Probab. Theory Relat. Fields 96:435–473 (1993)

Probability Theory and Related Fields

Hierarchical models of interacting diffusions: Multiple time scale phenomena, phase transition and pattern of cluster-formation

Donald A. Dawson[1] **and Andreas Greven**[2]

[1] Department of Mathematics, Carleton University, Ottawa, Canada K1S 5B6
[2] Institut für Mathematische Stochastik, Lotzestrasse 13, D-37083 Göttingen, Germany

Received October 28, 1991; in revised form March 4, 1993

Summary. The purpose of this paper is to explore the connection between multiple space-time scale behaviour for block averages and phase transitions, respectively formation of clusters, in infinite systems with locally interacting components. The essential object is the associated Markov chain which describes the joint distribution of the block averages at different time scales. A fixed-point and stability property of a particular dynamical system under a renormalisation procedure is used to explain this pattern of cluster formation and the fact that the longtime behaviour is universal in entire classes of evolutions.

Mathematics Subject Classification (1991): 60K35

0 Introduction

It is well-known that infinite systems with locally interacting components, in the sense of *infinite particle systems* or of *interacting diffusions* exhibit phenomena such as *phase transition* and *cluster formation* and both these properties are expected to be of a similar nature in a whole *universality class* of evolutions, (see [12]).

The purpose of this paper is to derive in a simplified situation a scenario which explains this behaviour by connecting three phenomena: multiple space-time scale behavior, appearance of slowly varying functionals of the process and the universality of the long term properties of the system. In particular the universality properties are related to a *fixed-point property* under a *renormalisation* procedure. The unique evolution corresponding to this fixed-point is also distinguished by having a simple *dual process* so that explicit calculations are possible in this case which lead to the determination of the behaviour of the system. Unfortunately all this can be done only by simplifying the interaction geometry and by passing to the mean-field limit of the interaction which is obtained by letting a certain parameter, N, tend to infinity. We shall however show in a precise sense that the mean-field limit predictions are correct for several aspects of the qualitative behaviour of the real system. Furthermore the models we cover are of interest in genetics [13, 14], in particular the results will produce important tools for the analysis of new types of genetic models to be treated in a forthcoming paper.

Let us give a heuristic preview of the scenario we have in mind. Consider the voter model on $\mathbf{Z}^d$, that is, the Markov process with state space $\{0,1\}^{Z^d}$ where the value of each site flips at a rate proportional to the number of neighbouring sites with a different value (see [12] for details). Consider the following spatial renormalization procedure. Combine sites (components) in blocks of $(2N+1)^d$-components, then group these blocks in groups of $(2N+1)^d$ blocks, etc. The state of each block is replaced by the average over the components of the block. During the evolution for a certain interval of time depending on N the block-average will remain almost constant, and the single components of the block will stabilize in some *quasi equilibrium*. However over larger time intervals the blocks will start to interact and only the group average over blocks will remain almost constant. On the level of single components a new quasi equilibrium, characteristic for the average of the block of size $(2N+1)^{2d}$, will develop. Since very large blocks maintain for a long time the initial density of 1's say θ', we expect that if the influence of components far apart is strong enough, or in other words, the influence of the average in large blocks is strong, then we shall approach an equilibrium state with density θ'. We call such a system *stable*. If the interaction is weak, then larger and larger blocks will become all 0 or all 1 and the whole system will converge weakly to a mixture of δ-measures on the two traps, corresponding to all components being 0, or all components being 1. This we call *clustering*. In the simple symmetric voter model the two cases correspond to $d \geqq 3$ and $d=1, 2$, respectively. In the case of clustering one has three regimes: large clusters, diffusive clustering and small clusters. The first two correspond to the behaviour of the simple symmetric voter model in dimension 1 respectively 2. See also [7, 12]. (The third regime for the voter model does not arise in the simple symmetric case.)

This picture of evolution displaying a hierarchy of *separated* time scales for the effective dynamics of *averages over blocks on different levels*, is in general difficult to prove for classes of lattice models and for fixed range of interaction. We shall therefore consider *hierarchical models* with infinitely many components in the limit of large interaction range so that we can verify above scenario rigorously.

The scenario we described for the voter model relates the interaction of the averages over blocks of sizes growing in time to the existence of a whole set of invariant measures. An important aspect of our analysis is that, by passing to the limit of large interaction range, we can associate in a rigorous fashion with our process a sequence of Markov chains, the so-called *interaction chains* with state space (in our case) [0, 1] (which is related to the conserved quantity). This interaction chain describes how larger blocks influence smaller ones and the time index of the chain "corresponds" to the block-size and the state to the block-average over the components. Stability and clustering properties of our process relate to properties of this Markov chain, such as existence of nontrivial entrance laws (stability). In addition the cluster formation and growth of clusters can be studied by looking at a scaling (in time) limit of this Markov chain. The special role of systems of interacting Fisher-Wright diffusions is then reflected by the fact that the associated Markov chains have beta distributions as transition kernels. Other situations in which the density over the components is not the only relevant parameter would require an associated chain with a more complex state space and transition mechanism. Nevertheless we believe the *interaction chain* provides insight into the longterm behaviour in general.

The importance of the above scenario also lies in the fact that it is intimately related with the *universality* of the dichotomy between stability versus clustering, described above in the case of the voter model, for a whole class of evolutions. This universality also holds for such phenomena as cluster formation at various scales (as first found in the 2-dimensional voter model), or the invariance principle (that is large clusters) found first for cluster formation in the 1-dimensional voter model (see Cox and Griffeath [8], Arratia [2]).

