrance law for $(K_k)_{k\in\mathbb{Z}^-}$.

We denote such an entrance law by $(Z^\infty_k)_{k\in\mathbb{Z}^-}$. Since $K_k(\theta,\cdot)$ has mean θ, $(Z^\infty_k)_{k\in\mathbb{Z}^-}$, is a martingale with values in $[0,\infty]$. Hence in particular we know that the following limit exists almost surely:

$$\text{(4.26)} \qquad Z^\infty_{-\infty} = \lim_{k\to-\infty} Z^\infty_k.$$

We have to prove next in step 2 and 3 in order to obtain Theorem 3:

$$\text{(4.27)} \qquad \sum_0^\infty c_k^{-1} = +\infty \Longrightarrow Z^\infty_{-\infty} \equiv 0,$$

Namely, by writing $Z^\infty_k = E(Z^\infty_0 | Z^\infty_k, Z^\infty_{k-1}, \cdots)$ we see by letting $k \to -\infty$ and using $Z^\infty_k \ge 0$, the relation $Z^\infty_{-\infty} = 0$ implies (by the uniform integrability of Z^∞_k, as a closable, by assumption, martingale) that almost surely $Z^\infty_0 = 0$ and hence almost surely:

$$\text{(4.28)} \qquad Z^\infty_k \equiv 0 \quad k \in \mathbb{Z}^-.$$

Step 2. The strategy now is to exclude the possiblity of $Z^{\infty}_{-\infty}$ taking a positive value, by exhibiting a contradiction to the convergence of Z^{∞}_k as $k \to -\infty$. For this purpose we collect in the present step 2 first a number of facts.

We start with the following first observation:

$$\{Z^{\infty}_{-\infty}\} \cap \{\theta > 0 | \limsup_{n\to\infty} \mathcal{F}^{(n)}(g)(\tilde{\theta}) = 0,\ \forall \tilde{\theta} \in U(\theta)\} = \emptyset, \quad Z^{\infty}_{-\infty} \text{ -a.s.,} \tag{4.29}$$

where U denotes neighbourhood and $\{X\}$ range of X. Namely note first the following three relations:

$$\begin{aligned} &\mathcal{F}^{(k)}(g)(\theta) = E(g(Z^{\infty}_0)|Z^{\infty}_{-k-1} = \theta) \\ &\inf_{0<\delta\le x\le \delta^{-1}} g(x) > 0 \qquad \forall \delta \in (0,1), \\ &EE(Z^{\infty}_0 | Z^{\infty}_{-j-1} = \theta) = \theta. \end{aligned} \tag{4.30}$$

This implies that starting in a θ contained in the second set in (4.29) the processes $(Z^j_k)_{k=-j-1,\dots,0}$ converge to a process with $Z^{\infty}_0 = 0$ and hence $Z^{\infty}_{-k} = 0$ for all $k \in I\!N$ with this starting point θ. Combined with the coupling property of Lemma 2.4 this gives (4.29) by contradiction.

We continue with the second observation. It is clear from explicit form of the transition kernel (as equilibrium of a diffusion with locally bounded drift, see (0.17)) that Z^{∞}_k cannot converge to a value θ for $k \to -\infty$ if $\inf_n \mathcal{F}^{(n)}(g)(\tilde{\theta}) > 0$ in a neighbourhood of θ and $\Sigma c_k^{-1} = +\infty$. Hence almost surely

$$Z^{\infty}_{-\infty} \notin \bigcup_{\delta>0} \{\theta | \liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\tilde{\theta}) > \delta,\ \forall \tilde{\theta} \in U(\theta)\}. \tag{4.31}$$

Step 3. Now we are ready to start the proof. Since the map $\theta \to \Gamma^{c,g}_{\theta}(\cdot)$ is *monotone* (in the stochastic order), recall (0.17) and compare via (2.22) processes starting in $\theta_1 > \theta_2$ and with drift parameters θ_1, θ_2 respectively, we know that $\liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\tilde{\theta}) = 0$ implies $\liminf \mathcal{F}^{(n)}(g)(\theta) = 0$ for all $\theta \le \tilde{\theta}$. Furthermore since g is continuous, strictly positive on $(0,\infty)$ and $E(Z^{(n)}_0 | Z^n_{-n-1} = \theta) = \theta$, we even know by (2.27) that for some $\varepsilon = \varepsilon(\delta) > 0$:

$$\liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\tilde{\theta}) \ge \delta > 0 \Longrightarrow \liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\theta) \ge \delta/2 \quad \forall \theta \in [\tilde{\theta} - \varepsilon, \tilde{\theta} + \varepsilon]. \tag{4.32}$$

Hence we can define

$$a^* = \inf(\theta | \liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\theta) > 0) \tag{4.33}$$

but we know that

$$a^* = \max(\theta | \liminf_{n\to\infty} \mathcal{F}^{(n)}(g)(\theta) = 0)$$

and conclude by (4.32) and (4.31) that

$$Z^{\infty}_{-\infty} \notin (a^*, \infty]. \tag{4.34}$$

On the set $[0, a^*)$ we can conclude that for a sequence $n(k) \uparrow \infty$ with $\mathcal{F}^{(n(k))}(g)(a^*) \to 0$ we have $\mathcal{F}^{(n(k))}(g)(\theta) = 0\ \forall \theta < a^*$, by the monotonicity mentioned above. By the coupling result of Lemma 2.4 we know therefore that in fact even

$$\mathcal{F}^{(n)}(g)(\theta) \underset{n\to\infty}{\longrightarrow} 0 \qquad \forall \theta < a^*. \tag{4.35}$$

Now we use (4.29) to conclude that

$$Z^{\infty}_{-\infty} \notin (0, a^*). \tag{4.36}$$

Finally we exclude $Z_{-\infty}^{\infty} = a^* > 0$ with positive probability by the coupling Lemma 2.4 which says that an entrance law with $Z_{-\infty}^{\infty} = a^*$ would have to be arbitrarily close to the trivial process, which is associated with every $\theta \in (0,\infty)\backslash\{a^*\}$.
Altogether with (4.35), (4.36) we have proved (4.28).

Proof of Theorem 4
(a) The proof of (i) (*local extinction*) can be obtained via Theorem 2 in [CFG]. The proof of (ii) (*persistence*) can be obtained by the same argument as in Cox and Greven (1994) or Proposition 6.1 in Dawson and Perkins (1991).
(b) An elementary calculation involving the application of Ito's lemma to$x_\xi(t)x_{\xi'}(t)$, and taking the limit $t \to \infty$ yields that under the law $\tilde{\nu}_\theta^N$

$$(4.37) \qquad Cov(x_\xi, x_{\xi'}) \leq \frac{2\theta H(1, \xi - \xi')}{c_0}$$

where $H(1,\xi)$ denotes the expected number of visits to zero by the discrete time hierarchical random walk. However from A4 we have that

$$H(1, \xi - \xi') \underset{d(\xi,\xi')\to\infty}{\longrightarrow} 0$$

uniformly in N. The first statement in (b) then follows as in the proof of Theorem 0.10(a) in [DGV]. The second statement of (b) follows from Theorem 11 part(c) which is given below. It can also be proved in the same way as in the proof of (0.71) in [DGV] but using (4.37).

5 Proof of Theorems 5 and 6

A) The diffusive clustering case, Proof of Theorem 5

The proof starts with the case $g(x) = dx$ and will make use of techniques and results obtained by J. Lamperti and P. Ney [1968] to study the behaviour of a critical Galton-Watson process conditioned on surviving till time n. By developing comparison methods we are able to study our time *inhomogeneous* Markov chains $(Z_k^j)_{k=j-1,\dots,0}$ describing $\mathbb{R}^+$-*valued masses* instead of particles, based on knowledge about time-homogenous branching particle systems. Using comparison methods we generalize finally to the case $g \in \mathcal{G}((d))$.

In section 5a) we formulate a key statement about systems of time inhomogeneous branching masses, which is proved in the lengthy section 5b), next section 5c) proves Theorem 5 in the case $g(x) = dx$ and section 5d) for $g \in \mathcal{G}((d))$.

a) Preparations

First we need as a major tool a result on the behaviour of a system of branching masses, namely the analog to the asymptotics of the extinction probability for $\mathbb{N}$-valued critical branching processes.

Suppose that $\{K_d(\theta, d\varphi),\ d \in (0,\infty)\}$ is a one parameter set of positive transition kernels on $[0,\infty) \times [0,\infty)$ satisfying for every $d \in (0,\infty)$ the following assumptions:

$$(5.1) \qquad K_d(\theta, [0,\infty)) = 1 \text{ and } K(0,\{0\}) = 1$$

$$(5.2) \qquad K_d(\theta_1 + \theta_2, d\varphi) = (K_d(\theta_1,\cdot) * K_d(\theta_2,\cdot))(d\varphi) \quad \text{for } \theta_1, \theta_2 \in [0,\infty)$$
$$\text{with } * \text{ denoting convolution of distributions}$$

$$(5.3) \qquad \int K_d(\theta, d\varphi)\varphi = \theta \qquad \text{for } \theta \in [0,\infty)$$

$$(5.4)\qquad \int K_d(\theta, d\varphi)\varphi^2 - \theta^2 = \frac{1}{d}\theta,$$

$$(5.5)\qquad \sup_{d\geq d_*} \int K_d(\theta, d\varphi)\varphi^3 = O(d_*^{-2}) \quad \text{as} \quad d_* \to 0,$$

$$(5.6)\qquad \sup_{d\geq d_*} \int K_d(\theta, d\varphi)\varphi^4 = O(d_*^{-3}) \quad \text{as} \quad d_* \to 0.$$

Consider now a sequence $(d_k)_{k\in\mathbb{N}}$ of real numbers satisfying either of the two conditions

$$(5.7)\qquad 0 < (d_k)^{-1} \leq d^* < \infty \quad \forall k \in \mathbb{N} \text{ and } \sum_{k=0}^{\infty} d_k^{-1} = +\infty$$

or:

$$d_j \sum_{k=0}^{j} d_k^{-1} \underset{j\to 0}{\longrightarrow} \infty \text{ and } \sum_{k=0}^{\infty} d_k^{-1} = +\infty$$

Now we have all ingredients to formulate a result on the extinction probability of branching masses.

Proposition 5.1 *Assume that $\{K_d\}_{d\in\mathbb{R}^+}$ satisfies (5.1)-(5.6) and the sequence (d_k) relation (5.7). Then for every $\varepsilon > 0$ and every $\theta > 0$*

$$(5.8)\qquad (\delta_\theta)K_{d_n} \circ K_{d_{n-1}} \circ \ldots \circ K_{d_o}([\varepsilon, \infty)) \underset{n\to\infty}{\sim} \frac{1}{2}\theta\Big(\sum_{k=0}^{n} \frac{1}{d_k}\Big)^{-1}. \qquad \square$$

Corollary 5.1 *For the backward composition we have for $\varepsilon > 0$:*

$$(5.9)\qquad (\delta_\theta)K_{d_0} \circ \cdots \circ K_{d_n}([\varepsilon, \infty)) \underset{n\to\infty}{\sim} \frac{1}{2}\theta\Big(\sum_{k=0}^{n} \frac{1}{d_k}\Big)^{-1}. \qquad \square$$

b) Decay of the nonextinction probability, Proof of Proposition 5.1.

Now we come to the proof of the crucial relation (5.8). The problem is that for a realization of Z_k^j of the chain starting in θ at times $-j-1$ and transitions according to K_{d_k} at time $-k$ we know that $\mathcal{L}(Z_k^j | Z_{-j-1}^1 = \theta)$ may be (and is in the case of interest to us) supported by $(0, \infty)$ for all k provided $\theta > 0$ and hence the point 0 is not reached. On the other hand the probability of being 0 is the quantity readily read of from the Laplace transform. The idea of the proof is now to construct a new sequence of transition kernels satisfying on the one hand $K(\theta, \{0\} \cup [\varepsilon, \infty)) = 1$ but is such that $\{0\}$ can be hit with positive probability and such that for $\varepsilon \to 0$ we get good approximations of the original situation. Here however we have to consider ε of the form $\varepsilon b(j,k)$ for the chain indexed j at time $-k$, in order to achieve this approximation.

Step 1 Construction of the new chain:
Denote by (Z_k^j) the Markov chain with transition kernels K_{d_k} at time k which starts in θ at time $-j-1$.

On the *same* probability space we shall construct $(Z_k^{j,\varepsilon})_{k=-j-1,\ldots,0}$ such that $Z_{-j-1}^{j,\varepsilon} = \theta$ and such that the following conditions are satisfied

$$(5.10)\qquad \begin{aligned} &\text{(i)} \quad \text{Prob}(Z_0^{j,\varepsilon} \neq 0) \sim \tfrac{1}{2}\theta\Big(\sum_{k=0}^{j} d_k^{-1}\Big)^{-1}(1 + O(\varepsilon)) \quad \text{as } j \to \infty \text{ and } \varepsilon \to 0 \\ &\text{(ii)} \quad \text{Prob}\Big(\Big\{\Big[\sup_{|k|\leq j} |Z_k^j - Z_k^{j,\varepsilon}|\Big] > \delta\Big) = o\Big(\sum_{k=0}^{j} d_k^{-1}\Big)^{-1} \quad \text{as } j \to \infty,\ \delta > 0 \\ &\text{(iii)} \quad \text{Prob}\Big(Z_0^{j,\varepsilon} \in (0,\delta) | Z_0^{j,\varepsilon} > 0\Big) \underset{j\to\infty}{\longrightarrow} 0 \text{ uniformly in } \varepsilon \quad \forall \delta > 0 \end{aligned}$$

Then we can finish the proof by observing that as $j \to \infty$ for all $\varepsilon > 0$:

$$
\begin{aligned}
(5.11) \qquad \mathrm{Prob}(Z_0^j \geq \delta) &= (1+O(\varepsilon))\mathrm{Prob}(Z_0^{j,\varepsilon} \neq 0) + o\Big(\sum_{k=0}^{j} d_k^{-1}\Big)^{-1} \\
&\underset{j\to\infty}{\sim} \frac{1}{2}\theta\Big(\sum_{k=0}^{j} d_k^{-1}\Big)^{-1}(1+O(\varepsilon)),
\end{aligned}
$$

which is the desired asymptotic relation.

The next step is therefore to construct the pair $((Z_k^j)_{k=-j-1,\dots,0}, Z^{j,\varepsilon}_{k=-j-1,j,\dots,0})$ satisfying (5.10). The key to that construction is the following fact. Due to the "branching property" (5.2) the measures $K_d(\theta,\cdot)$ are infinitely divisible (and supported by $[0,\infty)$!), hence, we can represent the Laplace transforms of the measures $K_d(\theta,\cdot)$ as follows:

$$(5.12) \qquad \int_0^\infty K_d(\theta, dy)e^{-\lambda y} = \exp(-\theta\psi_d(\lambda)), \quad \psi_d(\cdot) \geq 0.$$

Consequently by using this relation iteratively we get:

$$(5.13) \qquad \int_0^\infty (\delta_\theta K_{d_j} \circ \cdots \circ K_{d_0})(dy)e^{-\lambda y} = \exp(-\theta\psi_{d_j}(\psi_{d_{j-1}}(\cdots(\psi_{d_0}(\lambda))\cdots)).$$

Using again that the measures $K_d(\theta,\cdot)$ resp. $\delta_\Theta K_{d_j} \circ \cdots \circ K_{d_1}$ are *infinitely divisible* their Laplace transforms or better the functons ψ_d can be represented via the socalled canonical or *Levy measure*. Namely there exists a measure R_d on $[0,\infty)$ such that (recall that our random variable is nonnegative and hence there is no normal component)

$$(5.14) \qquad \psi_d(\lambda) = \int_0^\infty (1-e^{-\lambda u})R_d(du).$$

Having the relations (5.14) and (5.12), (5.13) in mind we shall specify the law of $Z_k^{j,\varepsilon}$ by transition kernels which will be defined via Levy-measures R_d^ε obtained from R_d. This way we preserve automatically the branching property (5.2) and the property (5.1) for the new transition kernel.

Define first for a fixed $\varepsilon > 0$ and a $d = d(\varepsilon)$ the measure R_d^ε on $[0,\infty)$ by

$$(5.15) \qquad R_d^\varepsilon : R_d^\varepsilon(A) = \alpha R_d(A \cap [\varepsilon,\infty))$$

where we choose $\alpha = \alpha(\varepsilon)$ such that

$$(5.16) \qquad \int_0^\infty uR_d^\varepsilon(du) = 1, \quad \alpha(\varepsilon) \downarrow 1 \text{ as } \varepsilon \downarrow 0.$$

Next let $\tilde{K}_d^\varepsilon$ be the kernel such that $\tilde{K}_d^\varepsilon(\theta,\cdot)$ has Laplace transform $\exp(-\theta\psi_d^\varepsilon(\lambda))$ with

$$(5.17) \qquad \psi_d^\varepsilon(\lambda) = \int_0^\infty (1-e^{-\lambda u})R_d^\varepsilon(du).$$

We can give a particular representation for that kernel as follows: Observe that a (generalised) Poisson process on $[0,\infty)$ with intensity measure $R(\cdot)$ has the property that a realization, written as $\sum_{i=0}^{\infty}\delta_{m_i}$ generates a random variable $Z = \sum_{i=0}^{\infty} m_i$ with Laplace-transform $\exp(-\int_0^\infty(1-e^{-\lambda u})R(du))$. This proves in particular that $\exp(-\theta\psi_d^\varepsilon(\lambda))$ is Laplace transform of a probability distribution. We define $\tilde{d}$ as the following function of $d : d \to (\int \tilde{K}_d^\varepsilon(1,d\rho)\rho^2 - 1)^{-1}$. Furthermore from the representation (5.17) it is easily checked that we have for fixed ε that:

$$(5.18) \qquad \{\tilde{K}_{\tilde{d}}^\varepsilon | d \in (0,\infty)\}$$

satisfies (5.1) - (5.6). Namely (5.2) follows since (5.17) implies that the law is infinitely divisible in Θ. Furthermore (5.15) and (5.16) give together (5.1) and (5.3). The relation (5.4) holds by the construction of the function $\tilde{d}$. Finally the representation of the measures via Poisson point processes together with the fact that R_d and R_d^ε agree on $[0,\infty)$ gives (5.4) and (5.5).

This makes now the construction of Z_k^j, $Z_k^{j,\varepsilon}$ on one probability space work as follows:

We shall construct two point processes on $[0,\infty)$. We need the following ingredients:

- $Y_{1,d}^{\varepsilon,\rho}$ is a Poisson point process with intensity measure $\rho(R_d|_{[\varepsilon,\infty)})$. (Here $|$ denotes restriction)
- $Y_{2,d}^{\varepsilon,\rho}$ is a Poisson point process with intensity measure $\rho(R_d|_{[0,\varepsilon)})$.
- $Y_{3,d}^{\varepsilon,\rho}$ is the Poisson point process with intensity measure $(\alpha-1)\rho(R_d|_{[\varepsilon,\infty)})$.

All the above three processes are independent. Define $(\tilde{Y}_d^{\varepsilon,\rho}, \approx{Y}_d^{\varepsilon,\rho})$ as follows

$$\tilde{Y}_d^{\varepsilon,\rho} = Y_{1,d}^{\varepsilon,\rho} \cup Y_2^{\varepsilon,\rho} \tag{5.19}$$

$$\overset{\approx}{Y}_d^{\varepsilon,\rho} = Y_{1,d}^{\varepsilon,\rho} \cup Y_3^{\varepsilon,\rho}. \tag{5.20}$$

Now we realize an independent collection of these random variables for $d = d_1, d_2, \ldots$ and for fixed j with ε of the form $\varepsilon b(j,k)$ for a given array b and for a particular value of k. In the notation we will however write ε rather than $\varepsilon b(j,k)$. For a simple point process Y represented as $Y = \sum_i \delta_{m_i}$ we write simply $I(Y) = \sum_i m_i$. We set

$$\begin{aligned} & Z_{-j-1}^j = Z_{-j-1}^{j,\varepsilon} = \Theta \\ & Z_{-j}^j = I(\tilde{Y}_{d_j}^{\varepsilon,\Theta}) \qquad Z_{-j}^{j,\varepsilon} = I(\overset{\approx}{Y}_{d_j}^{\varepsilon,\Theta}) \\ & Z_{k+1}^j = I(\tilde{Y}_{d_{k+1}}^{\varepsilon,Z_k^j}) \qquad Z_{k+1}^{j,\varepsilon} = I(\overset{\approx}{Y}_{d_{k+1}}^{\varepsilon,Z_k^{j,\varepsilon}}). \end{aligned} \tag{5.21}$$

We are now left with verifying (5.10) (i),(ii) and (iii), which is the content of steps 2,3, respectively step 4 and 5.

Step 2 (Nonextinction probability of the new chain, Proof of (5.10(i)).) In this section we fix an array $\varepsilon = \varepsilon b(j,k)$ as in step 1 and work with a given sequence $(d_k)_{k\in\mathbb{N}}$ satisfying (5.7). We assume $b(j,k) \le 1$ throughout this section. This means we have by the construction of the previous step a Markov chain $(Z_k^{j,\varepsilon})_{\{k\in -j-1,-j,\ldots,0\}}$ given.

The key to the proof is the analysis of the Laplace transform of the terminal distributions of the approximating Markov chains:

$$L_j^\varepsilon(\lambda) = \int_0^\infty e^{-\lambda u} P(Z_0^{j,\varepsilon} \in du), \quad \lambda \in [0,\infty). \tag{5.22}$$

It turns out to be useful to extend the domain of this function to complex values as well. For this purpose define

$$f_\theta^{d,\varepsilon}(z) = \int_0^\infty z^u \tilde{K}_d^\varepsilon(\theta, du), \; z \in \mathbb{C}, |z| \le 1.$$

so that $L_j^\varepsilon(\lambda) = f_\theta^{d,\varepsilon}(e^{-\lambda})$.

In order to analyse the expression (5.22) we need the logarithm of the transforms of the corresponding transition kernels denoted $\tilde{K}_{d_k}^\varepsilon$ at time $-k-1$:

$$\psi_\theta^{d,\varepsilon}(\lambda) = -\frac{1}{\theta} \log f_\theta^{d,\varepsilon}(e^{-\lambda}). \tag{5.23}$$

The function $L_j^\varepsilon(\cdot)$ can be obtained, due to the branching property (5.2) of the kernels $\{\tilde{K}_d^\varepsilon,\, d \in [0,\infty)\}$, as follows:

$$L_j^\varepsilon(\lambda) = \exp\,(-\theta\psi_1^{d_j,\varepsilon} \circ \ldots \circ \psi_1^{d_0,\varepsilon}(\lambda)). \tag{5.24}$$

Hence we have the formula (recall $Z_0^{j,\varepsilon} \to 0$ in law as $j \to \infty$)

$$\begin{aligned} &\text{Prob }(Z_0^{j,\varepsilon} > 0) = 1 - L_j^\varepsilon(\infty) \sim \theta G_\varepsilon^{(j)}(\infty) \text{ as } j \to \infty \\ &\qquad\qquad \text{with } G_\varepsilon^{(j)}(\lambda) = \psi_1^{d_j,\varepsilon} \circ \ldots \circ \psi_1^{d_0,\varepsilon}(\lambda). \end{aligned} \tag{5.25}$$

Therefore the remaining task is now to analyse the behaviour of $G_\varepsilon^{(j)}(\infty)$ as $j \to \infty$ uniformly in ε.

We define $d_k(\varepsilon)$ as the reciprocal variance of $\tilde{K}_{d_k}^\varepsilon(1,\cdot)$ and define $K_{d_k(\varepsilon)}^\varepsilon = \tilde{K}_{d_k}^\varepsilon$. More generally (cf. (5.18)) we can reparametrize in such a way that we get kernels $K_d^\varepsilon = K_{\tilde{d}}^\varepsilon$ such that d is the reciprocal variance of $K_d^\varepsilon(1,\cdot)$. Since K_d^ε satisfies (5.2) (which implies we have a Levy-Khinchine representation) it is easy to verify that K_d^ε satisfies (5.5) and (5.6).To now complete the agrument the main tools are the following three lemmata:

Lemma 5.2 *Assume that K_d^ε and the sequence $(d_k(\varepsilon))$ satisfies (5.1) - (5.7). Then for $\lambda \in (0,\infty)$ uniformly in ε:*

$$\frac{1}{G^{(n)}(\lambda)} = \frac{1}{2}\sum_{i=0}^n d_i^{-1}(\varepsilon) + o(\sum_{i=0}^n d_i^{-1}(\varepsilon)) \text{ as } n \to \infty. \tag{5.26}$$ $\square$

Lemma 5.3

$$E(\exp(-\lambda Z_0^{j,\varepsilon} | Z_0^{j,\varepsilon} > 0)) \underset{j\to\infty}{\longrightarrow} 0 \qquad \textit{uniformly in } \varepsilon. \tag{5.27}$$ $\square$

Lemma 5.4

$$\sum_0^n d_k^{-1} / \sum_0^n d_k^{-1}(\varepsilon) = (1 + O(\varepsilon)) \text{ uniformly in } n. \square$$

The relation (5.10) (i) is now obtained as follows. For $(d_k)_{k\in\mathbb{N}}$ and $\varepsilon b(j,k)$ we construct the chain $(Z_k^{j,\varepsilon})_{k=-j-1,\ldots,o}$ as in step 1 and note that $d_k(\varepsilon)$ is such that the conditional variance $\mathrm{Var}(Z_{k+1}^{j,\varepsilon}\ |Z_k^{j,\varepsilon} = \theta)$ in step $-k$ is given by $\theta/d_k(\varepsilon)$.

Note that with the abbreviation $q_n^\varepsilon = \text{Prob }(Z_0^{j,\varepsilon} > 0)$ we can write

$$L_j^\varepsilon(\lambda) = (1 - q_j^\varepsilon) + q_j^\varepsilon E(\exp\,(-\lambda Z_0^{j,\varepsilon})|Z_0^{j,\varepsilon} > 0). \tag{5.28}$$

Hence we get

$$1 - L_j^\varepsilon(\lambda) = q_j^\varepsilon(1 - E(\exp\,(-\lambda Z_0^{j,\varepsilon}|Z_0^{j,\varepsilon} > 0)). \tag{5.29}$$

According to Lemma 5.3 we know therefore

$$(1 - L_j^\varepsilon(\lambda))/(1 - L_j^\varepsilon(\infty)) \underset{j\to\infty}{\longrightarrow} 1. \tag{5.30}$$

Combining the relations (5.25) and (5.26) gives the desired relation (recall $\sum_0^n d_i^{-1}(\varepsilon)$ diverges as $n \to \infty$)

$$\text{Prob }(Z_0^{j,\varepsilon} > 0) \underset{n\to\infty}{\sim} \frac{\theta}{2}(\sum_{i=0}^n d_i^{-1}(\varepsilon))^{-1}, \tag{5.31}$$

uniformly in ε. By applying finally Lemma 5.4 we get (5.10)(i).

Step 3 Proof of Lemma 5.2 and Lemma 5.4:

The proof consists of two parts (I) and (II). First we show (5.26) for fixed $\lambda \in [0,\infty)$ and $\varepsilon > 0$. The second part consists in showing the uniformity in ε and Lemma 5.4.

(I) Since we keep θ and $\varepsilon > 0$ fixed here we repress it in the notation and write furthermore instead of $\psi_1^{d_j(\varepsilon),\varepsilon}$ simply ψ_j and instead of $d_j(\varepsilon)$ simply d_j. In this notation we get the recursion relation (see 5.24 and 5.25):

$$G^{(j)}(\lambda) = \psi_j(G^{(j-1)}(\lambda)) = \psi_j \circ \psi_{j-1}(G^{(j-2)}(\lambda)) = \ldots = \psi_j \circ \ldots \circ \psi_0(\lambda). \tag{5.32}$$

Since we know by assumption (5.6) that $\int u^4 \tilde{K}_d^\varepsilon(1,du) < \infty$ we know that $\int u^4 K_d^\varepsilon(1,du) < \infty$ and we can conclude that the function $\psi_n(\lambda)$ is four times differentiable and get (recall $\psi_n(0) = 0$)

$$\psi_n(\lambda) = \psi_n^{(1)}(0)\lambda + (\frac{\psi_n^{(2)}(0)}{2})\lambda^2 + (\frac{\psi_n^{(3)}(0)}{6})\lambda^3 + y_n(\lambda), \;\; y_n(\lambda) = O(\lambda^4). \tag{5.33}$$

We now make the Ansatz:

$$\frac{1}{\psi_n(\lambda)} = A_n \cdot \frac{1}{\lambda} + B_n + C_n\lambda + D_n\lambda^2 + E_n\lambda^3 + \delta_n(\lambda) \text{ with } \delta_n(\lambda) = O(\lambda^3) \tag{5.34}$$

and abbreviate

$$a_n = \psi_n^{(1)}(0), \; b_n = \frac{\psi_n^{(2)}(0)}{2}, \; c_n = \frac{\psi_n^{(3)}(0)}{6}. \tag{5.35}$$

We find by inserting (5.33) in (5.34) through comparison of coefficients that

$$\begin{aligned} A_n &= \frac{1}{a_n}, \quad B_n = -\frac{b_n}{a_n^2}, \quad C_n = -\frac{c_n}{a_n^2} + \frac{b_n^2}{a_n^3} \\ D_n &= 2\frac{b_n c_n}{a_n^3} - \frac{b_n^3}{a_n^4} - \frac{1}{a_n^2}\frac{y_n(\lambda)}{\lambda^4}, \quad E_n = 0 \end{aligned} \tag{5.36}$$

$$\begin{aligned} \delta_n(\lambda)\psi_n(\lambda) &= \left(-\frac{b_n}{a_n^2}\frac{y(\lambda)}{\lambda^4} + c_n C_n + b_n D_n\right)\lambda^4 \\ &+ \left(c_n D_n + C_n \frac{y_n(z)}{\lambda^4}\right)\lambda^5 \\ &+ \left(\frac{y_n(z)}{\lambda^4} D_n\right)\lambda^6 \end{aligned}$$

In order to relate the quantities appearing in (5.35) with the moments of the given laws $K_d^\varepsilon(1,\cdot)$ we derive some relations useful later on. Suppose that the random variable X has a Laplace transform

$$F(\lambda) = Ee^{-\lambda X} \tag{5.37}$$

which can be written as $F(\lambda) = \exp(-\psi(\lambda))$. Then

$$\left[(\frac{d}{d\lambda})^k F\right](0) = (-1)^k EX^k \tag{5.38}$$

and

$$\psi^{(1)}(0) = EX, \; \psi^{(2)}(0) = -\text{Var }(X), \; \psi^{(3)}(0) = EX^3 - (EX)^3 + 3 \text{ Var }(X)EX \tag{5.39}$$

$$\begin{aligned} \psi^{(4)}(\lambda) &= -e^{\psi(\lambda)}F^{(4)}(\lambda) + 4\psi^{(3)}(\lambda)\psi^{(1)}(\lambda) \\ &+ 3(\psi^{(2)}(\lambda))^2 - 3(\psi^{(1)}(\lambda))^2 + (\psi^{(2)}(\lambda) + (\psi^{(1)}(\lambda))^4. \end{aligned} \tag{5.40}$$

Observe next (recall $L_j(\lambda) \to 1$ as $j \to \infty$) $G^{(n)}(\lambda) \to 0$ as $n \to \infty$. Using the relations (5.32) and (5.34) in connection with (5.36) we are able to expand the function $(G^{(n)}(\lambda))^{-1}$ and we find:

$$
\begin{aligned}
(5.41) \qquad \frac{1}{G_n(\lambda)} &= \frac{1}{\psi_n(G^{(n-1)}(\lambda))} \\
&= \frac{1}{a_n}\cdot\frac{1}{G^{(n-1)}(\lambda)} - \frac{b_n}{a_n^2} + C_n G^{(n-1)}(\lambda) + D_n (G^{(n-1)}(\lambda))^2 + \delta_n(G^{(n-1)}(\lambda)).
\end{aligned}
$$

We know that (recall (5.1) - (5.4))

$$
(5.42) \qquad \psi_n(0) = 0,\ \psi_n^{(1)}(0) = 1,\ \psi_n^{(2)}(0) = -\frac{1}{d_n}.
$$

Therefore we obtain by iterating (5.41) and substituting a_n, b_n via (5.35) the expression:

$$
\begin{aligned}
(5.43) \qquad \frac{1}{G^{(n)}(\lambda)} &= \frac{1}{\psi_0(\lambda)} + \sum_{i=0}^{n} \frac{1}{d_i} + \sum_{i=0}^{n} C_i G^{(i-1)}(\lambda) \\
&+ \sum_{i=0}^{n} D_i (G^{(i-1)}(\lambda))^2 + \sum_{i=0}^{n} \delta_i (G^{(i-1)}(\lambda)).
\end{aligned}
$$

We have to show now that the expression

$$
(5.44) \qquad R_n(\lambda) = \sum_{i=0}^{n} C_i G^{(i-1)}(\lambda) + \sum_{i=0}^{n} D_i (G^{(i-1)}(\lambda))^2 + \sum_{i=0}^{n} \delta_i (G^{(i-1)}(\lambda))
$$

satisfies $R_n(\lambda) = o(\sum_{i=1}^n d_i^{-1})$ as $n \to \infty$, for every fixed $\lambda \in [0,\infty)$.

At this point we see that we now need information on c_n i.e. $\psi_n^{(3)}(0)$. For this purpose we note first that combining (5.35), (5.39) and Jensen's inequality that

$$
(5.45) \qquad c_n > 0.
$$

Therefore we know from (5.36) in connection with (5.35) and (5.42) that:

$$
(5.46) \qquad C_n \le (\frac{1}{d_n})^2, \quad C_n \ge -c_n
$$

$$
\begin{aligned}
D_n(\lambda) &\le (\frac{1}{d_n})^3 - \frac{y_n(\lambda)}{\lambda^4}, \\
D_n(\lambda) &\ge -\frac{y_n(\lambda)}{\lambda^4} - |b_n c_n| \ge -\frac{y_n(\lambda)}{\lambda^4} - \frac{\text{const}}{d_n^3}
\end{aligned}
$$

$$
\begin{aligned}
\delta_n(z) &= (\psi_n(\lambda))^{-1}\Big(\left[-3c_n(\frac{1}{d_n})^2 + c_n^2 + (\frac{1}{d_n})^4 + \frac{-2y_n(\lambda)}{d_n\lambda^4} \right] \lambda^4 \\
&+ \left[-c_n(\frac{1}{d_n})^3 + (\frac{2c_n^2}{d_n}) \ + 2c_n\frac{y_n(\lambda)}{\lambda^4} + \frac{-y_n(\lambda)}{{d_n}^2\lambda^4} \right] \lambda^5 \\
&+ \frac{y_n(\lambda)}{\lambda^4} \left[(-\frac{1}{d_n})^3 + \frac{2c_n}{d_n} + \frac{y_n(\lambda)}{\lambda^4} \right] \lambda^6 \Big).
\end{aligned}
$$

In order to exploit the (5.46) for general sequences $(d_n)_{n\in\mathbb{N}}$ we need to use information on c_n and $y_n(\lambda)$ that is on $\psi_{(0)}^{(3)}$ and $\psi_n^{(4)}(\lambda)$.

With the assumptions (5.5) and (5.6) (and the fact that those relations according to (5.18) remain true for the approximating kernels) we obtain the following:

$$
(5.47) \qquad c_n \le K(d_n^{-2} \vee 1) + K,\ K \ge 1
$$

$$|\frac{y_n(\lambda)}{\lambda^4}| \leq K(d_n^{-3} \vee 1) + K'.$$

In order to estimate $R_n(\lambda)$ we discuss separately the two cases

$$(5.48) \qquad \begin{aligned} &1.\ 0 < d_k^{-1} \leq d^* < \infty \quad \text{and} \quad \sum d_k^{-1} = \infty \\ &2.\ \ d_k \underset{k\to\infty}{\longrightarrow} 0 \quad \text{and} \quad d_k \cdot \sum_{i=0}^{k} d_i^{-1} \underset{k\to\infty}{\longrightarrow} \infty. \end{aligned}$$

Case 1 We use that $G^{(i)}(\lambda)$ converges to 0 as $i \to \infty$ in connection with the bounds on $C_n, D_n, \delta_n(\lambda)$ obtained by combining (5.44) - (5.46), (5.36) and bounding $c_n, y_n(\lambda)$ by (5.47).

Case 2 Since now the terms d_k^{-1} diverge we have to use the speed at which $G^{(k)}(\lambda)$ converges to zero to get out the desired estimate.

Suppose that along a subsequence $G^{(j)}(\lambda)(\sum_{k=0}^{j} d_k^{-1})$ diverges. Then $(G^{(j)}(\lambda))^{-1}$ is asymptotically smaller than $\sum_{k=0}^{j} d_k^{-1}$, at least along a subsequence, This would imply by (5.43) that $R_n(\lambda) \sim -\sum_{i=0}^{n} d_i^{-1}$ along that subsequence, which contradicts the bound on $R_n(\lambda)$ obtained by combining (5.44) - (5.46) and making use of (5.47). Hence we can conclude

$$(5.49) \qquad G^{(j)}(\lambda) \leq C\left(\sum_{k=0}^{j} d_k^{-1}\right)^{-1} \qquad \forall j \in I\!N.$$

Then the desired relation $R_n(\lambda) = o(\sum_{k=0}^{n} d_k^{-1})$ follows as above from (5.44) - (5.46), (5.36) and (5.47) by the fact that for case 2: $d_j \cdot \sum_{k=0}^{j} d_k^{-1} \to \infty$ as $j \to \infty$.

(II) We shall now display again ε in the notation for $d_k(\varepsilon)$, since both d_k and $d_k(\varepsilon)$ will appear. In order to prove the uniformity in ε and to prove Lemma 5.4 it suffices to show that for $k \in I\!N$: $d_k(\varepsilon)/d_k$, $(\psi_k^{\varepsilon,(i)}(\lambda)/\psi_k^{(i)}(\lambda)$ for $i = 3, 4$ converges as $\varepsilon \to 0$ to 1 uniformly in k.

Note that (recall (5.14) - (5.18))

$$\begin{aligned} \psi_k^{(j)}(\lambda) &= (-1)^j \int_0^\infty u^j e^{-\lambda u} R_{d_k}(du) \\ (\psi_k^\varepsilon)^{(j)}(\lambda) &= (1-\alpha(\varepsilon))(-1)^j \int_\varepsilon^\infty u^j e^{-\lambda u} R_{d_k}(du). \end{aligned}$$

Hence we need (as a sufficient condition) in order to get the assertion for $d_k(\varepsilon)/d_k$:

$$\int_0^\varepsilon u^2 R_{d_k}(du) / \int_0^\infty u^2 R_{d_k}(du) \underset{\varepsilon\to 0}{\longrightarrow} 0 \quad \text{uniformly in } k.$$

This suffices at the same time for the remaining points. This can always be achieved by choosing $\varepsilon = \varepsilon(j,k)$ sufficiently small as a function of j uniformly in k.