The idea is the following. Aggregating over more and more components and passing to larger and larger time scales will wash out more and more the specifics of the local evolution. Suppose now that the evolution of the blocks rescaled in time would be the very same at all levels in the case of some particular evolution mechanism. In this case we would expect the behaviour of this system to predict the behaviour of all systems which "converge" to this "fixed-point" evolution after aggregating and rescaling often enough. The analysis of this fixed-point evolution turns out to be easier due to the fact that *duality relations* replace asymptotic relations in this case and hierarchies of equations do decouple (instead of only in the limit ($t \to \infty$)).

We shall establish that it is precisely the systems of interacting hierarchical Fisher-Wright diffusions which have this fixed-point property in the mean-field limit. Indeed the dichotomy of stability versus clustering and the pattern of cluster-formation depend only on the interaction strength and not on any other specifics of the evolution, as long as we are in the domain of attraction of the fixed-point.

We treat interacting diffusions rather than infinite particle systems for technical reasons which will become apparent in the proofs. One reason is that the behaviour of blocks in the voter model in rescaled time leads to Fisher-Wright diffusions (see [6]), while the same procedure applied to interacting Fisher-Wright diffusions again produces Fisher-Wright diffusions. This hereditary property is of course very useful.

The organization of the paper is as follows: Section 1(a) introduces the model, 1(b) formulates the multiple time scale behaviour and the fixed-point property of Fisher-Wright systems, 1(c) states the results on stability and clustering and patterns of cluster-formation while 1(d) states the results on stability and clustering without passing to the mean-field limit. The Sects. 2–6 contain the proofs of the Theorems 1–7.

1 The main results

(a) The model

(i) We begin by defining the *hierarchical group*, which will play the role which $\mathbf{Z}^d$ plays for interacting particle systems or interacting diffusions indexed by lattice sites. This is a much more natural setup for genetics models, but the reader should keep in mind that our model is also designed to provide a good caricature of lattice models on $\mathbf{Z}^2$. (To develop the analogy, think of $\mathbf{Z}^2$ divided into squares of size N, groups of squares of size N etc. Every point could then be localized by a sequence of squares in which it is contained plus the final information where it is sitting in the $N \times N$ square.) In this spirit we define for every $N=2,3,4\ldots$ (denoting with $\mathbf{N}=\{0,1,2,\ldots\}$):

(1.1) $\Omega^N=\{(\xi_k)\colon k\in\mathbf{N},\ \xi_k\in\mathbf{N}, 0\leqq\xi_k\leqq N-1, \xi_j=0\ \forall j\geqq k_0 \text{ for some } k_0\in\mathbf{N}\}.$

(1.2) $$\Omega^\infty=\bigcup_N \Omega^N.$$

The set Ω^N furnished with the additive structure defined by componentwise addition modulo N forms a countable abelian group. Ω^∞ is a countable abelian group with componentwise addition.

Remark. Every element in Ω^N is contained in Ω^{N+k} for $k=1,2,\ldots$ and in Ω^∞.

In the sequel we denote by $\tilde{\xi}=(0,0,\ldots)$ the canonical tagged site (analogous to choice of the site 0 in $\mathbf{Z}^d$ as the natural reference point). On the set Ω^∞ (and by restriction on Ω^N) we define a metric $d(\xi,\xi')$,

(1.3) $$d(\xi,\xi')=\min(k\,|\,\xi_j=\xi'_j\ \ \forall j\geqq k).$$

The next object we shall need is the block of size r containing a given site ξ:

(1.4) $$\xi_N(r)=\{\xi'\in\Omega^N\colon d(\xi,\xi')\leqq r\} \qquad (\xi(0)=\xi).$$

(ii) For every fixed value of $N\geqq 2$ we define a stochastic process $X^N(t)=\{x_{\xi,k}(t),\ \xi\in\Omega^N,\ k\in\mathbf{N}\}$ with values in $[0,1]^{\Omega^N\times\mathbf{N}}$ by the following infinite system of stochastic differential equations and relations:

(1.5) $$dx_{\xi,0}(t)=\left(\sum_{k=1}^{\infty}\frac{c_{k-1}}{N^{k-1}}(x_{\xi,k}(t)-x_{\xi,0}(t))\right)dt+\sqrt{2g(x_{\xi,0}(t))}\,dw_\xi(t)$$

$$x_{\xi,k}(t)=\frac{1}{N^k}\Big(\sum_{\xi'\in\xi_N(k)}x_{\xi',0}(t)\Big), \qquad k=1,2,\ldots$$

$\mathscr{L}((x_{\xi,0}(0))_{\xi\in\Omega^N})=\mu$, with μ homogeneous product measure and $E(x_{\xi,0}(0))=\theta'$.

The three ingredients $w_\xi(t)$, $g(\cdot)$, $(c_k)_{k\in\mathbf{N}}$ are as follows:

(a) $\{(w_\xi(t))_{t\in\mathbf{R}^+}\}_{\xi\in\Omega^\infty}$ is an i.i.d. collection of standard Wiener processes.
(b) $g\colon[0,1]\to\mathbf{R}^+$ is Lipschitz continuous, $g(x)>0$ for $x\in(0,1)$ and $g(0)=g(1)=0$.
(c) The c_k are strictly positive numbers with $\sum_k c_k N^{-k}<\infty$.

Using a result of Shiga [14] it follows that the system (1.5) has a *unique strong solution.*

Remark. As initial distribution we could consider in our setting any homogeneous ergodic measure on $[0,1]^{\Omega_N}$, say μ_N. In order to be able to discuss the case $N\to\infty$, we then need $\mu_N\Rightarrow\mu$, with μ homogeneous ergodic on $[0,1)^{\Omega_\infty}$. To keep notation to a minimum we restrict our attention to the case of product measure.

Two further remarks are included to mention some applications and extensions of this model:

Remark. In population genetics hierarchical systems arise naturally. For example, ξ could represent an individual, $\xi(1)$ it's family, $\xi(2)$ it's clan, $\xi(3)$ it's village ..., etc. The functions g of interest in modelling the evolution of gene frequencies in time are $g(x)=\text{const}\cdot x(1-x)$ (resampling) and $g(x)=(x(1-x))^2$ (random selection). The c_k represent the intensity of emigration – immigration. For such applications compare [13].

Remark. If in the first line of (1.5) we take $c_k=N^{\varepsilon k}$ with $\varepsilon=\frac{\alpha}{d}-1$, with $\varepsilon \neq 0$ we obtain the caricature of a α-stable d-dimensional lattice system. In particular if $d=1$ or 2 then the system we consider is a caricature of systems with motion mechanism on the borderline of recurrence and transience. Depending on ($\varepsilon>0$, <0, respectively) we are in one of these regimes.

(b) Multiple time scale behaviour, renormalisation, fixed-point analysis

In this subsection we shall formulate in steps (i) and (ii) the multiple time scale behavior of our system. This will provide a setting for a the next step (iii) in which we introduce a renormalization procedure whose unique fixed point is determined and analysed from the point of view of stability.

(i) In the case of hierarchical systems in the limit where the range of interaction tends to infinity (i.e. mean-field limit, $N\to\infty$), we are able to analyse rigorously in what time scales the averages over the different sized blocks fluctuate and what limiting evolutions arise. In the case of systems with N fixed or in lattice systems, it is more difficult to formulate rigorously the scenario given in the introduction since the time scales for the different levels do *not separate*, that is they are of comparable orders of magnitude.

We begin by introducing the ingredients necessary to formulate the statements on multiple time scale behaviour and which are central notions for the whole paper.

In particular it will turn out in Theorem 1 that the longterm behaviour of the system in the mean-field limit $N\to\infty$ can be described in terms of the properties of Markov chains $(Z_k^j)_{k=0,\dots,j}$ with state space $[0,1]$, see III and IV below. We believe that these Markov chains are the *central objects in developing an understanding of the longterm behaviour of large interacting systems.* One special feature of our model is that the associated Markov chain has a simple state space.

The necessary ingredients are the following:

I. *Time scales* $\beta_k(N)=N^k$, $k=1,2,\dots$

II. The *quasi-equilibrium* on level k, $\Gamma_{\theta'}^k$, the *associated diffusion* in eqilibrium on level, k, $(Z_t^{k,\theta'})_{t\in\mathbf{R}^+}$, and finally the *diffusion function* in equilibrium F_k which governs the fluctuation of averages on the level k:

(1.6) $\Gamma_{\theta'}^k(\cdot)$ is the unique equilibrium of the diffusion $(Z_t^{k,\theta'})_{t\in\mathbf{R}^+}$, $k\in\mathbf{N}$, $\theta'\in[0,1]$,

$$dZ_t^{k,\theta'}=c_k(\theta'-Z_t^{k,\theta'})\,dt+\sqrt{2F_k(Z_t^{k,\theta'})}\,dw(t),\ k\in\mathbf{N},\qquad \theta'\in[0,1], \tag{1.7}$$

with $\mathscr{L}(Z_0^{k,\theta'})=\Gamma_{\theta'}^k$. Write $\tilde{Z}_t^{k,\theta'}$ if $Z_0^{k,\theta'}=\theta'$.

$$F_0(x)=g(x),\qquad F_{k+1}(x)=E^{\Gamma_x^k}(F_k(\cdot)),\qquad x\in[0,1],\qquad k\in\mathbf{N}. \tag{1.8}$$

III. Time-inhomogeneous Markov chains $(Z_n^j)_{n=-1,\dots,j}$ with state space $[0,1]$ and transition kernel $\Gamma_x^{j-n}(dy)$ at time $n-1$, starting in θ' at time $n=-1$. We call $(Z_n^j)_{n=-1,\dots,j}$ the *interaction chain* corresponding to level j.
IV. The one dimensional *marginals* $\{\mu_{\theta'}^{j,j-k}(\cdot)\}_{k=0,\dots,j}$ of the interaction chain $(Z_k^j)_{k=-1,\dots,j}$ are of particular importance. They are given by $\Gamma_{\theta'}^j(\cdot)$ for $k=0$ and for $1\leq k\leq j$ by:

$$\mu_{\theta'}^{j,j-k}(\cdot)=\int_{[0,1]}\dots\int_{[0,1]}\Gamma_{\theta'}^j(d\theta_j)\,\Gamma_{\theta_j}^{j-1}(d\theta_{j-1})\dots\Gamma_{\theta_{j-k+1}}^{j-k}(\cdot). \quad \Box \tag{1.9}$$

The significance of III, IV will become clear in subsection 1 (c).

(ii) Recall that the initial distribution μ of $X^N(t)$ has density θ' and for notational convenience we do *not* display the dependence on N for the components $x_{\xi,k}(t)$. By $\mathcal{L}(Y_n)\Rightarrow\mathcal{L}(Y)$ as $n\to\infty$ we denote weak convergence of random variables with values in $\mathbf{R}$, by $\mathcal{L}((Y_s^n)_{s\in\mathbf{R}^+})\Rightarrow\mathcal{L}((Y_s)_{s\in\mathbf{R}^+})$ as $n\to\infty$ weak convergence in the path space $C([0,\infty))$, see [10].