Step 4 (Proof of Lemma 5.3 and (5.10)(iii))

We know that with $a_n^\varepsilon = \frac{1}{2}\sum_{i=0}^{n} d_i^{-1}(\varepsilon)$ (recall (5.28)),

$$(5.50) \qquad a_j^\varepsilon(1 - L_j^\varepsilon(\lambda)) = a_j^\varepsilon \cdot q_j \cdot (1 - E(\exp(-\lambda Z_j^{j,\varepsilon} | Z_j^{j,\varepsilon} > 0))).$$

The l.h.s. of the equation above converges to 1 for every $\lambda \in (0, \infty)$ as a consequence of (5.26) using the definition (5.24), (5.25). Hence

$$(5.51) \qquad a_j^\varepsilon q_j (1 - E(\exp(-\lambda Z_j^{j,\varepsilon} | Z_j^{j,\varepsilon} > 0))) \underset{j\to\infty}{\longrightarrow} 1.$$

Since the r.h.s. is independent of λ we can conclude that vague limit points of $\mathcal{L}(Z_j^{j,\varepsilon} | Z_j^{j,\varepsilon} > 0)$ are convex combinations of δ_0, δ_∞ that is have the form $\beta\delta_0 + (1-\beta)\delta_\infty$. We need to exclude the case $\beta > 0$ (recall that $\bar{\varepsilon} = \varepsilon \cdot b(j,k)$, so that we do *not* have $Z_j^{j,\varepsilon} \geq \delta > 0$ for some δ for all $j \in I\!N$!). In the case of the vague limit $\beta\delta_0 + (1-\beta)\delta_\infty$ we must have that (along a subsequence) $a_n^\varepsilon q_n \to (1-\beta)^{-1}$. In order to establish $\beta = 0$, it would therefore according to (5.25) and (5.26) be enough to have uniform convergence of $G^{(j)}(\lambda)$ in λ, $\lambda \in [r, \infty)$ for some $r > 0$ as $j \to \infty$. Note for this purpose that

$$\psi_k^{\varepsilon,j}(0) = 0,\ (\psi_k^{\varepsilon,j})^{(1)}(0) = 1,\ (\psi_k^{\varepsilon,j})^{(2)}(\lambda) < 0,\ \psi_k^{\varepsilon,j}(\infty) < \infty. \tag{5.52}$$

Suppose now that $\psi_0^{\varepsilon,j}(\infty)$, (the first term in the composition of $G^{(j)}(\cdot)$ see (5.25)) is bounded also in j. Then $[0,\infty)$ would be mapped onto $[0, \psi_0^{\varepsilon,j}(\infty))$ under $\psi_0^{\varepsilon,j}$ and the images of the latter interval are bounded intervals in $[0, \psi_0^{\varepsilon,j}(\infty)]$. In such an interval we obtain then immediately (recall (5.33), (5.41)) the desired uniform convergence of $G^{(j)}(\lambda)$ for $\lambda \in [r, \infty)$, $r > 0$. The strategy of the proof is therefore to modify the last transition of our chain $(Z_k^j)_{k=-j-1,\dots,0}$, obtain the result and later remove the perturbation. The simplest modification is to put the Levy-measure of $\mathcal{L}(Z_0^{j,\varepsilon} | Z_{-1}^{j,\varepsilon} = 1)$ equal to zero on the interval $[0, \delta]$, where δ can be made arbitrarily small.

Step 5 Approximation, Proof of (5.10)(ii).

The basic strategy will be to construct an array $\bar{\varepsilon} = \varepsilon b(j,k) > 0$ such that (5.10)(ii) will hold. Recall that due to the representation (5.12) and (5.14) the two processes $(Z_k^{\varepsilon,j})_{k=-j-1,\dots,0}$ and $(Z_k^j)_{k=-j-1,\dots,0}$ can be represented via Poisson point processes with random intensity measure. Since the processes which we compare can be viewed as built from 3 Poisson point processes we will consider a 3-type branching process, where the two processes to be compared appear as the sum of two types each (one common) so that estimating the difference between the processes will amount to estimate the two populations associated with the types which do not occur in both Z^j and $Z^{j,\varepsilon}$, that is in the ones generated by the following parts of the canonical measures: $(1-\alpha)R_d 1_{(\varepsilon,\infty)}$, respectively $R_d 1_{(0,\varepsilon]}$.

For this purpose we have to choose for fixed j a sequence of ε-values determining the approximation in each step. Hence we have to choose $\bar{\varepsilon} = \{\varepsilon b(j,k),\ k \leq j,\ k,j \in I\!N\}$ appropriately. For a branching system generated by the Levy measure R we introduce masses of three types 1,2,3 according to the following rules:

- The "offspring" of mass of type 1 is of type 1, 2 or 3 and is generated by the Levy-measures $R_{(\varepsilon,\infty)}$ $(\alpha(\varepsilon)-1)R_{(\varepsilon,\infty)}$ respectively $R_{[0,\varepsilon]}$. ($R_{[a,b]}$ is R restricted to $[a,b]$)
- The "offspring" of mass of type 2 respectively 3 is generated by the Levy-measures $\alpha(\varepsilon)R_{(\varepsilon,\infty)}$ respectively $R_{(0,\infty)}$ and is always of the same type.

We denote by $(A_k^{j,\varepsilon})_{-j-1,\dots,0}$, $(B_k^{(j,\varepsilon)})_{-j-1,\dots,0}$ and $(C_k^{(j,\varepsilon)})_{-j-1,\dots,0}$ the population of types 1, 2, 3 and time k with respect to the mechanism prescribed for given j and $\bar{\varepsilon}$.

To conclude the desired asymptotics of (5.10) (ii) it suffices to prove that (recall that $B^{j,\varepsilon}$, $C^{j,\varepsilon}$ are submartingales!).

$$\mathrm{Prob}(B_0^{j,\varepsilon} \geq \frac{\varepsilon}{4} \text{ or } C_0^{j,\varepsilon} \geq \frac{\varepsilon}{4}) = o(\sum_0^j d_j^{-1}), \tag{5.53}$$

However the processes $(B_k^{j,\varepsilon})_{k=-j-1,\dots,j}$ and $(C_k^{j,\varepsilon})_{k=-j-1,\dots,j}$ have the structure of *critical processes* (recall (4.16)) with time inhomogeneous *immigration*, the latter created by an independent source. We use Chebychev to estimate

$$\begin{aligned} \mathrm{Prob}(B_0^{j,\varepsilon} \geq \frac{\varepsilon}{4}) &\leq \frac{16}{\varepsilon^2} E(B_0^{j,\varepsilon})^2 \\ \mathrm{Prob}(C_0^{j,\varepsilon} \geq \frac{\varepsilon}{4}) &\leq \frac{16}{\varepsilon^2} E(C_0^{j,\varepsilon})^2. \end{aligned} \tag{5.54}$$

The r.h.s. can be estimated by decomposing $B_0^{j,\varepsilon}$, $(C_0^{j,\varepsilon})$ according to the time the masses imigrated, which gives a decomposition of the form

$$(5.55)\qquad B_0^{j,\varepsilon} = \sum_{k=-j}^{-1} \tilde{B}_k^{j,\varepsilon}, \quad (\tilde{B}_k^{j,\varepsilon})_{k=-j,\dots,0}.$$

Here $\tilde{B}_k^{j,\varepsilon}$ is derived as $\tilde{B}_{k,0}^{j,\varepsilon}$ from processes $\tilde{B}_{k,\cdot}^{j,\varepsilon}$ which are independent and follow the branching mechanism of the process $B^{j,\varepsilon}$ but have starting time $-k$ and initial mass $\tilde{m}_k^{j,\varepsilon}$. The latter is random but independent of the evolution of $B^{j,\varepsilon}$ and $B_{\ell,\cdot}^{j,\varepsilon}$ $(\ell \neq k)$ and is given by the mass of type 2, which the process $A^{j,\varepsilon}$ generates at time $-k$. Similarly we proceed with $C_0^{j,\varepsilon}$.

Then (with Const, Const' independent of ε) and with $D_j = \sum_{k=0}^{j} d_k^{-1}$:

$$(5.56)\qquad \begin{aligned} E(B_0^{j,\varepsilon})^2 &= \sum_{k=0}^{j}(E(m_k^j))^2 + \sum_{k=0}^{j} E(m_k^j)\sum_{i=0}^{j-k}\frac{1}{d_i} \\ &\le \left(\sum_{k=0}^{j} E(m_k^j)\right)^2 + \left(\sum_{k=0}^{j} E(m_k^j)D_j\right) \\ &\le \left(\sum_{k=0}^{j} \varepsilon b(j,k)\theta\right)^2 + \left(\sum_{k=0}^{j} \varepsilon b(j,k)\theta D_j\right). \end{aligned}$$

Hence if we choose

$$(5.57)\qquad b(j,k) = \frac{\varepsilon^2}{j^2(vD_j)2^{k+1}16\cdot(\sum_{k=1}^{j} d_k^{-1})^2}$$

we can bound the r.h.s. of (5.54) by

$$(5.58)\qquad \text{Const'}\ \frac{\varepsilon}{(\sum_{k=1}^{j} d_k^{-1})^2},$$

which clearly is $o((\sum_0^j d_k^{-1})^{-1}$ for every ε (uniformly in ε).

The same reasoning applies to $(C_k^{j,\varepsilon})_{k=-j,\dots,0}$. This finishes the proof.

c) Proof of Theorem 5a) and 5b) for the case of branching diffusions

Step 1 We start by recalling a result from the literature on branching and extending it to our situation of branching masses. In a paper [1968] J. Lamperti and P. Ney proved the following result:
Let $(U_n)_{n\in I\!N}$ be a critical branching process (discrete time) on $I\!N$, with $U_0 = 1$ and with time homogeneous transition mechanism satisfying

$$(5.59)\qquad E(U_1) = 1 \quad E(U_1^2) = 1 + \sigma^2 < \infty.$$

Then the process

$$(5.60)\qquad \hat{U}_\alpha^{(n)} = U_{[\alpha n]}/n\sigma^2$$

satisfies (recall (1.8)-(1.9) for the definition of (Z_α))

$$(5.61)\qquad \mathcal{L}((\hat{U}_\alpha^{(n)})_{\alpha\in[0,1]}|\hat{U}_1^{(n)} > 0) \underset{n\to\infty}{\Longrightarrow} \mathcal{L}((Z_\alpha)_{\alpha\in[0,1]}).$$

The proof is by evaluating the Laplace-functionals of both sides of (5.61), using only three properties of the process U_n, namely, (i) the *branching property,* (ii) *criticality* of the branching and (iii) the following *asymptotics* of the time n *nonextinction probability*:

$$(5.62)\qquad \text{Prob}(U_n > 0) \sim \frac{1}{2\sigma^2}n^{-1} \quad \text{as } n \to \infty,$$

$$\text{Prob}(U_{[\alpha n]} > 0) \sim \alpha^{-1} \frac{1}{2\sigma^2} n^{-1}.$$

A similar asymptotic has been derived by Jagers [J] for a time inhomogeneous transition mechanism provided offspring distributions in each step are critical, uniformly square integrable and the variances σ_k^2 are bounded away from 0 uniformly in k. Namely if $EU_k^2 = 1+\sigma_k^2$ then

$$\text{(5.63)} \qquad \begin{aligned} &\text{Prob} \quad (U_n > 0) \sim \tfrac{1}{2}\Big(\sum_{k=0}^{n-1} \sigma_k^2\Big)^{-1}, \\ &\text{Prob} \quad (U_{[\alpha n]} > 0) \sim \alpha^{-1}\tfrac{1}{2}\big(\textstyle\sum_{k=0}^{n-1} \sigma_k^2\big)^{-1}. \end{aligned}$$

Now we can get, due to the fact only the properties (i) - (iii) were used in the proof, with the arguments in Lamperti and Ney the relation (5.61) if we replace (5.60) by

$$\text{(5.64)} \qquad \hat{U}_\alpha^{(n)} = U_{[f_\alpha(n)]} / \sum_{k=0}^{n-1} \sigma_k^2,$$

with

$$f_\alpha(n) : \sum_{k=0}^{f_\alpha(n)} \sigma_k^2 / \sum_{k=0}^{n} \sigma_k^2 \underset{n \to \infty}{\longrightarrow} \alpha.$$

Step 2 For us this discussion in step 1 means that all we have to do in order to carry out the same calculation as Lamperti and Ney [1968] is to establish the relation (5.8), which is the analog of (5.63) for masses instead of particles. Hence all we need to do now is to verify the assumptions of Proposition 5.1 which asserts (5.8). Here are the details:

First of all (5.7) is satisfied by assumption for Theorem 5 a). In order to verify (5.1) - (5.6) we now use that we restrict ourselves in this section to the case $g(x) =$ const. x. This means the set of kernels we need are given by (w.o.l.g. $g(x) = x$)

$$\text{(5.65)} \qquad K_d(\theta, \cdot) = \Gamma_\theta^{d,x}(\cdot).$$

However the distribution $\Gamma_\theta^{d,x}$ is the Gamma distribution with parameters $(\theta/d,\ d)$. Hence

$$\text{(5.66)} \qquad \Gamma_\theta^{d,x}(dz) = \frac{1}{\Gamma(\nu)} \alpha^{-2} z^{\nu-1} e^{-z/\alpha} dz$$

with $\nu = \theta/d$ and $\alpha = d$. This distribution is supported by $(0, \infty)$ hence (5.1) holds. The branching property (5.2) holds, which is checked using that the Laplace transform $L(\lambda)$ of $\Gamma_\theta^{d,x}(\cdot)$ given below is multiplicative in θ:

$$\text{(5.67)} \qquad L(\lambda) = \exp(-\theta d \int_0^\infty (1 - e^{-\lambda u/d}) \frac{e^{-u}}{u} du).$$

¿From this expression it can be derived that

$$\text{(5.68)} \qquad \begin{aligned} \int \Gamma_\theta^{d,x}(dz) z &= \theta \\ \int \Gamma_\theta^{d,x}(z) z^2 &= \theta^2 + \frac{1}{d}\theta = \frac{\theta}{d}(1 + \theta d) \\ \int \Gamma_\theta^{d,x}(z) z^3 &= \frac{\theta}{d^2}(2 + \theta d)(1 + \theta d) \\ \int \Gamma_\theta^{d,x}(z) z^4 &= \frac{\theta}{d^3}(3 + \theta d)(2 + \theta d)(1 + \theta d). \end{aligned}$$

Hence (5.3) - (5.6) hold and we are finished in the case $g(x) = \text{const} \cdot x$.

d) Proof of Theorem 5a) and 5b) for $g \in \mathcal{G}((d))$.

The proof is based on a comparison argument for expectations of certain convex functions with respect to branching and more general diffusions. A small technical problem arises if $g(x)$ vanishes too fast at 0, which we handle later.

¿From [BCGH2] we know that if $\liminf_{x \to 0} g(x)/x^2 > 0$ then $\mathcal{F}^{(n)}(g)$ has positive derivative at 0 for $n \geq n_0$. We can furthermore assume under these conditions according to that paper that $\mathcal{F}^{(n)}(g)(\theta)/\theta$ converges to d in the L_∞ norm. We assume for now: $g(x)/x^2 \geq \delta > 0$ for $x \in [0, \varepsilon)$ and we will later remove the restriction.

For such a function $g \in \mathcal{G}((d))$ we can find sequences d_n^-, d_n^+ such that

$$
\begin{array}{rcl}
d_n^- \theta & \leq & \mathcal{F}^{(n)}(g)_{(\theta)} \leq d_n^+ \theta \quad \theta > 0, n \in I\!N \\
d_n^+ & \geq & d_n^- \\
d_n^+ & - & d_n^- \underset{n \to \infty}{\longrightarrow} 0.
\end{array}
\tag{5.69}
$$

The crucial point now is that we do not have the branching property anymore (only for large θ asymptotically). We do have criticality, i.e. the martingale property of the markov chain. However in order to carry out the calculation needed to show the convergence of the Laplace transform of the multidimensional distributions (see Lamperti and Ney 1968) it suffices to establish that the following holds:

$$
\operatorname{Prob}(Z_0^j \geq \varepsilon | Z_{-j-1}^j = \theta) \sim \frac{\theta}{2} \sum_0^j c_k^{-1} \text{ as } j \to \infty
$$

$$
\mathcal{L}(Z_0^j (\frac{\theta d}{2} \sum_{k=0}^j c_k^{-1})^{-1} | Z_{-j-1}^j = \theta, Z_0^j \geq \varepsilon) \Longrightarrow \exp(1) \text{ as } j \to \infty.
$$

The last relation is however nothing but the assertion of the theorem for the special one dimensional marginal distribution at $\alpha = 1$. Hence the proof will be completed using the following, which we shall prove below:

Lemma 5.5 *Suppose that $g(x) \leq dx$ (resp. $g(x) \geq dx$). Define $f_\lambda(x) = e^{-\lambda x}$ for $\lambda \in [0, \infty)$, $x \in [0, \infty)$. Denote again by*

$$
K^{c,g}(\theta, d\varphi) = \Gamma_\theta^{c,g}(d\varphi).
$$

Then for all $m \in \{0, 1, \ldots, n\}$:

$$
\delta_\theta K^{c_n, \mathcal{F}^{(n)}(g)} \circ \ldots \circ K^{c_m, \mathcal{F}^m(g)}(f_\lambda) \underset{(\geq)}{\leq} \delta_\theta K^{c_n, dx} \circ \ldots \circ K^{c_m, dx}(f_\lambda)
\tag{5.70}
$$

$$
\delta_\theta K^{c_n, \mathcal{F}^{(n)}(g)} \circ \ldots \circ K^{c_m, \mathcal{F}^{(m)}_{(g)}}([\varepsilon, \infty)) \underset{(n-m) \to \infty}{\sim} \frac{\theta}{2} \sum_m^n c_k^{-1}. \qquad \square
\tag{5.71}
$$

To continue the proof we need only to observe that the statements of Theorem 5a), 5b) for $g(x) = dx$ were derived in the previous subsection using Laplace transforms and therefore we can proceed as follows (recall we only need to prove the result for the $\alpha = 1$ one dimensional marginal). First introduce

$$
L_g^{n,m}(\lambda) = 1 - E(\exp(-\lambda Z_{-m}^n) | Z_{-n-1}^n = \theta).
\tag{5.72}
$$

Then the relation (5.70) implies that for $m = m(n) \leq n$:

$$(5.73) \qquad L^{n,m}_{d_n^+ x}(\lambda) \leq L^{n,m}_g(\lambda) \leq L^{n,m}_{d_n^- x}(\lambda).$$

Consequently, if $m(n)$ and a_n, b_n are such that for a function L^* (depending possibly on $m(n)$) which is the Laplace transform of a distribution the following relation holds

$$(5.74) \qquad a_n L^{n,m}_{d(n)x}(b_n\lambda) \underset{n\to\infty}{\longrightarrow} L^*(\lambda), \text{ whenever } d(n) \underset{n\to\infty}{\longrightarrow} d \in (0,\infty)$$

then also for every $g \in \mathcal{G}((d))$

$$(5.75) \qquad a_n L^{n,m}_g(b_n\lambda) \underset{n\to\infty}{\longrightarrow} L^*(\lambda).$$

In order to finish from here the proof of Theorem 5 we need that the following is true. Let (X_n) be $\mathbb{R}^+$-valued random variables. A_n be the event $(X_n \geq \varepsilon)$ with $P(A_n) \to 0$ as $n \to \infty$. Then

$$(5.76) \qquad a_n(1 - E(\exp(-\lambda(X_n/b_n))) \quad \underset{n\to\infty}{\longrightarrow} \quad 1 - L(\lambda), \quad L(\lambda) = E^\mu(\exp(-\lambda x))$$

is equivalent to

$$(5.77) \qquad \mathcal{L}(X_n/b_n|A_n) \underset{n\to\infty}{\Longrightarrow} \mu, \;\; E^\mu(e^{-\lambda x}) = L(\lambda) \;\text{ and }\; a_n \sim [P(A_n)^{-1}].$$

Choose now

$$(5.78) \qquad a_n = d/2\sum_{k=0}^n c_k^{-1}, \quad b_n = d/2\Big(\sum_{k=m(n)}^n c_k^{-1}\Big)^{-1}.$$

Then we can conclude with (5.71) and the fact that we have proved the convergence for the case $g(x) = dx$ by reading (5.76) from right to left that

$$(5.79) \qquad (d/2\sum_0^n c_k^{-1})L^{n,m}_g(\lambda \cdot \sum_m^n \frac{d}{2}c_k^{-1}) \underset{n\to\infty}{\longrightarrow} L^*(\lambda).$$

This completes the proof with (5.76).