Next we formulate the *multiple time scale phenomenon* in the *mean-field limit* $N\to\infty$. This will imply in particular that the qualitative behaviour of our system can be read off from the *marginals of the interaction chains*, $\{\{\mu_{\theta'}^{j,k}\}, j, k\in\mathbf{N}, j\geq k\}$, alone.

Theorem 1 (Multiple time scale behaviour) *Let $j, k\in\mathbf{N}$. Then in the three possible cases $k=j$, $k>j$, $k<j$ we have the following behaviour for $X^N(t)=\{x_{\xi,k}(t)\mid\xi\in\Omega_N, k\in\mathbf{N}\}$ (here $\mathbf{R}^+=(0,\infty)$):*

$$\mathcal{L}((x_{\xi,j}(s\beta_j(N)))_{s\in\mathbf{R}^+})\underset{N\to\infty}{\Rightarrow}\mathcal{L}((\tilde{Z}_s^{j,\theta'})_{s\in\mathbf{R}^+}). \tag{1.10}$$

For all $k>j$:

$$\mathcal{L}((x_{\xi,k}(s\beta_j(N)))_{s\in\mathbf{R}^+})\underset{N\to\infty}{\Rightarrow}\delta_{\{Y_s\equiv\theta'\}}. \tag{1.11}$$

For all $k<j$ the following two relations hold:

$$\mathcal{L}((x_{\xi,k}(s\beta_j(N)+t\beta_k(N)))_{t\in\mathbf{R}^+})\underset{N\to\infty}{\Rightarrow}\mathcal{L}((Z_t^{k,\theta^*})_{t\in\mathbf{R}^+}), \tag{1.12}$$

where θ^* is independent of $w(t)$ driving Z_t^{k,θ^*} and $\mathcal{L}(\theta^*)=\mu_{\theta'}^{j,j-k+1}(\cdot)$.

$$\mathcal{L}([x_{\xi,k}(s\beta_j(N)+t\beta_k(N)))_{t\in\mathbf{R}^+}]\mid x_{\xi,k+1}(s\beta_j(N))=\theta,\mathcal{F}_{k+1}(\xi,s,N)) \underset{N\to\infty}{\Rightarrow}\mathcal{L}((Z_t^{k,\theta})_{t\in\mathbf{R}^+}), \tag{1.13}$$

where $\mathcal{F}_{k+1}(\xi,s,N)$ is the σ-algebra generated by $\{x_{\xi',0}(s\beta_j(N)), d(\xi,\xi')>k\}$.

Remark. This result is interpreted as follows: In the time scale $\beta_j(N)$ the averages over blocks $\xi_N(m)$ (with $m>j$) of volume bigger than N^j remain constant and equal to the initial density θ', while the block $\xi(j)$ has an average which fluctuates in the time scale $\beta_j(N)$ as a diffusion in external field with force $c_j(\theta'-x)$. The averages of smaller blocks can only fluctuate on smaller time scales namely $\beta_k(N)$ and the corresponding equilibrium distributions are given by the marginals of the interaction chain $(Z_k^j)_{k=-1,\ldots,j}$ with $Z_{-1}^j=\theta'$. (Everything is understood to be in the limit $N\to\infty$ of course.) This means that the interaction chain $(Z_k^j)_{k=-1,\ldots,j}$ describes, in the time scale $\beta_j(N)$, the *joint distribution* of the block averages $\left\{x_{\xi,l}\left(s(N)\,N^j+\sum_{k=0}^{j} s_k\,\beta_k(N)\right)\middle| l=j+1, j, \ldots, 0\right\}$ of the system for every vector $\vec{s}=(s_0, \ldots, s_j)$ with $s_j>0$, $s_k\geqq 0$, $k=0, \ldots, j-1$, $s(N)\to\infty$ and $s(N)/N\to 0$ (in the limit $N\to\infty$). In precise mathematical language:

For every fixed $\theta'\in(0,1)$, let $\{Z_k^j(N,\vec{s})\}_{k=0,\ldots,j}$ be defined by

$$\mathscr{L}((Z_{j-k}^j(N,\vec{s}))_{k=j,j-1,\ldots,0}\,|\,Z_{-1}^j=\theta')=$$
$$\mathscr{L}\left(x_{\xi,k}\left(s(N)\,N^j+\sum_{l=k}^{j} s_l\,\beta_l(N)\right)_{k=j,j-1,\ldots,0}\middle| x_{\xi,j+1}(s(N)\,N^j)=\theta'\right).$$

Then

$$(Z_k^j(N,\vec{s}))_{k=-1,0,\ldots,j} \underset{N\to\infty}{\Rightarrow} (Z_k^j)_{k=-1,0,\ldots,j}, \qquad \text{in law.}$$

Thus the whole point in taking the mean-field limit $N\to\infty$ is, that instead of the *space-time* picture given by the stochastic processes $(Z_k^j(N,\vec{s}))_{k=0,\ldots,j}$ which are non-Markovian, we obtain Markov chains in the limit, namely the interaction chains which are independent of $\vec{s}$.

(iii) In this paragraph we introduce a *renormalisation procedure* into the analysis. First we observe that the dynamics of $\{x_{\xi,k}(\cdot), \xi\in\Omega^N\}$, $k\in\mathbb{N}$ is determined in the time scale $\beta_k(N)$ and the mean-field limit $N\to\infty$ by the sequences $(c_0, c_1, \ldots)$ and $(F_0, F_1, F_2, \ldots)$. By absorbing c_k as well in the time scales $\beta_k(N)$ by passing to $c_k^{-1}\ \beta_k(N)$ and by replacing F_k by $\tilde{F}_k$ with $\tilde{F}_{k+1}=E^{\Gamma_x^k}\left(\frac{1}{c_{k+1}}\tilde{F}_k(\cdot)\right)$ $\tilde{F}_0(x)=c_0^{-1}g(x)$ we get a description completely in terms of the sequence of functions $(\tilde{F}_k)_{k\in\mathbb{N}}$.