Proof of Lemma 5.5 We start with proving (5.70). The proof builds on the fact that the map $\mathcal{F}$ is order preserving and $\mathcal{F}(dx) = dx$, see in [BCGH,2], so that for all $n \in \mathbb{N}$ $g(x) \leq dx$ (resp. $\geq$) implies

$$(5.80) \qquad \mathcal{F}^{(n)}(g)_{(\theta)} \leq d\theta \quad (\mathcal{F}^{(n)}(g)_{(\theta)} \geq d\theta) \quad \theta \in [0,\infty).$$

Now we can complete the proof by exhibiting a class $\mathcal{M}$ of functions, which contains $\{f_\lambda | \lambda \in [0,\infty)\}$ and the following two relations hold:

$$(5.81) \qquad \delta_\theta K^{c,g}(f) \leq \delta_\theta K^{c,dx}(f) \quad \forall\, \theta \geq 0, \;\; f \in \mathcal{M} \text{ whenever: } g(x) \geq dx, \; \forall x \in [0,\infty),$$

$$(5.82) \qquad f \in \mathcal{M} \Longrightarrow (\theta \to \delta_\theta K^{c,g}(f) \in \mathcal{M} \quad \text{for every } g \in \mathcal{G}).$$

Now (5.80), (5.81) and (5.82) clearly imply the assertion (5.70).

To complete the proof we have to choose $\mathcal{M}$ and verify (5.81) and (5.82). We set

$$(5.83) \qquad \mathcal{M} = \{f : [0,\infty) \to \mathbb{R}^+ | f \quad \text{is completely monotone and bounded}\}.$$

Recall that completely monotone bounded positive functions are Laplace transforms of positive measures [F]. Then the class $\mathcal{M}$ is clearly preserved since if f is Laplace transform of a distribution ν then the new function is Laplace transform of the distribution $\nu K^{c,g}$. Furthermore it suffices to verify (5.70) for all $f_\lambda : f_\lambda(x) = e^{-\lambda x}$ where λ runs through $[0,\infty)$, since the closed convex hull of this family of functions is $\mathcal{M}$ (with respect to the supnorm). That is we should show that

$$\delta_\theta K^{c,g}(f_\lambda) \leq \delta_\theta K^{c,dx}(f_\lambda). \tag{5.84}$$

We are now going to prove this relation (5.84).

Consider the two semigroups V_t and U_t which are defined on $L^\infty(\mathbb{R}, dx)$ by

$$V_t(f) = Ef(Y_t) \qquad U_t(f) = Ef(X_t) \tag{5.85}$$

where

$$\begin{aligned} dY_t &= c(\theta - Y(t))dt + \sqrt{2g(Y_t)}dw_t && (5.86)\\ dX_t &= c(\theta - X(t))dt + \sqrt{2dX_t}dw_t. && (5.87)\end{aligned}$$

If we know that $g(x) \geq dx$, then with G_U, G_V denoting the generators of U_t respectively V_t, we conclude

$$G_V(f) \geq G_U(f) \qquad \forall f \in C_2(\mathbb{R}), \quad f'' \geq 0. \tag{5.88}$$

For functions f such that $U_t(f) \in \mathcal{V}(G_U) \cap \mathcal{V}(G_V)$ ($\mathcal{V}$ = domain) we know that

$$V_t(f) = U_t(f) + \int_0^t V_{t-s}(G_V - G_U)U_s(f)ds. \tag{5.89}$$

Hence for every function f satisfying

$$U_s(f) \in C^2(\mathbb{R}), \qquad U_s(f) \quad \text{convex for all} \quad s \geq 0, \quad \|f\|_\infty < \infty, \tag{5.90}$$

we have

$$V_t(f) \geq U_t(f). \tag{5.91}$$

In particular denoting by Π_U, Π_V the unique equilibria of the semigroups $(U_t)_{t\geq 0}, (V_t)_{t\geq 0}$ we conclude that for every f satisfying (5.90) we know

$$\langle \Pi_U, f\rangle \leq \langle \Pi_V, f\rangle. \tag{5.92}$$

In order to prove (5.84) we need in view of (5.92) only prove that the function $f_\lambda : x \to e^{-\lambda x}$ satisfies the assumptions (5.90). The nontrivial part is to show that:

$$z \to E_z(\exp(-\lambda X_t)) \qquad \text{is convex for every} \quad t \in [0,\infty),\ \lambda \in [0,\infty). \tag{5.93}$$

The latter will be proved by giving a representation of the function in (4.93) which allows immediately to read of the convexity.

Define the process $(Z_t)_{t\geq 0}$ on $[0,\infty)$ by:

$$dZ_t = -cZ_t + \sqrt{2dZ_t}dw_t, \quad Z_0 = z \in [0,\infty). \tag{5.94}$$

This is a subcritical branching diffusion, which satisfies by the branching property, respectively, the coupling principle

$$\begin{aligned} &E_{z_1+z_2}\exp(-\lambda Z_t) = E_{z_1}\exp(-\lambda Z_t)E_{z_2}\exp(-\lambda Z_t).\\ &z \to E_z \exp(-\lambda Z_t) \quad \text{is continuous in} \ z. \end{aligned} \tag{5.95}$$

Since our process $(X_t)_{t\geq 0}$ is a subcritical branching process (of the above type) with additional *immigration* (at a rate independent of the current state) we have

$$E_z\exp(-\lambda X_t) = E_z\exp(-\lambda Z_t)E_0(\exp(-\lambda X_t)). \tag{5.96}$$

Combining (5.96) with (5.95) we can write using the branching property for Z_t:

$$E_z(\exp(-\lambda X_t)) \;=\; \exp(-z\psi_t(\lambda))\exp(-\Phi_t(\lambda)) \quad \text{for a nonnegative function } \Phi_t(\cdot). \tag{5.97}$$

¿From the last relation we can conclude immediately

$$(\frac{\partial}{\partial z})^2 E_z(\exp(-\lambda X_t)) = (\psi_t(\lambda))^2 E_z \exp(-\lambda X_t) \geq 0 \tag{5.98}$$

which proves (5.93) and hence completes the proof of (5.84).

Finally we prove (5.71). We use (5.76) and Proposition 5.1, which gives the result for $g(x) = dx$. For the latter case observe that the variance of $\Gamma_\Theta^{c,dx}$ is simply $(d/c)\Theta$ so that we get (5.71) for $g(x) = dx$. Now by (5.76) read right to left in combination with (5.74) and (5.75) we obtain with reading (5.76) left to right the conclusion for $g \in \mathcal{G}((d))$.

In the case, where $g(x)/x$ is not bounded below as $x \to 0$ we use for comparison from below functions $g_\delta(x)$ of the form

$$g_\delta(x) = d(x-\delta)^+.$$

This generates for fixed δ a process $(Z_k^j)_{k=j-1,\dots,0}$ with the property that $(Z_k^j-\delta)_{k=j-1,\dots,0}$ is a process generated by dx and then all the arguments carry over since we can choose δ arbitrarily small due to the fact that for every $x \in [0,\infty)$ (see BCGH 2)

$$\mathcal{F}^n(g)(x) \underset{n\to\infty}{\longrightarrow} dx, \; \sup_{x\geq\delta}(\mathcal{F}^{(n)}(g)(x)/dx) \underset{n\to\infty}{\longrightarrow} 1 \quad \text{for all } \delta > 0.$$

(B) Proof of Theorem 6

The key to this theorem is the representation of the interaction chain via the Levy-measure associated with Γ_θ^k in the case where $g(x) = \text{const } x$ and a scaling argument using the explicit form of $\Gamma_\theta^k(\cdot)$ in that case. The generalization to $g \in \mathcal{G}((d))$ is as in 5 d) and we will not give the details again.

Proof of part (a)

The relation (1.23) is a direct consequence of the relation (2.6) and the assumption (1.21). From here (1.22) is immediate. Since (1.23) implies that the variance of $Z^j_{f_j(\tilde\alpha)}$ given $Z^j_{f_j(\alpha)}$ with $0 < \alpha < \tilde\alpha \leq 1$ vanishes after scaling with $\sum_0^j c_k^{-1}$ the limiting process is constant. This constant must be 0 since Z^j_{-j} scaled goes to 0.

Proof of part (b)

(i) Assume that $g(x) = x$ first. Consider the Markov chain $\tilde Z_k^j = c^j Z_k^j;\; k \in \{-j-1,-j,\dots,0\}$ and denote for $k \in I\!N$, deviating from earlier conventions, the corresponding transition kernels at time $-k-1$ by K^j_{-k}. We begin by calculating the Laplace transform of $K^j_{-k}(\theta,\cdot)$. The original chain $(Z_k^j)_{k=-j-1,-j,\dots,0}$ had transition kernels K_{-k} with the property that the law $K_{-k}(\theta,\cdot)$ is infinitely divisible with a Laplace transform $L_{k,\theta}(\lambda)$ given by

$$\begin{aligned} L_{k,\theta}(\lambda) &= \exp(-\theta\psi_k(\lambda)) \\ \psi_k(\lambda) &= c_k \int_0^\infty (1-e^{-\lambda u/c_k})\frac{e^{-u}}{u}du. \end{aligned} \tag{5.99}$$

Hence if we write $L^j_{k,\theta}(\lambda)$ for the Laplace transform of the transformed kernel $K^j_{-k}(\theta,\cdot)$ then

$$L^j_{k,\theta}(\lambda) = \exp(-\theta\psi_k^j(\lambda)) = \exp(-\Theta c^{-j}\psi_k(c^{-j}\lambda)). \tag{5.100}$$

Then one has

$$\psi_k^j(\lambda) = \int_0^\infty (1-e^{-\lambda u})c^{k-j}\exp(-c^{k-j}u)\frac{du}{u}. \tag{5.101}$$

The next step is to investigate the behaviour of the transformed chain. We consider first a starting point *not* scaled. For this purpose let $(X_n^j)_{n=-j-1,-j,\dots,0}$ be the Markov chain, which starts in θ with transition kernels K_{-k}^j. Note that

$$K_{-k}^j(\theta,\cdot) = \Gamma_\theta^{c^{k-j},x}(\cdot). \tag{5.102}$$

Put

$$\tilde{X}_n^j = X_{n-j-1}^j \quad n = 0,1,\dots,j+1. \tag{5.103}$$

This object we can imbedded in the chain $(\widehat{X}_n^\theta)_{n\in I\!N}$, which is a Markov chain with initial value resp. transition kernel

$$\begin{aligned}\widehat{X}_0 &= \theta \\ \widehat{K}_n(\theta,\cdot) &= \Gamma_\theta^{c^{-n},x}(\cdot), \quad n \in I\!N.\end{aligned} \tag{5.104}$$

We know that

$$\widehat{X}_n^\theta = \tilde{X}_n^j \quad \text{for } n = 0,1,\dots,j+1. \tag{5.105}$$

Since $(\widehat{X}_n^\theta)_{n\in I\!N}$ is uniformly integrable (Var $(\widehat{X}_n^\theta) \le \sum_{n=0}^\infty ((\frac{1}{c})^n)^{-1} = \sum_0^\infty c^n < \infty,\ c < 1!$) and $E\widehat{X}_n^\theta = \theta$, $(\widehat{X}_n^\theta)_{n\in I\!N}$ is a uniformly integrable positive Martingale. Hence

$$\widehat{X}_n^\theta \underset{n\to\infty}{\longrightarrow} \widehat{X}_\infty^\theta \text{ a.s. }, E\widehat{X}_\infty^\theta = \theta. \tag{5.106}$$

Since $\mathcal{L}(\widehat{X}_n^\theta)$ is infinitely divisible, so is $\mathcal{L}(\widehat{X}_\infty^\theta)$ and we can write due to the branching property and $\mathrm{Var}(\widehat{X}_\infty^\theta) > 0$ for $\theta > 0$:

$$\begin{aligned}E\exp(-\lambda\widehat{X}_\infty^\theta) &= \exp(-\theta\chi(\lambda)) \\ \chi(\lambda) &= \int_0^\infty (1-e^{-\lambda u})R_\infty(du).\end{aligned} \tag{5.107}$$

(ii) As a first consequence we get from (5.106) that

$$\mathcal{L}(\tilde{Z}_0^j|\tilde{Z}_{-j-1}^j = \theta) \underset{j\to\infty}{\Longrightarrow} \mathcal{L}(\widehat{X}_\infty^\theta). \tag{5.108}$$

Since

$$E(\exp(-\lambda\tilde{Z}_0^j)|\tilde{Z}_{-j-1}^j = \theta) = \exp(-\theta\psi_j^j \circ \dots \psi_0^j(\lambda)) \tag{5.109}$$

we get

$$\psi_j^j \circ \dots \circ \psi_0^j(\lambda) \underset{j\to\infty}{\longrightarrow} \int_0^\infty (1-e^{-\lambda u})R_\infty(du). \tag{5.110}$$

Returning to our original problem (recall $\tilde{Z}_{-j-1}^j = c^j\theta$) we find

$$1 - E(\exp(-\lambda\tilde{Z}_0^j)|Z_0^j = \theta) \underset{j\to\infty}{\sim} c^j\theta\int_0^\infty (1-e^{-\lambda u})R_\infty(du). \tag{5.111}$$

This proves (1.24) in the case $g(x) = x$.

In order to prove (1.25) we show first that $\int R_\infty(du) = +\infty$. From the explicit form of the Levy-measure in the case $g(x) = dx$ we know that $\psi_j^j \circ \dots \circ \psi_0^j$ has a representation

$$\psi_j^j \circ \dots \circ \psi_0^j(\lambda) = \int (1-e^{-\lambda u})R_j(du)$$

with $R_j([0,\varepsilon)) = +\infty$. The fact that this carries over to R_∞ follows however from the historical representation in Theorem 8, which implies that the limiting point process must also have realizations with countably many points. Therefore we know that $\lim_{j\to\infty}(1 - E\exp(-\lambda \tilde{Z}_0^j | \tilde{Z}_{j-1}^j = \theta)$ cannot be represented as $1 - L^*(\lambda)$ and L^* being the Laplace transform of a probability distribution. If along a subsequence $P(Z_0^j > \varepsilon)c^{-j}$ converges to a number in $(0,\infty)$ then by (5.76) the r.h.s. of (5.11) would be of the form $1 - L^*(\lambda)$ with L^* the Laplace transform of a probability distribution. Due to the assumpotics of (1.23) this implies $c^{-j}P(Z_0^j > \varepsilon) \to \infty$ as $j \to \infty$, which proves (1.25) in the case $g(x) = x$.

The proof of the relation (1.26) follows by working with $\widehat{X}_m^\theta$ instead of $\widehat{X}_\infty^\theta$ in (5.107) - (5.111). The relation (1.27) follows from $\widehat{X}_m^\theta \to \widehat{X}_\infty^\theta$ ((5.106)) and $E\exp(-\lambda \widehat{X}_m^\theta) = \exp(\int_0^\infty (1 - e^{-\lambda u}) R_m(du)$.

(iii) The generalization to $g \in \mathcal{G}((d))$ is immediate with the relation (5.73) since everything is expressed in part b) of the Theorem 6 in terms of the Laplace transform.

Proof of part (c)
The relations (1.28), (1.29) are proved using moment calculations and Laplace transform methods in conjunction with comparison arguments. First assume that $g(x) = x$, later we will generalize.

Start by proving the convergence relation of (1.28). Put $L_j(\lambda) = E\exp(-\lambda \tilde{Z}_0^j)$. Then

$$L_j(\lambda) = \exp(-\theta c^j \int_0^\infty (1 - e^{-\lambda u}) R_j(du)). \tag{5.112}$$

Define by $\widehat{F}_j$ the distribution of $\tilde{Z}_0^j$ under the Palm measure. If we now define the corresponding Laplace transform $\widehat{L}_j(\lambda) = \int_0^\infty e^{-\lambda u} \widehat{F}_j(du)$, we obtain

$$\widehat{L}_j(\lambda) = -\theta^{-1} c^{-j} \frac{\partial}{\partial \lambda} L_j(\lambda).$$

¿From the above representation and since we already proved in part b) that $R_j \Longrightarrow R_\infty$ weakly as $j \to \infty$:

$$\begin{aligned}\widehat{L}_j(\lambda) \quad &= \quad (\int_0^\infty u e^{-\lambda u} R_j(du)) \exp(-\theta c^j \int_0^\infty (1 - e^{-\lambda u}) R_j(du)) \\ &\underset{j \to \infty}{\longrightarrow} \int_0^\infty e^{-\lambda u} u R_\infty(du) := \widehat{L}_\infty(\lambda).\end{aligned}$$

The next observation is that

$$\int_0^\infty u R_\infty(du) = 1$$

and hence $\widehat{L}_\infty(\lambda)$ is the Laplace transform of a probability measure whose distribution function we call $\widehat{F}_\infty$. This proves, (via the characterisation of weak convergence by the convergence of Laplace transforms and the fact that $\widehat{F}_\infty$ has no positive atoms) the first part of relation (1.28) for the case $g(x) = dx$. Again since everything is expressed in terms of the Laplace transform the comparison methods of section 5d) gives the result for $g \in \mathcal{G}((d))$. (See (5.69) - (5.73)).

In order to establish the formula for the Laplace transform of $\widehat{F}_\infty$ we prove below the following behaviour of the k-th moment of Z_0^j as $j \to \infty$ (the numbers D_k are defined below).

$$E(Z_0^j)^k \underset{j\to\infty}{\sim} D_k \theta (c^{-j})^{k-1} + R_{k,\theta}(j), \quad k = 1, 2, \ldots \tag{5.113}$$

$$\sum_{k=0}^\infty \frac{\lambda^k}{k!} [R_{k,\theta}(j)(c^{-j})^{k-1}] \underset{j\to\infty}{\longrightarrow} 0. \tag{5.114}$$

The numbers $(D_k)_{k\in\mathbb{N}}$ are determined as follows: Define first the numbers $(A_k^j)_{j=1,\dots,k},\ k\in\mathbb{N}$ by:

$$\theta(1+\theta)\dots((k-1)+\theta)=\sum_{j=1}^{k}A_k^j\theta^j. \tag{5.115}$$

Then define D_k recursively by

$$D_1=1 \tag{5.116}$$

$$D_k=\frac{1}{1-c^{k-1}}(A_k^{k-1}D_2+A_k^{k-2}c^{k-1}D_2+\dots+A_k^1c^{k-1}D_{k-1}).$$

With the above information we continue as follows. Put

$$F^j(du)=Prob(\tilde Z_0^j\in du) \tag{5.117}$$

and recall that

$$\widehat{F}^j=(E\tilde Z_0^j)^{-1}uF^j(du).$$

Then according to (5.113)

$$\int_0^\infty u^k\widehat{F}^j(du)\underset{j\to\infty}{\longrightarrow}D_{k+1}$$

and with (5.114) we have

$$\int_0^\infty e^{-\lambda u}\widehat{F}^j(du)\underset{j\to\infty}{\longrightarrow}\sum_0^\infty(-1)^k\frac{\lambda^k}{k!}D_{k+1}$$

which completes the proof. It remains therefore to verify the asymptotics of the k-th moments, as expressed in (5.113) and (5.114).

We consider the case where $g(x)=x$, since the assertion is as we already proved independent of g within $\mathcal{G}((1))$. In the special case where $g(x)=x$, we can use relation (5.68) to obtain with $A_k^k=1$ for all $k\in\mathbb{N}$:

$$\begin{aligned}\int\Gamma_\theta^{d,x}(z)z^k &= \frac{\theta}{d^{k-1}}((k-1)+\theta d)\dots(1+\theta d)\\ &= \sum_{j=1}^{k}A_k^jd^{-(k-j)}\theta^j.\end{aligned} \tag{5.118}$$

Define for $\ell\in\mathbb{Z}^-$

$${}^jM_\ell^{(k)}=E(Z_\ell^j)^k\quad\text{with}\quad Z_{-j-1}^j=\theta. \tag{5.119}$$

Then we obtain by conditioning on Z_ℓ^j in connection with (5.118) for the k-th moment at time $\ell+1$, that:

$${}^jM_{\ell+1}^{(k)}=(\sum_{n=1}^{k-1}A_k^n(\frac{1}{c_{\ell+1}})^{k-n}\,{}^jM_\ell^{(n)})+{}^jM_\ell^{(k)}. \tag{5.120}$$

Now we turn to the representation of $\tilde Z_0^j$ from the previous step of the proof, namely (5.99) - (5.107). We conclude that for ℓ fixed (note $A_k^1=1$)

$${}^jM_{\ell+1}^{(k)}/{}^jM_\ell^{(k)}\underset{j\to\infty}{\longrightarrow}\frac{1}{c^{k-1}}. \tag{5.121}$$

Then the first asymptotic relation follows by induction over k. The induction is started in $k=2$ which has been asymptotically evaluated already in part a) of the Theorem 5. ¿From the asymptotic relation for $g(x) = x$ we obtain by scaling (recall $\Gamma_\theta^{c,dx} = \Gamma_\theta^{c/d,x}$) the result for $g(x) = dx$. Then again we generalize with the comparison lemma of section 5 d).

Next we prove the relation (5.114). As before it suffices to consider the case $g(x) = x$. The agrument proceeds in two steps. First one shows that

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}[R_{k,\theta}(j)(c^{-j})^{k-1}] \underset{j\to\infty}{\longrightarrow} 0, \text{ for } \theta \in (0,\infty) \text{ sufficiently small.} \tag{5.122}$$

The next step is to extend the definition of $\Phi_j(\theta) = E\exp(-\theta Z_0^j)$ to values $\theta \in \mathbb{C}$ with $\mathcal{R}e\theta > 0$. If we can show that

$$|\Phi_j(\theta)| \leq \text{Const } |\theta| \tag{5.123}$$

then we have a family of holomorphic functions on the right half plane, which is bounded on compact sets. Combining then relations (5.122) and (5.123) gives the assertion, since such a family converges uniformly on compact sets if it converges on one interior point.

The two missing estimates in order to establish above relations are obtained as follows: The relation (recall 5.111)

$$|1 - E\exp(-\theta\tilde{Z}_0^j)| \leq |\theta c^j \int (1-e^{-\lambda u})R_j(du)| \tag{5.124}$$

together with $R_j \Longrightarrow R_\infty$ gives (5.123). For the relation (5.122) we want to use Lebesgue dominated convergence theorem for $|\theta|$ small enough. Hence we need that we can find a constant C such that

$$E((Z_0^j)^k | Z_{-j-1}^j = \theta) \leq c^{-jk} C^k k!. \tag{5.125}$$

This can be derived from the representation from (5.99) - (5.107) using that $E(\widehat{X}_n^\theta)^k \leq E(\widehat{X}_\infty^{\theta,k})^k$ and the variable $\widehat{X}_\infty^\theta$ satisfies $E\exp(\lambda\widehat{X}_\infty^\theta) < \infty$ for $\lambda \in (0,\varepsilon)$ with $\varepsilon > 0$ suitable. The latter follows from the explicit form of the transition kernel $\widehat{K}_n$ of (5.104).

In order to prove (1.29) we follow the same strategy as before, that is we use Laplace transforms in connection with the markovian structure. In particular it suffices to consider the case $g(x) = x$. In fact we have already established the asymptotics of the marginal distributions in part b) of this theorem. From the special representation in (5.102) - (5.107) we can in fact read off immediately that weak limit points of $\widehat{P}((\tilde{Z}_{-j-1+m}^j)_{m=1,\dots,k} \geq (u_m)_{m=1,\dots,k})$ is the Markov Process given by the expression of the r.h.s. of (1.29) (Recall the definition of Palm distributions preceeding Theorem 6). This completes the proof.

6 Proof of Theorems 7 - 12

The main work here has to be done for Theorems 8 - 12, since Theorem 7 can be treated in analogy to previous work. The most essential tools in this chapter are the properties of the Gamma distribution and the relation to resampling systems as recalled in the Appendix, the tools available for infinitely divisible random measures on polish spaces including Palm measures and finally previous result from the multiple space-time scale analysis in [DGV].

(i) Proof of Theorem 7

The process $(X^N(t))_{t\geq 0}$ can be viewed as a *super random walk* with respect to a transient walk. The transience of the random walk follows from the hypotheses $\sum_k \frac{1}{c_k} < \infty$ and (0.3) (cf. Appendix A4). Parts (a)-(d) of Theorem 7 are obtained by a straightforward modification of Theorem 6.3 of Dawson and Perkins (1991) for superprocesses on $\mathbb{R}^d$ instead of a superprocess on the discrete group Ω_N. The main difference is the replacement of the symmetric stable process in $I\!R^d$ there by the random walk in Ω_N in the present case. The state space E plays the role of the space $M_p(I\!R^d)$ there. The key property required of the state space is that the semigroup associated to the random walk maps the state space into itself. But the assumption (0.7) implies that for $x \in E$, $||a(x)|| = \sum_{i,j} a(i,j)x_j\alpha(i) \leq M||x||$ and hence the semigroup a_t of the random walk has, as a map $E \to E$, norm bounded by e^{Mt}, so that the random walk semigroup maps the space E continuously into itself.

(ii) Proof of Lemma 1, Theorem 8 and Corollaries 1,2

Proof of Lemma 1

(a) Computing second moments from the Gamma distribution we obtain for $\ell > k \geq 0$:

$$(6.1) \qquad E[(z^{\ell}_{-k}(u) - z^{\ell+1}_{-k}(u))^2]$$
$$= E[E[(z^{\ell}_{-k}(u) - z^{\ell+1}_{-k}(u))^2 | z^{\ell+1}_{k}(u)]$$
$$= \frac{d}{c_\ell} E[z^{\ell+1}_{-k}(u)] = \frac{du}{c_\ell}$$

Therefore for each u,

$$\lim_{j'\to\infty} \sup_{j\geq j'} E[(z^{j}_{-k}(u) - z^{j'}_{-k}(u))^2]$$
$$\leq \lim_{j'\to\infty} \sup_{j\geq j'} \text{ const } \sum_{\ell=j'}^{j} \frac{1}{c_\ell} = 0.$$

Then $z^{j}_{-k}(\theta)$ converges in distribution and in L^2 as $j \to \infty$. In addition the limit $z^{\infty}_{-k}(\theta)$ is a reverse martingale with mean measure θ. The almost sure convergence follows from the reverse martingale convergence theorem.

(b) By construction $z^{\infty}_{-k}(\theta)$ conditioned on $\{z^{\infty}_{-j}\}_{j>k}$ has the gamma distribution $\Gamma^{k}_{z^{j}_{-k-1}(\theta)}$. This means that the sequence $\{z^{\infty}_{k}\}_{k=-j-1,\ldots,0}$ is a Markov chain with Gamma transition kernels K_{-k} and therefore is a version of the entrance law for the interaction Markov chain $(z^{\infty}_{k})_{k\in Z\!\!Z^-}$ whose existence was established in Theorem 2.

Proof of Theorem 8

The properties (i) and (ii) are immediate consequences of the following construction: We begin with the sequence of independent gamma processes $(\gamma_k)_{k\in I\!N}$. For each $k \in Z\!\!Z^+$ let $\{m_{-k}(i)\}_{i\in I\!N}$ denote the *size ordered* jumps in $\{\gamma_k(u) : 0 \leq u \leq z^{\infty}_{-k-1}\}$, that is, $m_{-k}(i)$ is the *ith* largest jump of $\gamma_k(\cdot)$ in the interval $[0, z^{\infty}_{-k-1}]$. (Note that without loss of generality we can ignore ties.) Property (iii) follows from the basic property of the Moran gamma process described in (A3.3) under basic fact. This implies that conditioned on the independent gamma processes $(\gamma_j)_{j>k}$ the random variable $\gamma_k(z^{\infty}_{-k-1}(\theta))$ and the random vector $\{\frac{m_{-k}(i)}{z^{\infty}_{-k}}\}_{i\in I\!N}$ are independent. Moreover $\gamma_k(z^{\infty}_{-k-1}(\theta))$ has the distribution $\Gamma^{k}_{z^{\infty}_{-k-1}(\theta)}$ and $\{\frac{m_{-k}(i)}{z^{\infty}_{-k}}\}_{i\in I\!N}$ has the Poisson Dirichlet distribution with parameter $\frac{z^{\infty}_{-k-1}c_k}{d}$.

Property (iv) will follow immediately from the following construction: Starting with level zero we now make the following definitions introducing the "father-son" relation between jumps according to the time interval a jump occurs. However for this purpose we have to consider the size ordered jumps in the father generation, since the canonical measure of γ_k is infinite (near 0) and formally we proceed as follows. For each $i \in I\!N$, let $G(1,i)$ be the unique integer $n \in I\!N$ such that the jump of size $m_0(i)$ in γ_0 occurs in the interval

$$[\sum_{\ell<n} m_{-1}(\ell), \sum_{\ell\leq n} m_{-1}(\ell) \ ).$$

Similarly let $G(2,i)$ be the unique integer n such that the jump of size $m_{-1}(G(1,i))$ in γ_1 occurs in the interval $[\sum_{\ell<n} m_{-2}(\ell), \sum_{\ell\leq n} m_{-2}(\ell))$. Continuing in this way we inductively construct a sequence $G(1,i), G(2,i), \ldots$.

Property (v) is proved as follows. First consider only jumps in the gamma process γ_{k-1} of size ε or larger. These follow a Poisson process on the time interval of length z_{-k}^{∞} and the times of jumps are given by i.i.d. uniform random variables in $[0, z_{-k}^{\infty}]$. Therefore the probability that a given jump of size ε or larger occurs in the interval of length $m_{-k}(i)$ is equal to $\frac{m_{-k}(i)}{z_{-k}^{\infty}}$. We then let $\varepsilon \to 0$.

Property (vi). Given $\{m_{-k-1}(\cdot), m_{-k-2}(\cdot), \ldots\}$, the collection $(\sum_i m_{-k}(i) 1_{G_{-k-1}(i)=j})_{i\in I\!N}$ is given by the sums of the jumps of γ_k over disjoint intervals of lengths $m_{-k-1}(j)$ and therefore are independent $\Gamma(\frac{c_{k+1}m_{-k-1}(j)}{d}, \frac{d}{c_{k+1}})$ random variables.

Proof of Corollary 1

We first prove the result when $j=1$. By Theorem 8(c,iii), conditioned on the sequence Z_{-k}^{∞}, $\frac{m_{-k}(\cdot)}{Z_{-k}^{\infty}}$ are independent Poisson Dirichlet with parameters $\lambda_k := \frac{Z_{-k-1}c_k}{d}$. Moreover by Theorem 8(v), $m_{-k}(G(k,1))$ is obtained by size-biased sampling from $\frac{m_{-k}(\cdot)}{Z_{-k}^{\infty}}$. This means (cf. A3.4) that $\frac{m_{-k}(G(k,1))}{Z_{-k}^{\infty}}$ has distribution with density

$$\lambda_k(1-p)^{\lambda_k-1}, \quad 0\leq p\leq 1. \tag{6.2}$$

Hence the density of the rescaled random variable $\lambda_k \frac{m_{-k}(G(k,1))}{Z_{-k}^{\infty}}$ is

$$(1-\frac{p}{\lambda_k})^{\lambda_k-1}, \quad 0\leq p\leq \lambda_k.$$

But since $c_k\to\infty$ and $Z_{-k}^{\infty}\to\theta$, $\lambda_k\to\infty$. Noting that $\frac{Z_{-k-1}^{\infty}}{Z_{-k}^{\infty}}\to 1$, it is then elementary to verify that as $k\to\infty$,

$$\begin{aligned}
\mathcal{L}(\frac{c_k}{d}m_{-k}(G(k,1))) &\quad \Rightarrow \mathit{Exponential}(1)\\
E(\frac{c_k}{d}m_{-k}(G(k,1))) &\quad \to 1.
\end{aligned}$$

We obtain the result for $j>1$ in a similar way. If $i_1,\ldots,i_j$ belong to different families, then $m_{-k}(G(k,i_1)),\ldots,m_{-k}(G(k,i_j))$ are obtained by size-biased sampling (without replacement) from $\{\frac{m_{-k}(\cdot)}{Z_{-k}^{\infty}}\}$ and therefore has the same law as $v_{k,1}, (1-v_{k,1})v_{k,2},\ldots,(1-v_{k,1})\cdots(1-v_{k,j-1})v_{k,j}$ where $v_{k,r}$ are independent and each of their distributions has density (6.2) (cf. the GEM representation - A3.4). The proof is completed as above with the additional observation that $\max_{i=1,\ldots,j} v_{k,i} \Longrightarrow 0$ as $k\to\infty$ and therefore $\prod_{\ell=1}^{i-1}(1-v_{k,\ell}) \Longrightarrow 1$ for each $i=1,\ldots,j$ and $\lambda_k v_{k,i} \Longrightarrow \mathit{Exponential}(1)$.

Proof of Corollary 2

(a) This follows from the representation (1.39). Conditioned on $\{M_\ell^*(\cdot)\}_{\ell\leq -j}$, the random variables $\{M_{-(j-1)}^*(i)\}_{i\in I\!N}$ are given by the sum of the jumps of the Gamma process γ_{j-1} which occur in disjoint intervals of lengths $\{M_{-j}^*(i)\}_{i\in I\!N}$ and therefore have independent Gamma distributions $\{\Gamma_{M_{-j}^*(i)}^{j-1}\}_{i\in I\!N}$. This proves that they are independent inhomogeneous Markov chains with transition functions $\{\Gamma_\theta^{j-1}\}_{j\in I\!N}$. The martingale property follows since $\int x\Gamma_\theta^{j-1}(dx)=\theta$.

(b) By definition (in this step $j, k, \ell \in \mathbb{N}$ and i labels the families)

$$M^*_{-k}(i) = \sum_{j \equiv_k \ell^{-1}(i)} m_{-k}(G_{-k}(j)).$$

that is, $M^*_{-k}(i)$ consists of the mass of the subpopulation at level k having a common ancestor at some level $\ell \geq k$. Therefore it can be decomposed as

$$M^*_{-k}(i) = \sum_{\ell=k}^{\infty} M^*_{-k,-\ell}(i)$$

where $M^*_{-k,-\ell}(i)$ denotes the mass at level k of the subfamily of i which has a last common ancestor at level ℓ. We will compute the law of the $M^*_{-k,-\ell}(i)$ conditioned on the $\{m_{-\ell}(G_{-\ell}(i))\}_{\ell \geq k}$. We have $M^*_{-k,-k}(i) = m_{-k}(G_{-k}(i))$ and $M^*_{-k,-\ell}(i)$ for $\ell > k$ denotes the mass of the descendents at level k of $m_{-\ell}(G_{-\ell}(i))$ excluding descendents of $m_{-\ell+1}(G_{-\ell+1}(i))$. But conditioned on the $\{m_{-\ell}(G_{-\ell}(i))\}_{\ell \geq k}$, the descendents at level $\ell-1$ of $m_{-\ell}(G_{-\ell}(i))$ excluding descendents of $m_{-\ell+1}(G_{-\ell+1}(i))$ are given via $\gamma_{\ell-1}(m_{-\ell}(G_{-\ell}(i))$ conditioned on the event that $\gamma_{\ell-1}$ has a jump of size $m_{-\ell+1}(G_{-\ell+1}(i))$ in $[0, m_{-\ell}(G_{-\ell}(i))$. But (using in A2.3 the relation $(P)_x = P * Q_x$) and the identification of Q_x) the conditional law satisfies

$$\begin{aligned}(6.3)\qquad &\mathcal{L}[\gamma_{\ell-1}(m_{-\ell}(G_{-\ell}(i))) - m_{-\ell+1}(G_{-\ell+1}(i)) \\ &\qquad |\gamma_{\ell-1} \text{ has a jump of size } m_{-\ell+1}(G_{-\ell+1}(i))] \\ &= \Gamma^{\ell-1}_{m_{-\ell}(G_{-\ell}(i))}.\end{aligned}$$

At levels $\ell' < \ell - 1$ the mass of these descendents evolve by the interaction chain with transition function $\Gamma^{\ell'}$. Therefore $M^*_{-k,-\ell}(i) = \hat{Z}^{\ell}_{-k}(m_{-\ell}(G_{-\ell}(i)))$ where $\hat{Z}^{\ell}_{-k}(m_{-\ell}(G_{-\ell}(i))$ are independent copies of the interaction chain starting at level ℓ in $m_{-\ell}(G_{-\ell}(i))$.

(iii) Proof of Theorem 9

The key to this theorem are properties of special distributions arising in sampling systems. By Corollary 2(b), conditioned on $\{M^*_{-\ell} : \ell > k\}$, $\{M^*_{-k}(j)\}_{j \in \mathcal{I}}$ are independent $Gamma(\frac{c_{k+1} M^*_{-k-1}(j)}{d}, \frac{c_{k+1}}{d})$ random variables. On the other hand for the case of interacting Fleming Viot processes the corresponding kernel is given by a GEM-distribution, see (6.8) below (and the corresponding order statistics have Poisson Dirichlet distribution). We need some facts about these three special distributions.

We first recall the relation between Beta- and Gamma-distributions. Suppose that X, Y are both Gamma-distributed with parameter (α_1, β), respectively (α_2, β) and are independent. Then $X+Y$ and $X/(X+Y)$ are independent and $X/(X+Y)$ is Beta-distributed with parameter (α_1, α_2), while $X+Y$ is Gamma-distributed with parameter $(\alpha_1 + \alpha_2, \beta)$.

This means that we can characterize the $\mathcal{P}(\mathbb{N})$-valued random variable Q^*_k as follows. If $A \subseteq \mathbb{N}$, let

$$(6.4)\qquad Q^*_k(A) := \sum_{\ell \in A} Q^*_k(\ell) \quad k \in \mathbb{Z}^-.$$

Then for all $A \subset \mathbb{N}$, $k \in \mathbb{Z}^-$:

$$(6.5)\qquad Q^*_k(A) \text{ is Beta-distributed with parameter } (Q^*_{k-1}(A)/\gamma_k, (1 - Q^*_{k-1}(A))/\gamma_k)$$

and for any two disjoint subsets A_1, A_2 of $\mathbb{N}$,

$$(6.6)\qquad Q^*_k(A_1) + Q^*_k(A_2) \text{ and } \frac{Q^*_k(A_1)}{Q^*_k(A_1) + Q^*_k(A_2)} \quad \text{are independent}$$

and $\dfrac{Q_k^*(A_1)}{Q_k^*(A_1)+Q_k^*(A_2)}$ has a Beta distribution with parameter:

$$(\frac{Z^\infty_{-k+1}c_k}{d}Q^*_{k-1}(A_1), \frac{Z^\infty_{-k+1}c_k}{d}Q^*_{k-1}(A_2)).$$

This means that if we divide families into two groups, then the relative frequencies of the two groups evolve according to a kernel given by a Beta distribution.

It remains to show that $(q_k^*)_{k\in \mathbb{Z}^-}$ has the same structure. Under the hypothesis $\sum_k \frac{1}{c_k} < \infty$, $Z^\infty_{-k} \to \theta$, $d'_k \to d^*$ where $0 < d^* < \infty$ and $\frac{c'_k}{c_k} \to \frac{d^*}{d}\theta$ as $k \to \infty$ and hence $\sum \frac{d'_k}{c'_k} < \infty$. Therefore by Theorem 0.5 of [DGV] the associated interacting Fleming-Viot system has a unique non-degenerate entrance law with mean measure Θ. The forward transition mechanism $\Gamma_\theta^{c'_k,d'_k}$ on $\mathcal{P}([0,1])$ for the interaction chain of the Fleming-Viot system is given by the GEM representation (see Theorem 0.3, (0.25) of [DGV]) which is defined as follows:

(a) Let $(U_i)_{i\in \mathbb{N}}$ be an i.i.d. sequence with marginal distribution $\Theta \in \mathcal{P}([0,1])$.

(b) Let $(V_i^k)_{i\in \mathbb{N}}$ be an i.i.d. sequence with marginal Beta$(1, \frac{Z^\infty_{-k-1}c_k}{d})$.