We can go further. In [9, Theorem 2] it was shown that for $g(x)=d(x(1-x))$ and $c_0=c_1=1$ we have the following fixed-point property:

$$\tilde{F}_1(x)=\frac{d}{1+d}(x(1-x)),$$

so that in this case we can reduce the description by the functions $(\tilde{F}_k)_{k\in\mathbb{N}}$ to that by a sequence a_k defined by $\tilde{F}_k(x)=a_k(x(1-x))$. Therefore in this situation (i.e. $g(x)=\text{const}\cdot x(1-x)$) important properties of the system can be described by a *sequence of numbers* rather than by a sequence of functions $\tilde{F}_k$. However one expects this to be also (approximately) true for g for which the iterated application of the map $g=\tilde{F}_0\to\tilde{F}_1\to\tilde{F}_2\to\ldots$ brings the normalised image close to $\text{const}\cdot x(1-x)$ in a suitable sense. That is, for large k the $\tilde{F}_k$ will tend to look like $a_k\,x(1-x)$ so that properties involving the longterm behaviour on a spatially macroscopic level remain the same as for the system with $g(x)=\text{const}\cdot x(1-x)$. For this very reason we shall now investigate the *fixed-point properties* of the Fisher-Wright diffusions (i.e. $g(x)=\text{const}\cdot x(1-x)$) further. Ideally one would like to prove that $x(1-x)$ is the unique fixed point of this renormalization procedure and that the $a_k^{-1}\,\tilde{F}_k$ will produce something close for $k\to\infty$. The proof of global convergence involves a serious piece of nonlinear analysis, therefore in this paper we establish only that $x(1-x)$ is the *unique fixed point* is also *locally stable* and a weak form of *convergence* takes place. The desired strong results are in Baillon et al. [3]. To make our description precise we have to introduce the maps $\hat{F}$ and F^* below. (Later we shall use a whole sequence of these maps, belonging to the various values of the parameters (c_k).)

Define the following maps $\hat{F}$, F^* acting on the function g:

$$F^*(g)_{(x)} = \int \Gamma_x^0(dy)\, g(y)$$

$$\hat{F}(g)_{(x)} = \int \Gamma_x^0(dy)\, g(y) \left(\frac{a}{1+a/c_0}\right)^{-1} \quad \text{with } a := \sup_{y\in(0,1)} \left(\frac{g(y)}{y(1-y)}\right). \tag{1.14}$$

Then according to [9, Theorem 2]:

$$\hat{F}(x(1-x))\,(y) = y(1-y), \tag{1.15}$$

that is $x(1-x)$ is a *fixed-point* of the map $\hat{F}$. This map $\hat{F}$ is of course *nonlinear* since Γ_x^0 in (1.14) does depend on g! As mentioned above we expect that the iteration of the procedure in (1.14) will produce functions looking more and more like multiples of $x(1-x)$. For a computer simulation see Fig. 1. Mathematically we capture this by investigating *stability* properties of $\hat{F}$ around this fixed-point and by proving a convergence result. First note that using the explicit form of the density of $\Gamma_x^0(dy)$ (see (3.15)) we can define the map $\hat{F}$, F^* for all positive continuous functions g. We introduce the following normed space:

$$H = \left\{h\colon [0,1] \to \mathbf{R} \mid \sup_{x\in(0,1)} \left(\frac{|h(x)|}{x(1-x)}\right) < \infty\right\} \cap C([0,1]) \tag{1.16}$$

$$H^+ = \{h \in H \mid h(x) > 0 \ \text{ for } x \in (0,1),\ h \text{ is Lipschitz continuous}\}$$

$$\|h\| = \sup_{x\in(0,1)} \left(\frac{|h(x)|}{x(1-x)}\right).$$

The space H is in fact a Banach space. H^+ is the domain of the map $\hat{F}$, F^*. Both maps are continuous since $\Gamma_x^0(\cdot)$ is a continuous function of x and F^* does not increase the Lipschitz constant [9, Lemma 2.2].

For $h \in H^+$ the map $\hat{F}$ admits a tangential map $D_{\hat{F}}(h)_{(\cdot)}$ (not necessarily linear!) such that for all $f \in H$ (for proofs see Sect. 3)

$$\|[(\hat{F}(h+\varepsilon f) - \hat{F}(h) - \varepsilon D_{\hat{F}}(h)_{(f)}]\| = o(\varepsilon)\, \|f\|. \tag{1.17}$$

For an operator $A\colon H \to H$, define $\|A\| = \sup(\|A(x)\|\, \|x\|^{-1},\ x \in H \setminus \{0\})$.

Theorem 2 (Uniqueness, stability and convergence to the fixed-point) *Functions of the form $g(x) = dx(1-x)$ are the only functions in H^+ satisfying: there exists a function $L(c)$, such that $F^*(cg) = L(c)\, g$ for all $c \in \mathbf{R}^+$.*

For any $d > 0$ the function $g(x) = dx(1-x)$ is a fixed-point of $\hat{F}$ in H^+, and if $c_0 = 1$, then

$$\|D_{\hat{F}}(dx(1-x))\| = \frac{1}{1+d} < 1. \tag{1.18}$$

For every $g \in H^+$ with F^{*n} denoting the n-fold iteration of the function F^*:

$$(1.18') \qquad \frac{n}{c_0}(F^{*n})(g)(x) \text{ converges to } x(1-x) \text{ for all } x \in [0,1].$$

This means that $x(1-x)$ a locally stable fixed-point of $\hat{F}$ and if we apply F^* n-times then after normalisation it converges pointwise to $x(1-x)$.

In the sequel we shall need a *sequence* of compositions of maps associated with (c_k) where in particular the sequence (c_k) is an arbitrary sequence in $\mathbf{R}^+$. Define the following generalized version of (1.14):

$$\hat{F}^{(n)}(g)_{(x)} = d_n^{-1} F_n(x),$$

where

$$F_0(x) = g(x)$$
$$F_{n+1}(x) = \int \Gamma_x^n(dy) F_n(y)$$

and $d_{n+1} = \dfrac{d_n c_n}{c_n + d_n}$, $d_0 = \|g\|$.

Corollary. *Assume $c_k \equiv c$, then $g(x) = x(1-x)$ is the unique joint fixed-point of the maps $\hat{F}^k$, $k \in \mathbf{N}$ in H^+ and no other function $g \in H^+$ can have the property $\hat{F}^k(g) = b_k g$ for all k. Furthermore*

$$(1.19) \qquad \|D_{\hat{F}^{(k)}}\| < 1.$$

Now we discuss the case of varying (c_k). This case is more complicated and in general (1.18′) will not hold; for example if the c_k grow sufficiently rapidly, then the $\hat{F}^k(g)$ will not change much in shape for large k. Furthermore, the proof of (1.18) will reveal that

$$\|D_{\hat{F}^{(n)}}\| \xrightarrow[n \to \infty]{} 1 \quad \text{if } \frac{a_k}{c_k} \to 0$$

where $a_k = \|\hat{F}^{(k)}\|$. We shall see later on that this is the case when $\limsup c_k > 0$. Therefore, even for the linearization of $\hat{F}^{(n)}$, it is not always true that

$$\|D_{\hat{F}^{(n)}}(x(1-x))_{(h)}\| \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore we don't expect exponentially fast convergence. The asymptotic properties will depend on (c_k).

Later on we shall use the following sets of functions, which are the candidates for natural *universality classes*. For a fixed sequence (c_k) define:

$$(1.20) \qquad \mathscr{F}_0 = \{g \in H^+ \mid \lim_{n \to \infty} \|\hat{F}^{(n)}(g) - x(1-x)\|_\infty = 0\},$$

$$\mathscr{F}_1 = \{g \in H^+ \mid \sup_n \|\hat{F}^{(n)}(g) - x(1-x)\| < 1\},$$

$$\mathscr{F}_2 = \{g \in H^+ \mid \lim_{n \to \infty} \|\hat{F}^{(n)}(g) - x(1-x)\| = 0\}.$$

The question whether $\mathscr{F}_i$, $i = 1, 2$ contains a full neighbourhood of $x(1-x)$ for particular $(c_k)_{k \in \mathbf{N}}$ involves of course an analysis of the second order approxima-

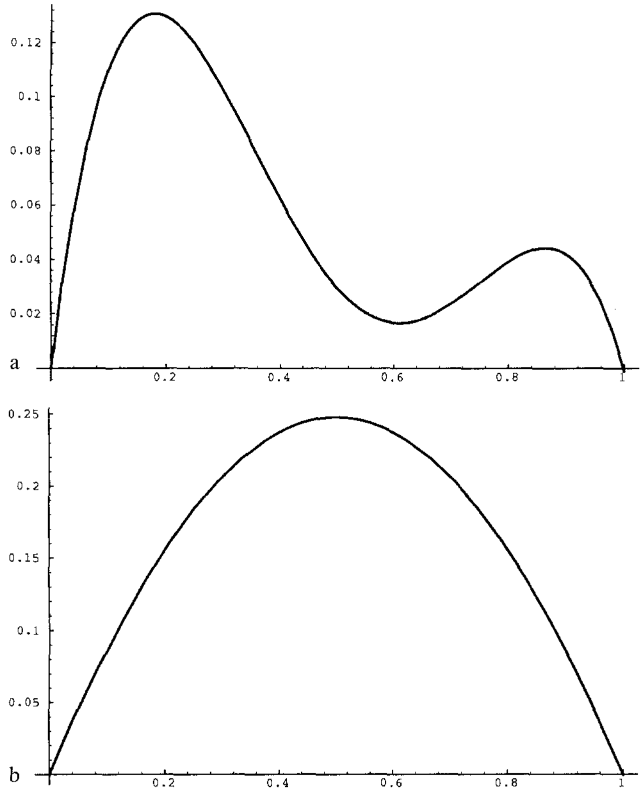

Fig. 1. Numerical simulation. $c_k = 1\ \forall k$. **a** g, **b** $\hat{F}^{20}$

tion of the map $\hat{F}$ which turns out to be quite involved and we defer this to a paper dealing with the study of the nonlinear map $\hat{F}$. In fact we believe more to be true and in a forthcoming paper by Baillon et al. [3] will establish the following:

Conjecture.

(a) If $\sum c_k^{-1} = \infty$, then $\mathscr{F}_0 \supseteq H^+$.

(b) If $\sum c_k^{-1} = +\infty$, $\int\limits_0^1 \frac{x(1-x)}{g(x)}\, dx < \infty$ then $g \in \mathscr{F}_2$.