(c) Construct $(U_i)_{i\in \mathbb{N}}$, $(V_i^k)_{i\in \mathbb{N}}$ as independent processes.

Set

$$\Gamma_\theta^{c'_k,d'_k} = \mathcal{L}\Big(\sum_{i=1}^\infty \Big[V_i^k \prod_{j=1}^{i-1}(1-V_j^k)\Big]\delta_{U_i}\Big) \in \mathcal{P}(\mathcal{P}([0,1])). \tag{6.7}$$

The representation (1.56) of the transition function for $(q_k^*)_{k\in\mathbb{Z}^-}$ is then obtained from (6.7) by taking a decomposition of the interval $[0,1]$ into disjoint subintervals of lengths $\{p_\ell\}_{\ell\in\mathbb{N}}$ as follows:

By combining the $m+1, m+2, \ldots$ pieces of this partition we obtain a *finite* partition. Note that for any such finite partition the distribution is a Dirichlet process corresponding to the measure $\{\tilde p_\ell\}_{\ell=1,\ldots,m}$ where $\tilde p_\ell = p_\ell$ if $\ell < m$ and $\tilde p_m = 1-\sum_{\ell=1}^{m-1} p_\ell$. Then $\mathcal{L}\Big(\sum\limits_{i=1}^\infty \Big[V_i^k \prod_{j=1}^{i-1}(1-V_j^k)\Big]\Big)$ has a Dirichlet distribution with parameters $\alpha_i = \frac{Z^\infty_{-k+1}c_k}{d}\tilde p_i$. Therefore for any two disjoint subsets A_1, A_2 of $\mathbb{N}$, $q_k^*(A_1)+q_k^*(A_2)$ and $\frac{q_k^*(A_1)}{q_k^*(A_1)+q_k^*(A_2)}$ are independent and $\frac{q_k^*(A_1)}{q_k^*(A_1)+q_k^*(A_2)}$ has a Beta distribution with parameter

$$(\frac{Z^\infty_{-k+1}c_k}{d}q^*_{k-1}(A_1), \frac{Z^\infty_{-k+1}c_k}{d}q^*_{k-1}(A_2)). \tag{6.8}$$

To see this, represent $q_k^*(A_1), q_k^*(A_2)$ and $q_k^*(\mathbb{N}\cap(A_1\cup A_2)^c)$ in terms of independent Gamma random variables (cf A3.2). Then $\frac{q_k^*(A_1)}{q_k^*(A_1)+q_k^*(A_2)}$ and the pair $q_k^*(A_1)+q_k^*(A_2), q_k^*(\mathbb{N}\cap(A_1\cup A_2)^c)$ are independent. This yields the required independence and the fact that $\frac{q_k^*(A_1)}{q_k^*(A_1)+q_k^*(A_2)}$ is Beta follows from the relation between the Beta and Gamma distributions referred to above.

(iv) Proof of Theorem 10

(a) By Theorem 9, the $\frac{M_k^*(i)}{Z^\infty_{-k}}$ have the same law as the entrance law $q_k^*(i)_{k\in\mathbb{N}}$ associated to the interacting Fleming-Viot system with coefficients $(c'_k)_{k\in\mathbb{N}}$ and $(d'_k)_{k\in\mathbb{N}}$ where the latter as given in terms of the $(c_k)_{k\in\mathbb{N}}$ and Z^∞_{-k} by (1.51). Therefore, since $Z^\infty_{-k}\to\theta$ as $k\to\infty$, it suffices to show that

$$\max_i q_k^*(i) \underset{k\to\infty}{\longrightarrow} 0.$$

But this now follows immediately from Theorem 0.7(a) of [DGV].

(b) Combining Theorem 9 and Theorem 0.7(b) of [DGV], we get

$$\lim_{\alpha\to 0}\lim_{k\to\infty}\sum_{i=1}^{\alpha c^k}\widehat{M}^*_{-k}(i)=0$$

and

$$\lim_{\alpha\to \infty}\lim_{k\to\infty}\sum_{i=1}^{\alpha c^k}\widehat{M}^*_{-k}(i)=1.$$

This second equality implies the tightness statement and the first equality implies that the limit is nondegenerate, that is, not δ_0.

(c) By Corollary 1, (1.43),

$$E(\frac{c_k}{d}m_{-k}(G_{-k}(i)))\underset{k\to\infty}{\longrightarrow}1, \tag{6.9}$$

and by Corollary 2(c) ,

$$M^*_{-k,\ell}(i)=\widehat{Z}^\ell_k(m_{-\ell}(G_{-\ell}(i))).$$

Since $(\widehat{Z}^\ell_k)_{k=-\ell-1,\dots,0}$ is a martingale

$$E[M^*_{-k,\ell}(i)]=E[(m_{-\ell}(G_{-\ell}(i)))].$$

Together with (6.11) this yields

$$\begin{aligned} E(M^*_{-k}(i)) &= \sum_{\ell\geq k}E(M^*_{-k,\ell}(i)) \\ &= \sum_{\ell\geq k}E(m_{-\ell}(G_{-\ell}(i)))\sim\sum_{\ell\geq k}\frac{d}{c_\ell}. \end{aligned} \tag{6.10}$$

(v) Proof of Theorem 11 and 12

Preparations

We begin with a Lemma that characterizes the interaction chain under the Palm measure $(P^\infty)_0$ defined by (1.62).

Lemma 6.1

(i) Under the Palm measure $(P^\infty)_0$ defined in (1.62), the interaction chain $\{Z^\infty_k\}_{k\in\mathbb{Z}^-}$ is a Markov chain with transition functions

$$P(Z^\infty_k\in\cdot|Z^\infty_{k-1}=\theta)=\widehat{\Gamma}^{k-1}_\theta(\cdot) \tag{6.11}$$

where $\widehat{\Gamma}^{k-1}_\theta(\cdot)$ is the Gamma$(1+\frac{c_{k-1}}{d}\theta,\frac{c_{k-1}}{d})$ distribution.

(ii) Under the Palm measure $(P^\infty)_0$, conditioned on $\{Z^\infty_k\}_{k\in\mathbb{Z}^-}$, the $\{\frac{m_{-k}(\cdot)}{Z^\infty_{-k}}\}_{k\in\mathbb{Z}^-}$ are independent and for each k, $\{\frac{m_{-k}(i)}{Z^\infty_{-k}}\}_{i\in\mathcal{I}}$ has the Poisson Dirichlet distribution with parameter $\frac{Z^\infty_{-k-1}c_k}{d}$.

(iii) The transition kernel of the interaction chain under the Palm measure has the following representation:

$$\widehat{\Gamma}^{k-1}_\theta=\Gamma^{k-1}_\theta * E_{k-1} \tag{6.12}$$

where E_{k-1} is the exponential distribution with mean $\frac{d}{c_k}$. $\square$

Proof (i) Since $\{Z_k^\infty\}_{k\in\mathbb{Z}^-}$ is both a Markov chain and a martingale, the function $h(x)=x$ is a harmonic function for all the kernels Γ_θ^{k-1}. Hence we can consider the h-transform of Doob (see Doob, p. 566), that is, the Markov chain with transition kernel p_k^h given by $p_k^h(x,y)=h(y)(h(x))^{-1}p_k(x,y)$ instead of p_k. This process has in our situation (i.e. $(Z_k^\infty)_{k\in\mathbb{Z}^-}$) exactly the law induced by $(P^\infty)_0$, as is immediately checked writing down the probabilities of a path.

Hence we can explicitly determine the h-transformed transition functions as follows. We write down the calculation for $K=-2$ first. Let $p_{(0,-1,-2)}(z_0,z_{-1},z_{-2})$ denote the joint density of $Z_0^\infty, Z_{-1}^\infty, Z_{-2}^\infty$, etc. and by $p(Z_0|Z_{-1},Z_{-2})$ the conditional density of Z_0^∞ given Z_{-1}^∞, Z_{-2}^∞ etc.
Under the Palm measure $(P^\infty)_0$, the conditional density of Z_{-1}^∞ given Z_{-2}^∞ is then given by

$$e\frac{z_{-1}p_{(-1|-2)}(z_{-1}|z_{-2})}{z_{-2}}.$$

Since conditioned on Z_{-2}^∞, Z_{-1}^∞ has distribution $\frac{d}{c_1}Gamma(\frac{c_1}{d}Z_{-2}^\infty,1)$, this means that under the Palm measure it has distribution $\frac{d}{c_1}Gamma(\frac{c_1}{d}Z_{-2}^\infty+1,1)$. Continuing in this way we get that the transition function $Z_{-k}^\infty\to Z_{-k+1}^\infty$ is given by $\text{Gamma}(\frac{c_{k-1}}{d}Z_{-k}^\infty+1,\frac{c_{k-1}}{d})$.

(ii) Recall from Theorem 8(iii) that the entrance law, conditioned on $\{Z_k^\infty\}_{k\in\mathbb{Z}^-}$, has the property that the $\{\frac{m_{-k}(\cdot)}{Z_{-k}^\infty}\}_{k\in\mathbb{Z}^-}$ are independent and for each k, $\{\frac{m_{-k}(i)}{Z_{-k}^\infty}\}_{i\in\mathcal{I}}$ has the Poisson Dirichlet distribution with parameter $\frac{Z_{-k-1}^\infty c_k}{d}$. Let F_1, F_2 be bounded measureable functions defined on $\sigma(\{Z_k^\infty\}_{k\in\mathbb{Z}^-})$, $\sigma(\{\frac{m_{-k}(\cdot)}{Z_{-k}^\infty}\}_{k\in\mathbb{Z}^-})$, respectively. Then

$$\begin{aligned}(6.13)\qquad E^{(P^\infty)_0}(F_1\cdot F_2) &= E^P(Z_0^\infty F_1\cdot F_2)\\ &= E^P[E^P(F_2|\sigma(\{Z_k^\infty\}_{k\in\mathbb{Z}^-})\cdot Z_0^\infty F_1)]\\ &= E^{(P^\infty)_0}[E^P(F_2|\sigma(\{Z_k^\infty\}_{k\in\mathbb{Z}^-})\cdot F_1)]\end{aligned}$$

where E^P, $E^{(P^\infty)_0}$ denotes expectation with respect to the probability laws P (recall that the processes γ_k are defined on $(\Omega,\mathcal{F},P)$),and $(P^\infty)_0$ respectively. Therefore

$$E^{(P^\infty)_0}(F_2|\sigma(\{Z_k^\infty\}_{k\in\mathbb{Z}^-})=E^P(F_2|\sigma(\{Z_k^\infty\}_{k\in\mathbb{Z}^-})$$

which implies (ii).
(iii) This relation follows from the above facts since the sum of two independent random random variables one exponential with mean $\frac{d}{c_{k-1}}(=\text{Gamma}\,(1,\frac{c_{k-1}}{d}))$ and the other $\text{Gamma}(\frac{c_{k-1}\theta}{d},\ c_{k-1})$ yields a $\text{Gamma}(\frac{c_{k-1}\theta}{d}+1,\frac{c_{k-1}}{d})$ random variable. This completes the proof of the lemma.

Remark In fact we get that under the Palm measure the system of gamma processes have the same parameters but each runs for an additional time interval of length $\frac{d}{c_k}$. If we then choose a sequence of jumps, the first has the exponential distribution and the remainder have the same distribution as in the original process. Therefore we can identify the system under the Palm measure as the sum of two independent random objects, one the family measure given in the theorem and the other is simply the original historical process. Note that this is analogous to what we get from the Palm of the historical process at site $\xi=0$ (cf. Theorem 7(d)).

Proof of Theorem 11 (a)

We have to show that under the Palm measure, $(P^\infty)_0$, the $\{\bar m_k\}_{k\in\mathbb{Z}^-}$ are independent exponential random variables with means $\frac{d}{c_{|k|}}$. By the Lemma 6.1(ii) for $k \in \mathbb{N}$ we have that conditioned on $\{Z^\infty_{-\ell}\}_{\ell\geq k+1}$, Z^∞_{-k} and $\{\frac{m_{-k}(\cdot)}{Z^\infty_{-k}}\}$ are independent and have under the Palm measure $(P^\infty)_0$ distribution $\frac{d}{c_k}\mathrm{Gamma}(\frac{c_k}{d}Z^\infty_{-(k+1)}+1,1)$, respectively, the Poisson Dirichlet distribution with parameter $\frac{Z^\infty_{-k-1}c_k}{d}$. By definition $\bar m_{-k} = m_{-k}(G_{-k}(i))$ where i is the index of a randomly chosen individual at level zero. By Theorem 8(v)

$$P(G_{-k}(i)=j) = \frac{m_{-k}(j)}{Z^\infty_{-k}}.$$

Therefore, by A3.4, $\frac{\bar m_{-k}}{Z^\infty_{-k}}$ has distribution $\mathrm{Beta}(1,\frac{c_k}{d}Z^\infty_{-(k+1)})$ and is independent of Z^∞_{-k} which has distribution $\frac{d}{c_k}\mathrm{Gamma}(\frac{c_k}{d}Z^\infty_{-(k+1)}+1,1)$.

Therefore conditioned on the $\{Z^\infty_{-k}\}$ we obtain using A3.1(2) that the $\bar m_{-k}$ have distributions $\frac{d}{c_k}\mathrm{Gamma}(1,1)$, that is, exponential with means $\frac{d}{c_k}$ and therefore they are also independent of the $\{Z^\infty_{-k}\}$.

Proof of Theorem 11 (b)

By the construction of $M^*_{-k,-\ell}$ preceeding Theorem 11, it follows that $\bar M^*_{-k,-\ell}$ denotes the mass of the descendents at level k of the mass $m_{-\ell}(G_{-\ell}(i)))$. Part (b) then follows from Lemma 6.1, equations (6.12), (6.13) since this says that under the Palm measure the family masses still evolve via the interaction chain.

In addition, by Lemma 6.1(iii) the interaction chain under the Palm measure is a Markov chain with transition function $\widehat\Gamma^{k-1}_\theta$. Then starting at level L we obtain

$$M^*_{-k} = z^L_k(Z^\infty_{-L-1}) + \sum_{\ell=k}^{L} \hat z^\ell_k(E_\ell)$$

where the $z^L, \hat z^\ell$ are independent copies of the Gamma processes (1.39) and the $\{E_\ell)\}$ are independent exponential random variables with means $\frac{d}{c_\ell}$. Taking the limit as $k\to\infty$ gives then the desired decomposition of the entrance law under the Palm measure.

We will denote the law of $\sum_{\ell=0}^\infty \hat z^\ell_0(E_\ell)$ by $(\widehat P^\infty)_0$. The identification of $(\widehat P^\infty)_0$ as the Palm measure of the canonical measure of the infinitely divisible measure P^∞ follows from the general result relating these two Palm measures (see A2.2).

Proof of Theorem 12 (a)

We start by first preparing the basic ingredients to prove this theorem. By Theorem 7(d) the Palm measure, $(P^N)_y$ of the equilibrium historical process $H^{N,-\infty}_0$ at y is the distribution of the sum of two independent random measures on $D((-\infty,\infty),\Omega_N)$ one of these has the same law as $H^{N,-\infty}_0$ and a second which is concentrated on the $y-clan$ and has law $(\widehat P^N)_y$ defined by (1.30).

We now consider the law of the $\bar M^N_{0,-k}(\xi)$ under the Palm distribution. By the defining property of the Palm measure if $B\in\sigma(\{\bar M^N_{0,k}(\xi)\}_{k\in\mathbb{Z}^-})$, then

$$(6.14)\qquad \begin{aligned}(P^N)_{\xi,0}(B) &= \theta^{-1}E[H^{N,-\infty}_0(\{y' : y'(0)=\xi\})1_B(H^{N,-\infty}_0)]\\ &= \mathcal{L}(H^{N,-\infty}_0) * (\int \Pi^\infty_{0,\xi}(dy)(\widehat P^N)_y(B)).\end{aligned}$$

Without loss of generality we can take $\xi=0$ and we will do so below.

Now let P^∞ be the law constructed in our historical representation of the entrance law. Then the modified Palm measure of $\mathcal{H}$ was defined by the analogue of (6.16), namely,

$$(6.15)\qquad (P^\infty)_0(B) := \theta^{-1}\int Z^\infty_0\cdot 1_B(\mathcal{H})P^\infty(d\mathcal{H}).$$

In the following discussion we will consider only the restriction of this measure to $\sigma(\{M_k^{*N}(\cdot)\}_{k\in \mathbb{Z}^-})$.

We will now prove part (c) of Theorem 11 which asserts that

$$(P^N)_{\xi,0} \underset{N\to\infty}{\Longrightarrow} (P^\infty)_0.$$

Recall that the above distributions are infinitely divisible and compare A.2.2. By Theorem 10.4 and Lemma 10.8 in Kallenberg (1976), it suffices to show that the mean measures converge and that the Palm measures of the canonical measures converge, that is,

$$(\widehat{P}^N)_{\xi,0} \underset{N\to\infty}{\Longrightarrow} (\widehat{P}^\infty)_0.$$

Since by construction,

$$E[H^N(\{y' : y'(0) = 0\})] = \theta = E[Z_0^\infty]$$

the convergence of the mean measures is automatic. Therefore we will focus on the convergence of the Palm measures of the canonical measures.

By Theorem 7, part (d) and (1.34) under the Palm distribution the contribution to the family mass of an individual (that is one chosen at random) has Laplace functional

$$(6.16)\qquad (\widehat{P}^N)_y = \prod_k \exp(-2d\int_{-\tau_k^N}^{-\tau_{k-1}^N} (V_{r,0}^N\phi)(y^r)dr)$$

where y denotes the ancestral "backbone" and τ_k^N denotes the exit time of the backbone of the randomly chosen "individual" from the ball of radius k (in reverse time). Therefore this can be represented as the sum of independent contributions, $\bar{M}_{0,-k}^N(0)$, which occur in the time intervals between the first exits of the "backbone" y from the balls of radii $k+1$ and k respectively.

The proof is divided in three steps (i) - (iii) and the strategy of proof is as follows: We will show that the k-th such contribution to the family mass at the fixed site 0, $\bar{M}_{0,-k}^N(\{0\})$, converges in distribution to the distribution of $\bar{M}_{0,-k}^*$. To do this we first show that the (normalized) masses of the families, $\bar{m}_{-k}^N$, produced in the interval $[\tau_{k-1}^N, \tau_k^N)$ along the backbone converge as $N\to\infty$ to independent exponential random variables. We then show that all but a negligible portion of this mass lies at a distance k from 0. Finally we will show that $\bar{M}_{0,-k}^N(0)$, the descendents at level 0 at site 0 of $\bar{m}_{-k}^N$ are described by the quasiequilibria of Theorem 1 and therefore described by the interaction chain in the limit. In order to carry this program out we need information about the limiting behaviour of the random walks as $N\to\infty$, and this will be obtained in the first step below.

Step (i) Associate with the random walk on Ω_N generated by the transition kernel $a(\cdot,\cdot)$ of (0.8) the continuous time Markov chain on Ω_N which gives the distance from the element 0 and then observe it at times of the form $s(N)N^k$ with k running through the natural numbers. (Recall that the transitions within distance k are uniformly distributed) Remember that the random walk on Ω_N with kernel $(a,(\cdot,\cdot)$ takes a jump on a randomly chosen point in a k-block with rate $\frac{c_{k-1}}{N^{k-1}}$. Recall that $\Pi_{0,\xi}^\infty$ denotes the law of the random walk starting at ξ at time 0.

Lemma 6.2.
(i)

$$\frac{c_k}{N^k}\tau_k^N \underset{N\to\infty}{\Longrightarrow} \textit{Exponential(1)},\quad E\Big(\frac{c_k}{N^k}\tau_k^N\Big) \underset{N\to\infty}{\longrightarrow} 1,$$

and

$$\Pi_{0,0}^\infty(d(0,y(\tau_k^N)) = k+1) \underset{N\to\infty}{\longrightarrow} 1.$$

(ii)

$$\frac{\int_0^{\tau_k^N} 1_{(d(y_s,0)<k)}}{\tau_k^N} \underset{N\to\infty}{\longrightarrow} 0 \text{ in probability.}$$

(iii) Define the return times $\sigma_k^N := inf\{t > \tau_{k-1}^N, y_t \in B_{k-1}\}$, *where* $B_k = \{\xi | d(0,\xi) \leq k\}$. *Then*

$$P((\sigma_k^N - \tau_{k-1}^N)/N^k < \frac{\varepsilon}{N^\alpha}) \to 0$$

for any $\alpha > 0$ $\square$.

Proof (i) The random walk exits the ball of radius k by making a jump to $k+\ell$, $\ell \geq 1$, and the rate of such a jump is $\frac{c_{k+\ell-1}}{N^{k+\ell-1}}$. Therefore $\frac{c_k}{N^k}\tau_k^N$ is exponential with parameter $\sum_{\ell \geq k} \frac{c_\ell}{N^\ell}$. By our assumptions, $c_k \leq C2^k$ for some constant C. Therefore

$$\begin{aligned}\frac{c_k}{N^k} \leq \sum_{\ell\geq 1} \frac{c_{k+\ell-1}}{N^{k+\ell-1}} \qquad &\leq \frac{c_k}{N^k} + C\sum_{\ell \geq 1} \frac{2^k+\ell}{N^{k+\ell}} \\ &\leq \frac{c_k}{N^k} + C\frac{2^{k+1}}{N^{k+1}}\frac{1}{1-\frac{2}{N}} \sim \frac{c_k}{N^k} \text{ as } N \to \infty.\end{aligned}$$

This yields the result. In addition the probabilty that the jump is to $k+1$ is bounded below by

$$\frac{\frac{c_k}{N^k}}{\frac{c_k}{N^k} + C\frac{2^{k+1}}{N^{k+1}}\frac{1}{1-\frac{2}{N}}}.$$

(ii) In the N^k time scale the exit time of the ball of radius k, τ_k^N/N^k, is asymptotically exponential with parameter $(c_k)^{-1}$. During this time it can jump to a lower level with rate c_{k-1}. However if it jumps to any lower level $k-\ell$ then it returns to level k in a time of order $O(\frac{1}{N})$.

(iii) This follows since since $(\sigma_k^N - \tau_{k-1}^N)$ dominates an exponential random variable with mean $\frac{c_{k-1}}{N^k}$.

Remark Lemma 6.2 implies that as $N \to \infty$ in the time scale N^k the embedded discrete time Markov chain describing the distance of the walk from 0, in the limit has nontrivial probabilities to leave the point k. Denote with p_k resp. q_k the probabilities to jump from state k to $k+1$ resp. $k-1$. Then $p_k = c_k$ and $q_k = c_{k-1}$ since the probability that a jump $k \to k+m$, resp. $k \to k-m$ with $m > 1$ occurs before a jump to $k \pm 1$ is of order N^{-1}. Therefore the distribution of $\tau_k^N - \tau_{k-1}^N$ is asymptotically exponential with parameter $\frac{1}{c_k}N^k$. Part (ii) of the lemma says that in all but a negligible fraction of the interval $[0, \tau_k^N)$ the backbone is at distance k.

Step (ii) We now determine the distribution of the mass, $\bar{m}_{-k}^N(\cdot)$ at time $-\tau_{k-1}^N$ which arises from the production at rate $2d$ in the interval $[\tau_{k-1}^N, \tau_k^N)$.

Lemma 6.3
Let $B_k := \{\xi' : d(\xi', 0) \leq k\}$. *Then*

$$N^{-k}\bar{m}_{-k}^N(B_k) \underset{N\to\infty}{\Longrightarrow} \bar{m}_{-k} \tag{6.17}$$

$$N^{-(k+1)}\bar{m}_{-k}^N(B_{k+1}) \underset{N\to\infty}{\Longrightarrow} 0 \tag{6.18}$$

$$N^{-k+1}\bar{m}_{-k}^N(B_{k-1}) \underset{N\to\infty}{\Longrightarrow} 0 \qquad \square \tag{6.19}$$

Proof
The proof will proceed in three steps labelled 1-3. First we introduce an approximating system ($N \to \infty$) and then in step 2 calculate corresponding log-Laplace transforms which are then in step 3 used to complete the argument.

Recall the relation (1.30), which tells us that we can express the Laplace transform of $(\widehat{P})_y$ in terms of the object $V_{s,t}$ defined by the historical process of the original system. Therefore we now focus on that system first. Note furthermore that for functionals of the path process which depend only on one time point, the semigroup action of $V_{s,t}$ is given via the action of the semigroup of the interacting system itself.

Step 1
Return now to the system $(x_\xi(t))_{\xi\in\Omega_N}$. The normalized mass in a k-block at time $N^k t$ is given by

$$y_k^N(t) := x_{\xi,k}(N^k t)$$

where $x_{\xi,k}$ is as defined section 0(c). We shall derive now a one dimensional approximating (as $N \to \infty$) diffusion in the case where the initial state satisfies $x_\xi^N(0) = 0$ for $d(\xi,0) \geq k+1$. Adding the corresponding equations and changing to time $N^k t$ in (0.2) (when $g(x) = dx$) yields the equations

$$(6.20) \qquad \begin{aligned} dy_k^N(t) &= -c_k y_k^N(t)dt + \sqrt{2dy_k^N(t)}dw_k(t) \\ &+ c_k y_{k+1}^N(t) + \sum_{\ell=k+2}^{\infty} \frac{c_\ell}{N^{\ell-1+-k}}(y_\ell(t) - y_k^N(t))dt, \quad k \in I\!N. \end{aligned}$$

This describes the flow of mass in and out of a k-block as well as the branching in the k-block.

The following calculations which keep track of the mass which has entered and exited the k-block can be made precise using the historical process or by introducing a multitype modification of the process but we will carry this out at an informal level. In particular we will now show that in order to determine the mass in the k-block it suffices to consider the branching and the flow out of the k-block but that the contribution of the mass that reenters the k-block is negligible.

We first note that if $\sum_{d(\xi',\xi)>k} x_{\xi'}(0) = 0$, then $\zeta_k^N(t) := N^{-k}\sum_{\xi'} x_{\xi'}(N^K t)$ is a critical Feller branching diffusion with generator $d(\frac{\partial}{\partial x})^2$ and hence is a martingale with a law independent of N and therefore

$$P(\sup_{0\leq t\leq T} \zeta_k^N(t) > K) \leq const\ \frac{1}{K}.$$

If initially $\sum_{d(\xi',\xi)>k} x_{\xi'}(0) = 0$, then at later times t we can get the estimate

$$\sum_{d(\xi',\xi)>k} x_{\xi'}(N^k t) \leq e^{c_k t} \sup_{0\leq s\leq t} \zeta_k^N(s),$$

since the migration mechanism is given by a deterministic system of differential equations. Recall that the rate of flow of mass from the complement of a k-block back into the k-block (in time scale $N^k t$) is bounded above by

$$N^k \sum_{\ell=k+1}^{\infty} \frac{c_\ell}{N^\ell} \leq N^k C \sum_{\ell=k+1}^{\infty} \frac{2^\ell}{N^\ell}.$$

Therefore the total mass to first leave the k-block and then return to this block in the time interval $[0, N^k T]$ is bounded above by

$$\frac{const}{N} T e^{c_k T} \cdot \sup_{0\leq t\leq T} \zeta_k^N(t).$$

Thus we have the evolution of the normalized mass in the k-block, in time scale tN^k, which was denoted $y_k^N(t)$ can, if $x_\xi^N(0) = 0$ for $d(0,\xi) \geq k+1$, be written in the form

$$y_k^N(t) = y_k(t) + O(\frac{1}{N}) \tag{6.21}$$

uniformly in $t \in [0,T]$ as $N \to \infty$, where y_k is continuous *subcritical branching diffusion*

$$dy_k(t) = -c_k y_k(t)dt + \sqrt{2dy_k(t)}dw_k(t). \tag{6.22}$$

If we consider our initial states $x_\xi^N(0)$ we know that $x_{\xi,k+1}(tN^k)$ (see Theorem 1) converges to the constant path equal to θ. In this case we approximate by the system $\tilde{y}_k(t)$ given by the *subcritical branching diffusion with immigration*:

$$d\tilde{y}_k(t) = c_k\theta dt - c_k\tilde{y}_k(t) + \sqrt{2d\tilde{y}_k(t)}dw(t). \tag{6.23}$$

Since our system starts in a configuration with $x_\xi^N(0)$ positive on all of Ω_N, we have to study the process given in (6.24) with additional immigration. Due to the branching property of the diffusion $y_k(t)$ this is no problem since the $\tilde{y}_k$ process can be disintegrated into independent contributions following an evolution mechanism as given in (6.23). Therefore it is of particular interest to calculate the probability that $y_k(t) > 0$ as a function of t and of the initial mass. This will be done in the next step.

Step 2
Then we define for $y_k(t)$ log-Laplace function $v_k(t,u)$ by $E[(\exp(-\lambda y_k(t))|y_k(0) = m)] = \exp(-mv_k(t,\lambda))$.

Then the log-Laplace function $v_k(t,u)$ (cf. [D] Section 4.5) satisfies

$$\frac{\partial v_k}{\partial t} = -dv_k^2 - c_k v_k \quad v_k(0) = \lambda. \tag{6.24}$$

This equation has the explicit solution given by:

$$v_k(t,\lambda) = \frac{c_k\lambda e^{-c_k t}}{c_k + \lambda d(1 - e^{-c_k t})}. \tag{6.25}$$

Using (6.27) we get that the probability that an initial mass m will produce a non-zero cluster of age t is

$$\begin{aligned}
(6.26)\qquad & & 1 - \lim_{\lambda\to\infty} E[(\exp(-\lambda y_k(t))|y_k(0) = m)] \\
(6.27)\qquad & = & 1 - \lim_{\lambda\to\infty} \exp(-mv_k(t,\lambda)) \\
& = & 1 - \lim_{\lambda\to\infty} \exp(-\frac{mc_k\lambda e^{-c_k t}}{c_k + \lambda d(1 - e^{-c_k t})}) \\
& = & 1 - \exp\Big(-\frac{mc_k e^{-c_k t}}{d(1-e^{-c_k t})}\Big) \sim \frac{mc_k e^{-c_k t}}{d} \text{ as } t \to \infty.
\end{aligned}$$

As a side remark note that, then (cf. [DP], Prop. 3.3 (b)) the non-zero cluster of age t has Laplace transform

$$\begin{aligned}
(6.28)\qquad & \lim_{m\to 0} \frac{E(e^{-\lambda y_k(t)}|y_k(0) = m) - \exp(-\frac{mc_k e^{-c_k t}}{d(1-e^{-c_k t})})}{1 - \exp(-\frac{mc_k e^{-c_k t}}{d(1-e^{-c_k t})})} \\
= & \frac{c_k}{(c_k + \lambda d(1 - e^{-c_k t}))}.
\end{aligned}$$

Therefore the cluster is exponentially distributed with mean $\frac{d(1-e^{-c_k t})}{c_k}$.

Step 3
We now turn to the Palm distribution and determine the distribution of the normalized

mass at time $-\tau^N_{k-1}$ produced in a k-block whose last common ancestors lie in $[\tau^N_{k-1},\tau^N_k)$. By (6.17) the contribution under the Palm distribution (cf. Theorem 7(d)) to $M^N_{-k}(\xi)$ which arises from the subfamilies that branch off in the interval $[\tau^N_{k-1},\tau^N_k)$ has Laplace transform

$$(6.29)\qquad \exp\Big(-2d\int_{-\tau^N_k/N^k}^{-\tau^N_{k-1}/N^k}(V^N_{r,\tau^N_{k-1}}\phi)(y^r)dr\Big).$$

For special choices of ϕ, we can use the approximation result of (6.31) below. Namely if we let $\phi_k=\frac{1}{N^k}1_{d(\xi',\xi)\le k}$, then since y^s remains in the k-block during this interval we obtain (recall (1.33)) for $0\le a<b$:

$$(6.30)\qquad \begin{aligned}&\exp\Big(-\int_{-a}^{-b}2d(V^N_{s,-b}\lambda\phi_k)(y^s))ds\Big)\\ &\underset{N\to\infty}{\longrightarrow}\exp\Big(-\int_{-a}^{-b}2dv_k(t,\lambda)dt\Big)\\ &=\Big(\frac{c_k}{c_k+\lambda d(1-e^{-c_k(b-a)})}\Big)^2\end{aligned}$$

Observe that the r.h.s. of (6.31) is a bounded continuous function of the variable $(b-a)\in[0,\infty)$. Now by Lemma 6.2 $[\tau^N_k/N^k-\tau^N_{k-1}/N^k]$ is exponential with mean $c_k^{-1}+O(\frac{1}{N})$. Therefore we get from (6.31) with $\phi=\lambda\phi_k$

$$(6.31)\qquad E\Big(\exp\Big(-\int_{-\tau^N_k/N^k}^{-\tau^N_{k-1}/N^k}2d(V^N_{s,-\tau^N_{k-1}/N^k}\lambda\phi_k)(y^s))ds\Big)\Big)$$

$$(6.32)\qquad \begin{aligned}&\underset{N\to\infty}{\longrightarrow}\int_0^\infty c_ke^{-c_kt}\Big(\frac{c_k}{c_k+\lambda d(1-e^{-c_kt})}\Big)^2dt\\ &=\frac{c_k}{c_k+\lambda d}.\end{aligned}$$

That is, we obtain the Laplace transform of an exponential with mean $\frac{d}{c_k}$ thus identical to the distribution of $\bar m_{-k}$ obtained in (a). This completes the proof of (6.19) of Lemma 6.3.

Since $\bar m^N_{-k}(B_{k+1}\cap B_k^c)$ is produced from the mass which emigrates from B_k, it is of order $O(N^k)$ which implies (6.20).

By (6.28) and Lemma 6.2 (iii) any clusters in B_{k-1} produced in the subinterval of $[\tau^N_{k-1},\tau^N_k)$ while the backbone has returned to B_{k-1} has negligible probability $(e^{-c_kN^\alpha})$ (with $0<\alpha<1$) of survival. Moreover by Lemma 6.2(ii) the backbone spends a negligible fraction of the sojourn time in B_{k-1}. Hence applying (6.30) and (6.26) with ϕ_k replaced by ϕ_{k-1} (and noting that in this calculation we are working with the solution of (6.24) with k replaced by $k-1$ but in the time scale N^kt) we get as $N\to\infty$

$$(6.33)\qquad \begin{aligned}&E\Big(\exp\Big(-\int_{-\tau^N_k/N^k}^{-\tau^N_{k-1}/N^k}2d1_{B_{k-1}}(y^s)(V^N_{s,-\tau^N_{k-1}/N^k}\lambda\phi_{k-1})(y^s))ds\Big)\Big)\\ &\sim E\Big(\exp\Big(-\int_{-\tau^N_k/N^k}^{-\tau^N_{k-1}/N^k}2d1_{B_{k-1}}(y^s)\frac{c_{k-1}\lambda e^{-Nc_{k-1}(|s|-\tau^N_{k-1}/N^k)}}{c_{k-1}+\lambda d(1-e^{-Nc_{k-1}(|s|-\tau^N_{k-1}/N^k)})}ds\Big)\Big)\\ &\underset{N\to\infty}{\longrightarrow}0.\end{aligned}$$

The last step follows because $1_{B_{k-1}}(y^s)=0$ for $|s|-\tau^N_{k-1}/N^k\le\tau^*_{k-1}$ where τ^*_{k-1} is the time until the first return to B_{k-1} after τ^N_{k-1} and is exponentially distributed with mean $\frac{1}{c_{k-1}}$ (in the N^k time scale). This completes the proof of (6.20) and of Lemma 6.3.

Step (iii) Recalling the discussion preceding (6.17) we can now complete the proof as follows.

Lemma 6.4

$$(6.34) \qquad (\widehat{P}^N)_{0,0}|_M \underset{N\to\infty}{\Longrightarrow} (\widehat{P}^\infty)_0|_M.$$

where $|_M$ denotes the restriction to the $\sigma\{M^*_{0,-\ell}\}_{\ell\in I\!N}$. □

Proof We must now compute $\bar{M}^N_{0,-k}(\{0\})$, that is, the distribution of the mass of individuals at site 0, due to the family branches produced in the interval $[\tau^N_{k-1}, \tau^N_k)$. By Lemma 6.3, in the $N \to \infty$ limit all but an asymptotically negligible fraction of this mass, $\bar{m}^N_k(B_k)$, belongs to families which broke off while the backbone was at distance k from site 0. Some mass of these families can reach site 0 by having immigrants enter successively smaller balls and then producing subfamilies. Now observe that the branching property implies that systems started in intitial state $\tilde{x}^N(0)$ and $\tilde{\tilde{x}}^N(0)$ and evolving independently can be added to form a version of the system started in $\tilde{x}^N(0)+\tilde{\tilde{x}}^N(0)$. Furthermore the distribution of the ancestral mass over the $(k+1)$-block is not relevant for the probability laws governing the immigration, since the jump distribution to a point depends only on the distance from this point. Therefore using this relation (6.30) we are back into the setup of Theorem 1c and the successive immigrations will be described in terms of the quasiequilibria which in the limit as $N \to \infty$ converge to the interaction chain (cf. 0.13) and (0.14). This has been made precise in (6.22) - (6.24). Therefore with (6.18) the contribution converges (as $N \to \infty$) to $\mathcal{L}(\bar{m}_k(B_k))$ which equals $\mathcal{L}(z^k_0(E_k))$ with E_k being an exponential variable with mean $\frac{d}{c_k}$ thus yielding (6.35).

Proof of Theorem 12(b) Note that (1.66) is a consequence of part (a) since this implies that the Palm measures converge and as pointed out above the equality of the mean measures is automatic. The proof of (1.67) then follows along same lines as the proof of (0.70) in section 5(f) of [DGV] and will not be given in detail here. The main additional step involves second moment bounds uniform in N analogous to (0.68) in [DGV]. The necessary bound is given by (0.30) in Theorem 4(b).

Appendix 1: Tools from the Laplace transformation.

Define (for $b_n \to \infty$ denoting a sequence increasing to ∞) the random variable $\widehat{X}_n$ by

$$\mathcal{L}(X_n/b_n | X_n \geq \varepsilon) = \mathcal{L}(\widehat{X}_n).$$

Assumption (i) $\mathcal{L}(\widehat{X}_n) \underset{n\to\infty}{\Longrightarrow} \mathcal{L}(\widehat{X})$, independent of ε (ii) $P(X_n \geq \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$ for all $\varepsilon > 0$. (iii) $a_n = O(b_n)$.