(c) If $\sum\limits_k c_k^{-1} < \infty$, $\hat{F}^k(g) \underset{k \to \infty}{\not\to} x(1-x)$ but $\mathscr{F}_1 \supseteq H^+$.

Remark. From Dawson and Greven [9] we know that if $g(x) \sim x^\alpha$ for $x \to 0$, and $g(x) \sim (1-x)^\alpha$ for $x \to 1$ with $\alpha > 2$ then in fact $\hat{F}(g)(x) \sim x^\alpha$ as $x \to 0$ (similarly for $x \to 1$). That means that $\|\hat{F}^{(k)}(g) - x(1-x)\| \geqq 1\ \forall k$. We still have in this regime that $\|\hat{F}^{(k)}(g) - x(1-x)\|_\infty \to 0$ as $k \to \infty$.

(c) Phase-transition and cluster-formation in the mean-field limit

We shall now investigate the qualitative behaviour of our mean-field hierarchical system for interaction with long range ($N \to \infty$) and in very large time scales, that is, for times $\gg N^R$ but $o(N^{R+1})$ for arbitrary values of R. According to (1.12) we know that the behaviour of the k-th order average at time $s\beta_j(N) + t\beta_k(N)$ with $j>k$ is described by $\mu_{\theta'}^{j,k}$(recall $\gg N^j$ means $s\to\infty$). Since we are interested primarily in the components of the system we shall focus in this section on $j=R$, $k=0$ and study the behaviour of $\mu_{\theta'}^{R,0}$ for $R\to\infty$.

In subsection (c)(i) we are concerned with the dichotomy *stability* versus *clustering*. We shall make precise the idea that if the c_k remain big enough for large k, then the system should preserve it's initial density θ' even in the limit of infinite times and have a nontrivial limiting state for each value $\theta'\in(0,1)$. On the other hand, if the interaction is too weak, that is, the (c_k) are too small for large values of k, then the system should eventually end up locally in one of the traps $x_{\xi,0}\equiv 0$ or $x_{\xi,0}\equiv 1\ \forall \xi\in\Omega^\infty$, since the large blocks are not able to force the smaller blocks back to an average value θ' over the block. In the subsequent subsections we study for the latter case the way in which clustering occurs in more detail. Namely in the subsection (c)(ii) we shall investigate *self-similar cases* and in (c)(iii) cases where we see clusters of 0 or 1's grow on a *variety of scales*, a phenomenon first observed by Cox and Griffeath [8] for the 2-dimensional (simple) voter model.

(i) We start by introducing the notion *stability* and *clustering*. In this discussion we take the function g in (1.5) to be fixed and the sequence (c_k) to be a free parameter. We call the evolution associated with the (c_k) *stable in the mean-field limit* iff

$$\mu_{\theta'}^{R,0}(\cdot) \underset{R\to\infty}{\Rightarrow} \mu_{\theta'}^{\infty}(\cdot), \text{ with } \mu_{\theta'}^{\infty}((0,1))=1 \qquad \forall \theta'\in(0,1). \tag{1.21}$$

We say the system *clusters in the mean-field limit* iff

$$\mu_{\theta'}^{R,0}(\cdot) \underset{R\to\infty}{\Rightarrow} \theta'\delta_1+(1-\theta')\delta_0. \tag{1.22}$$

Theorem 3 (Stability versus clustering) *Assume that* $g(x)=dx(1-x)$ *for some* $d\in(0,\infty)$. *Then*

(1.23) *The system is stable in the mean-field limit if* $\sum_k \frac{1}{c_k}<\infty$.

(1.24) *The system clusters in the mean-field limit if* $\sum_k \frac{1}{c_k}=+\infty$.

Corollary (Universality) *The dichotomy in* (1.23), (1.24) *holds for all* $g\in H^+$.

(ii) We now turn to a further analysis of the evolutions which cluster in the mean-field limit and consider self-similarity and clustering on different time scales. We begin with the case in which such large clusters form (growth like range of a symmetric walk) that we can even have *self-similarity*. This case corresponds to the one-dimensional simple voter model, compare Arratia [2], where in the space-time scaling limit we get clusters of 0 and 1 whose boundaries form a system of annihilating brownian motions. We introduce first some terminology. Again g is considered fixed and (c_k) the free parameter.

The system is called *self-similar in the mean-field limit* iff

(1.25) $$\mu_{\theta'}^{j,k}(\cdot)=\mu_{\theta'}^{j+\ell,k+\ell}(\cdot) \quad \forall k<j, \ell=0,1,\dots$$

The system is said to be in the *domain of attraction* of self-similarity in the mean-field limit iff

(1.26) $$\mu_{\theta'}^{j+\ell,k+\ell}(\cdot) \underset{\ell\to\infty}{\Rightarrow} \tilde{\mu}_{\theta'}^{j,k}(\cdot) \quad \text{and} \quad \tilde{\mu}_{\theta'}^{j+i,k+i}(\cdot)=\tilde{\mu}_{\theta'}^{j,k}(\cdot) \quad \forall k<j, i\in\mathbf{N}.$$

Theorem 4 (Self-similarity) *If* $g(x)=x(1-x)$, *the system is self-similar in the mean-field limit iff*

(1.27) $$c_k=c_0\left(\frac{c_0}{1+c_0}\right)^k$$

and in the domain of attraction of self-similarity iff the c_k *satisfy for some* $p\in(0,1)$ *the relation*

(1.28) $$c_k=c_0\,p^k b_k \quad \textit{with} \quad b_{k+1}/b_k\to 1 \quad \textit{as} \quad k\to\infty.$$

Corollary (Universality) *The assertion* (1.28) *holds for all* $g\in\mathscr{F}_2$.