Let F_n denote the distribution of X_n and let $a_n^{-1} = P(X_n > \varepsilon)$. Then we set

$$L_n(\lambda) = \int_0^\infty e^{-\lambda x} F_n(dx), \qquad L(\lambda) = \int_0^\infty e^{-\lambda x} F(dx).$$

Next calculate as follows:

$$L_n(\lambda) = \int_0^\varepsilon e^{-\lambda x} F_n(dx) + \int_0^\infty e^{-\lambda x} F_n(dx)$$

hence

$$1 - L_n(\lambda) = (1 - \int_0^\varepsilon e^{-\lambda x} F_n(dx)) - \int_\varepsilon^\infty e^{-\lambda x} F_n(dx)$$

and

$$a_n(1 - L_n(\lambda)) = a_n(1 - \int_0^\varepsilon e^{-\lambda x} F_n(dx)) - \int_\varepsilon^\infty e^{-\lambda x} \frac{F_n(dx)}{a_n^{-1}}.$$

Then

$$(*) \quad a_n(1 - L_n(\lambda/b_n)) = a_n(1 - \int_0^\varepsilon e^{-\lambda x/b_n} F_n(dx)) - \int_\varepsilon^\infty e^{-\lambda x/b_n} \frac{dF_n(x)}{a_n^{-1}}.$$

By the assumption we know that

$$\int_\varepsilon^\infty e^{-\lambda x/b_n} \frac{dF_n(x)}{a_n^{-1}} \underset{n\to\infty}{\longrightarrow} L(\lambda).$$

On the other hand rewrite the frist summand on the r.h.s. of (*) as

$$a_n[1 - (1 - a_n^{-1}) \int_0^\varepsilon e^{-\lambda x/b_n} \cdot \frac{dF_n(x)}{1 - a_n^{-1}}]$$

For $b_n \to \infty$ we know that we can rewrite this as

$$a_n[1 - (1 - a_n^{-1}) \int_0^\varepsilon (1 - \lambda x/b_n) \frac{dF_n(x)}{1 - a_n^{-1}}] + O(\frac{a_n}{b_n^2}),$$

Use $a_n = o(b_n^2)$ to get

$$a_n(1 - 1 - \frac{1}{a_n b_n} \lambda \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} + \frac{1}{a_n} + \frac{\lambda}{b_n} \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}}) + o(1)$$

$$= -\frac{1}{b_n} \lambda \cdot \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} + 1 + \frac{\lambda a_n}{b_n} \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} + o(1).$$

Next use $a_n = O(b_n)$ and assumption (ii) to conclude

$$\int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} = E(X_n | X_n \le \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$$

to obtain that the first summand in the r.h.s. of (*) converges to 1 as $n \to \infty$. Hence:

$$a_n(1 - L_n(\lambda/b_n)) \underset{n\to\infty}{\longrightarrow} 1 - L(\lambda), \; L(\lambda) = \int e^{-\lambda x} F(dx).$$

Assumption (i) $a_n(1 - L_n(\lambda/b_n)) \underset{n\to\infty}{\longrightarrow} 1 - L(\lambda)$ (ii) $L(\lambda) = \int e^{-\lambda x} F(dx)$. (iii) $a_n \to \infty$, $b_n \to \infty$, $a_n/b_n \underset{n\to\infty}{\longrightarrow} 1$.

Use the same decomposition as in the argument above to get via (i).

$$-\frac{1}{b_n} \lambda \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} + 1 + \lambda \frac{a_n}{b_n} \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} - \int_\varepsilon^\infty e^{-\lambda x/b_n} \frac{dF_n(x)}{a_n^{-1}} \underset{n\to\infty}{\longrightarrow} 1 - L(\lambda)$$

Since $L(\lambda)$ is a Laplace transform i.e. $L(\lambda) \underset{\lambda\to\infty}{\longrightarrow} 0$ we know $\frac{a_n}{b_n} \int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} \underset{n\to\infty}{\longrightarrow} 0$ and hence

$$\int_0^\varepsilon x \frac{dF_n(x)}{1 - a_n^{-1}} \underset{n\to\infty}{\longrightarrow} 0 \text{ i.e. } F_n(x) \Rightarrow \delta_0.$$

Then we must have

$$\int_\varepsilon^\infty e^{-\lambda x/b_n} \frac{dF_n(x)}{a_n^{-1}} \underset{n\to\infty}{\longrightarrow} L(\lambda), \; \int_\varepsilon^\infty \frac{dF_n(x)}{a_n^{-1}} \underset{n\to\infty}{\longrightarrow} 1.$$

This means for $\mathcal{L}(\widehat{X}) = F(\cdot)$,

$$\mathcal{L}(\widehat{X}_n) \underset{n\to\infty}{\longrightarrow} \mathcal{L}(\widehat{X}), \; a_n^{-1} \underset{n\to\infty}{\sim} \text{Prob } (X_n \ge \varepsilon).$$

Appendix 2: Some tools on infinitely divisible random measures

A2.1 Canonical Representation of infinitely divisible random measures and processes

Let E be a Polish space and let $M_{LF}(E)$ denote the space of locally finite (nonnegative) measures on E. An infinitely divisible random measure on E with no deterministic component has a *canonical representation* such that

$$-\log E(e^{-<X,\phi>}) = \int_{M_{LF}(E)} (1 - e^{-<\nu,\phi>}) R(d\nu)$$

where the *canonical* or *Lévy* measure R is a measure on $M_{LF}(E)$ satisfying

$$\int (1 \wedge \mu(A)) R(d\mu) < \infty$$

for every bounded set A.

A2.2 Campbell measures and Palm distributions

Let P be the law of a random measure on the Polish space E with locally finite intensity (denoted $\mathcal{M}_{LF}(E)$) Define the intensity measure I by $I(B) := \int \mu(B) P(d\mu)$. The associated *Campbell measure* $\bar{P}$ is a measure on $\mathcal{M}_{LF}(E) \times E$.

$$\bar{P}(B \times A) = \int_B \mu(A) P(d\mu).$$

The associated Palm distributions $\{(P)_x : x \in E\}$ are a collection of measures on the Borel-σ-algebra of $\mathcal{M}_{LF}(E)$ such that

(i) $x \longrightarrow (P)_x(B)$ is $\mathcal{B}(E)$-measurable for all sets $B \in \mathcal{B}(\mathcal{M}_{LF}(E))$.

(ii) For every $x \in E : (P)_x(\cdot)$ is a probability measure on $(\mathcal{M}_{LF}(E),\ \mathcal{B}(\mathcal{M}_{LF}(E)))$.

(iii) For every measurable bounded function g on $\mathcal{M}_{LF}(E) \times E$, the following holds

$$\int I(dx) \int (P)_x(d\mu) g(\mu, x) = \int \bar{P}(d\mu, dx) g(\mu, x).$$

If the random measure, X, is infinitely divisible law P with canonical measure R, then the Palm distributions of X and R are related as follows (see Kallenberg (1976), Lemma 10.6)):

$$(P)_x = P * (R)_x.$$

For example, in the case $E = I\!N$, and $Z = \sum Z_i \delta_i$, using the representation of the Laplace transform with the canonical measure we can write:

$$E(\exp(-\sum \lambda_i Z_i)) = \exp\Big(- \int (1 - e^{(\sum \lambda_i \mu(i))}) R(d\mu)\Big).$$

Then

$$\begin{aligned}
\frac{E(Z_0 \exp(-\sum \lambda_i Z_i))}{E[Z_0]} &= \frac{1}{E[Z_0]} \frac{\partial}{\partial \lambda_0} E(Z_0 \exp(-\sum \lambda_i Z_i))\Big|_{\lambda_0 = 0} \\
&= \int \exp\big(-\sum \lambda_i \mu(i)\big) \frac{\mu(0) R(d\mu)}{E[Z_0]} \\
&\cdot \quad \exp\Big(- \int (1 - e^{(\sum \lambda_i \mu(i))}) R(d\mu)\Big)
\end{aligned}$$

and the last expression is the Laplace functional of $\frac{(\mu(0)R(d\mu))}{\int \mu(0)R(d\mu)} * \mathcal{L}(Z)$.

A2.3 Moran Gamma Process

The Moran-Gamma-process was defined in (1.38) and (1.39) as a specific nondecreasing

jump process with independent stationary increments. We will view this process here as a random measure on $[0,\infty)$ as follows:
Let G be an infinitely divisible locally finite random measure on $[0,\infty)$ with canonical (Lévy) measure:

$$R(\{u\delta_x : u \in (u_1,u_2], x \in A\}) = |A| \int_{u_1}^{u_2} \frac{e^{-u}}{u} du$$

where $|A|$ is the Lebesgue measure of A. Consider the law, P, of the random measure G. Then the Palm distribution $(P)_x$ can be calculated as follows:

$$(P)_x = P * Q_x$$

where $Q_x(\{\mu = u\delta_x : u \in B)\} = \int_B e^{-u} du = (R)_x$ for $B \in \mathcal{B}(\mathbb{R}^+)$.
Then the infinitely divisible process defined by

$$\gamma(u) := G([0,u])$$

is called the *standard Moran Gamma process*. Note that the random variable $X := \gamma(m)$ is $Gamma(m,1)$ distributed (see A.3.1).

Appendix 3: Some tools on sampling systems

A3.1. Beta distribution and its relation to the Gamma Distribution
The Beta(ℓ,m) distribution on $[0,1]$ has probability density function

$$B(\ell,m)^{-1}x^{\ell-1}(1-x)^{m-1}$$

where

$$B(\ell,m) = \frac{\Gamma(\ell)\Gamma(m)}{\Gamma(\ell+m)}$$

The Gamma (ℓ,m) distribution on $[0,\infty)$ has density

$$\frac{m^\ell}{\Gamma(\ell)} x^{\ell-1}\exp(-mx).$$

Also for this distribution

$$E[X] = \frac{\ell}{m}, \quad Var[X] = \frac{\ell}{m^2}.$$

Basic Facts
(1) Let X and Y be independent random variables with Gamma distributions $\Gamma(\ell,1)$ and $\Gamma(m,1)$. Then $\frac{X}{X+Y}$ has the Beta(ℓ,m) distribution.
(2) Let U,V be independent random variables, U have distribution $B(\ell,m)$ and V have distribution Gamma $(\ell+m,1)$. Then $Z = UV$ has distribution Gamma (ℓ). (See Moran p. 330).

A3.2 Dirichlet Distributions The k-variate Dirichlet distribution with parameters $\alpha_1,\dots,\alpha_{j+1}$ is a distribution on the simplex $\{(x_1,\dots,x_k) : x_i \geq 0, i = 1,\dots,k, \sum_{i=1}^k x_i \leq 1\}$ with joint probability density function:

$$f(x_1,\dots,x_k) = \frac{\Gamma(\alpha_1+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})} x_1^{\alpha_1-1}\cdots x_k^{\alpha_k-1}(1-x_1-\cdots-x_k)^{\alpha_{k+1}-1}.$$

Basic Fact
If $X_1, \ldots, X_{k+1}$ are independent Gamma$(\alpha_1, 1), \ldots,$ Gamma $(\alpha_{k+1}, 1)$ then

$$Y_i := \frac{X_i}{X_1 + \cdots + X_{k+1}}, \quad i = 1, \ldots, k$$

is k-variate Dirichlet with parameters $(\alpha_1, \ldots, \alpha_k, \alpha_{k+1})$ and independent of $\sum_{j=1}^{k+1} X_j$.

A3.3 Poisson Dirichlet Distributions
Assume that $(X_1^{(n)}, \ldots, X_k^{(n)})$ is k-variate Dirichlet with parameters $(\alpha_1^{(n)}, \ldots, \alpha_k^{(n)}, \alpha_{k+1}^{(n)})$. Let $(X_{[1]}^{(n)}, X_{[2]}^{(n)}, \ldots)$ denote the decreasing order statistics of $(X_1^{(n)}, \ldots, X_k^{(n)})$. Assume that

$$\max(\alpha_1^{(n)}, \ldots, \alpha_n^{(n)}) \to 0$$

and that

$$\lambda^{(n)} = \alpha_1^{(n)} + \cdots + \alpha_n^{(n)} \to \lambda \quad \text{as } n \to \infty.$$

Then the joint distribution of $(X_{[1]}^{(n)}, X_{[2]}^{(n)}, \ldots)$ converges to the *Poisson Dirichlet distribution* with parameter λ.
On the other hand if $X_{[1]}^{(\infty)}, X_{[2]}^{(\infty)}, \ldots$ has a Poisson Dirichlet distribution with parameter λ and Y_i if a sequence of i.i.d. $\{1, \ldots, m\}$-valued random variables with distribution $p_1, \ldots, p_m$, then

$$\{\sum_{i=1}^{\infty} X_{[i]}^{(\infty)} 1_j(Y_i)\}_{j=1\ldots,m}$$

has a Dirichlet distribution with parameters $\lambda p_1, \ldots, \lambda p_m$.

A key tool in relating the branching and Fleming-Viot systems is the relation between the standard Moran gamma process and the Poisson Dirichlet distribution.

Basic Fact
Denote by $\{\Delta\gamma(u)\}_{u \in [0,\lambda]}$ the size-ordered jumps of the nondecreasing process γ in the time interval $[0, u]$. The distribution of the normalized order jumps $\{\frac{\Delta\gamma(u)}{\gamma(\lambda)}\}_{u \in [0,\lambda]}$ of the Moran Gamma process has the Poisson Dirichlet distribution with parameter λ and it is independent of $\gamma(\lambda)$. (See Kingman (1993), Chapter 9).

A3.4 Size-biased sampling
Let $\{p_k\}_{k \in \mathbb{N}}$ be a random probability vector with the Poisson Dirichlet distribution with parameter λ. Now let N be a random natural number satisfying

$$P(N = k) = p_k, \quad \forall k.$$

Then the probability mass, $q_1 := p_N$ of the chosen point has the Beta $(1, \lambda)$ distribution with probability density function

$$\lambda(1 - x)^{\lambda - 1}, \quad 0 \leq x \leq 1$$

If we chose a sequence $q_1, q_2, \ldots$ in this way ("without replacement") then this sequence has the *GEM representation*:

$$q_1 = v_1, q_2 = (1 - v_1)v_2, \ldots$$

where the v_i are i.i.d. Beta$(1, \lambda)$ random variables.

Appendix 4: The hierarchical random walk

We consider the discrete time hierarchical random walk on Ω_N. The transition probability $\xi \to \xi'$, $Q(\xi - \xi')$, is given by $\frac{p_j}{N^j}$ where $j = d(\xi, \xi')$ and p_j resp. $\tilde{c}_j$ is defined by

$$\frac{p_j}{N^j} = \frac{\tilde{c}_j}{N^{2j}} = \sum_{k=j}^{\infty} \frac{c_{k-1}}{N^{2k-1}} \text{ if } j = d(\xi, \xi') \geq 1$$

Introduce the following transform in the time variable

$$nH(s, \xi) = \sum_{n=1}^{\infty} s^n Q_n(\xi) = -N^{-k} \frac{sf_k^2}{(1 - sf_k^2)} + \sum_{j=k+1}^{\infty} \frac{(N-1)}{N^j} \frac{sf_j^2}{(1 - sf_j^2)}$$

where $k = d(\xi, 0)$ and

$$\begin{aligned} f_k &= p_1 + \ldots + p_{k-1} - \frac{p_k}{N-1} \\ p_k &= \frac{\frac{\tilde{c}_k}{N^k}}{\sum_{k=1}^{\infty} \frac{\tilde{c}_k}{N^k}} \\ \tilde{c}_m &= \sum_{k=m}^{\infty} c_{k-1} N^{-(2(k-m)-1)} \geq N c_m \end{aligned}$$

Sawyer-Felsenstein (1983) proved that the random walk is transient if

$$\sum_{1}^{\infty} \frac{1}{N^{j-1}} \frac{1}{1 - f_j} < \infty.$$

But

$$1 - f_j = \frac{p_j}{N} + \sum_{k=j}^{\infty} p_k \geq const \cdot \sum_{k=j}^{\infty} \frac{\tilde{c}_k}{N^k} \geq const \cdot \frac{N c_{j-1}}{N^j}.$$

Therefore

$$\sum_{j=1}^{\infty} \frac{1}{N^{j-1}} \frac{1}{1 - f_j} \leq const \cdot \sum_{j=1}^{\infty} \frac{1}{c_j}.$$

Moreover the mean number of visits to 0 starting at ξ is given by

$$H(1, \xi) = \sum_{n=1}^{\infty} Q_n(\xi) \leq \sum_{j=k+1}^{\infty} \frac{1}{N^{j-1}} \frac{1}{(1 - f_j)} \leq \text{const} \cdot \sum_{j=d(0,\xi)}^{\infty} \frac{1}{c_j}$$

and this is uniform in N.

References

[BCGH1-2] J. Baillon, P. Clement, A. Greven, F. den Hollander (1995). 1. On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions: The compact case. Canadian Journal of Mathematics 47, 3-27. 2. The noncompact case. Manuscript 1995.

[CFG] J.T. Cox, K. Fleischmann, A. Greven (1996). Comparison of interacting diffusions and an application to their ergodic theory. To appear Probab. Theory Rel. Fields.

[CGS] J.T. Cox, A. Greven (1995). T. Shiga, Finite and infinite systems of interacting diffusions. Probab. Theory Rel. Fields.

[D] D.A. Dawson (1993). Measure-valued Markov Processes. In: École d'Été de Probabilités de Saint Flour XXI, Lecture Notes in Mathematics 1541, 1-261, Springer-Verlag.

[DG1] D.A. Dawson and A. Greven (1993a). Multiple time scale analysis of hierarchically interacting systems. In: A Festschrift to honor G. Kallianpur, 41-50, Springer-Verlag.

[DG2] D.A. Dawson and A. Greven (1993b). Multiple time scale analysis of interacting diffusions. Probab. Theory Rel. Fields 95, 467-508.

[DG3] D.A. Dawson and A. Greven (1993c). Hierarchical models of interacting diffusions: multiple time scale phenomena. Phase transition and pattern of cluster-formation. Probab. Theory Rel. Fields, 96, 435-473.

[DM] D.A. Dawson and P. March (1995). Resolvent estimates for Fleming-Viot operators and uniqueness of solutions to related martingale problems, J. Funct. Anal. 132, 417-472.

[DGV] D.A. Dawson, A. Greven, J. Vaillancourt (1995). Equilibria and Quasiequilibria for Infinite Collections of Interacting Fleming-Viot processes. Transactions of the American Math. Society, volume 347, no. 7, 2277-2360.

[DP] D.A. Dawson and E.A. Perkins (1991). Historical Processes, Memoirs of the A.M.S. 454.

[DK] P. Donnelly and T. G. Kurtz (1991). A countable representation of the Fleming-Viot measure-valued diffusion, preprint.

[DO] J.L. Doob (1984). Classical Potential Theory and Its Probabilistic Counterpart, Springer-Verlag.

[DU 1] R. Durrett (1979). An infinite particle system with additive interactions, Adv. Appl. Probab. 11, 355-383.

[DU 2] R. Durrett (1993). Ten Lectures on Particle Systems, In: École d'Été de Probabilités de Saint Flour XXIII, Lecture Notes in Mathematics , Springer-Verlag.

[DN] R. Durrett and C. Neuhauser (1993). Particle systems and reaction diffusion equations, Ann. Probab., to appear.

[F] W. Feller (1992). An Introduction to Probability Theory and its Applications Vol. II, Wiley & Sons.

[Fl] J. Fleischmann (1978). Limiting distributions for branching random fields, Trans. Amer. Math. Soc. 239, 353-389.

[FG] K. Fleischmann and A. Greven (1994). Diffusive clustering in an infinite system of hierarchically interacting diffusions. Probab. Theory Rel. Fields 1994, 98, 517-566.

[G] G. Gauthier (1994). Multilevel systems of bilinear stochastic differential equations. Preprint: Technical Report Series of the Laboratory for Research in Statistics and Probability, No. **254**

[GW] L.G. Gorostiza and A. Wakolbinger (1992) Convergence to equilibrium of critical branching particle systems and superprocesses and related nonlinear partial differential equations, Acta Appl. Math. 27, 269-291.

[J] P. Jagers (1975). Branching Processes with Biological Applications. J. Wiley.

[Kal] O. Kallenberg (1976). Random Measures, Akademie-Verlag and Academic Press.

[KI] J.F.C. Kingman (1980). Mathematics of Genetic Diversity, SIAM, Philadelphia.

[KII] J.F.C. Kingman (1993) Poisson Processes, Clarendon Press, Oxford.

[K] A. Klenke (1995). Different clustering regimes in systems of hierarchically interacting diffusions. To appear Ann. of Probability.

[LN] J. Lamperti and P. Ney (1968). Conditioned branching processes and their limiting diffusions, TPA 13, 128-139.

[MO] P.A.P. Moran (1968). An Introduction to Probability Theory, Clarendon Press, Oxford.

[SF] S. Sawyer and J. Felsenstein (1983). Isolation by distance in a hierarchically clustered population, J. Appl. Prob. 20, 1-10.

[SS] K. Shiga and A. Shimizu (1980). Infinite dimensional stochastic differential equations and their applications. J. Math. Kyoto Univ., 20, 395-415.

[S] T. Shiga (1992). Ergodic theorems and exponential decay of sample paths for certain interacting diffusion systems, Osaka J. Math. 29, 789-807.