(iii) Next we develop a detailed quantitative analysis of *cluster-formation.* Two types of behaviour are known from studies of the voter model. One case is analogue to the above one, that in which there are large clusters and by rescaling time and space using a fixed scaling relation, we can describe the formation of clusters. The one-dimensional voter model is the typical example, rescaling time by $1/\varepsilon$ and space by $(\sqrt{\varepsilon})^{-1}$ gives for $\varepsilon\to 0$ an invariance principle for the right (left) border of the cluster containing 0. The second possibility is that clusters are smaller and grow at different time scales, for example this is the case in the 2-dimensional voter model in which clusters grow at rate $t^{\alpha/2}$, $\alpha\in(0,1]$ with α a random variable.

Note that in our hierarchical case and $N\to\infty$ this behaviour is reflected in the behaviour of the rescaled interaction chain $(Z_k^j)_{k=0,\dots,j}$. We focus on the second case discussed above. Under a proper simultaneous rescaling of n and j we obtain a nontrivial limit process which hits 0 or 1 with nondegenerate probability, the value of the latter depending on how many levels starting from j we look down. In other words *clusters of* 0 *and* 1*'s form at different (random) time scales* or *at given time scale clusters of different (random) orders of magnitude form.* This suggests the following:

Definition. If there exists a set $\{f_\alpha(\cdot)\}_{\alpha\in[0,1]}$ of functions $\mathbf{N}\to\mathbf{R}^+$ which satisfy: $f_\alpha(k)$ is increasing in α for all $k, f_0(k)\geqq 0, f_1(k)=k$ and the following property holds

(1.29) $$\mathscr{L}((Z^k_{[f_\alpha(k)]})_{\alpha\in[0,1]}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((Z_\alpha)_{\alpha\in[0,1]}),$$

then we call $f_\alpha(\cdot)$ a *set of cluster scales* and Z_α the *cluster process.*

This means that $f_\alpha(k)$ describes the blocksize such that at time N^k the probability of having a block with $N^{[f_\alpha(k)]}$ components ε-close to all 0 or all 1 is given in the limit $N \to \infty$ by the probability of Z_α being ε-close to 0 or 1 for every $\varepsilon > 0$.

One can give now a more systematic account of the formation of clusters than in the case of the voter model.

We can distinguish three *regimes in the cluster-formation* (by a cluster we always mean something expanding, in particular small, large, etc. refers to the speed of that expansion):

$$f_\alpha(k)/k \xrightarrow[k \to \infty]{} 0 \quad \forall \alpha \quad \textit{small clusters,}$$

$$f_\alpha(k)/k \xrightarrow[k \to \infty]{} \alpha \quad \forall \alpha \quad \textit{diffusive clustering,}$$

$$f_\alpha(k)/k \xrightarrow[k \to \infty]{} 1 \quad \forall \alpha \quad \textit{large clusters.}$$

At time N^k, in the first regime clusters are smaller than $(N^k)^\beta$ for all $\beta > 0$ while still growing to ∞ as time goes on, in the second regime they have size $(N^k)^\alpha$ with α a random variable with values in $[0, 1]$ provided $\mathscr{L}(Z_\alpha)$ is nondegenerate, while in the third regime they are of order N^k. In the case of the simple voter model only analogues of the last two cases appear (in dimension 2, dimension 1, respectively).

In the first and third case above it is of interest to study the correction terms and write for some h sublinear

$$f_\alpha(k) = h(k^\alpha) \qquad \text{(small clusters)}$$

$$f_\alpha(k) = k - (1-\alpha)\, h(k) \qquad \text{(large clusters).}$$

This allows us to determine what range occurs for the size of clusters on a finer scale. In order to find h define

$$\hat{h}(x) = \sum_{1}^{[x]} \frac{d_k}{c_k}, \qquad d_{k+1} = \frac{c_k\, d_k}{c_k + d_k}, \qquad d_0 = 1$$

and

$$h(k) = \hat{h}^{-1}(\log k) \qquad \text{in case of large clusters}$$

$$h(k) = \hat{h}^{-1}(\log \log k) \qquad \text{in case of small clusters.}$$

Theorem 5 (Pattern of cluster-formation) *Assume that $g(x) = dx(1-x)$. The system has for the three cases below, sets of cluster scales and a cluster process. The cluster process is given in all cases by*

(1.30) $Z_\alpha = Y\left(\log \dfrac{1}{1-\alpha}\right)$, *where $Y(t)$ is the Fisher-Wright diffusion*

on $[0, 1]$ which is generated by $1/2(x(1-x))\left(\dfrac{\partial}{\partial x}\right)^2$ and starts at θ',

and the sets of cluster scales are given by:

Case 1 *(diffusive clustering)*

(1.31) $$f_\alpha(k)=\alpha k \quad \textit{if } c_k \to c>0 \quad \textit{as } k\to\infty.$$

Case 2 *(large clusters)*

(1.32) $$\left.\begin{array}{l} f_\alpha(k)=k-(1-\alpha)\,h(k) \\ \textit{with } h(k)/k\to 0 \textit{ as } k\to\infty \end{array}\right\} \quad \textit{if } \exists\beta>0 \ \overline{\lim_{k\to\infty}}\, c_k k^\beta<\infty.$$

Case 3 *(small clusters)*

(1.33) $$\left.\begin{array}{l} f_\alpha(k)=h(k^\alpha) \\ \textit{and } h(k)=o(k) \end{array}\right\} \quad \textit{if } c_k \gg c,\, c_k \ll k\log(k))^\beta \textit{ for } \beta\leqq 1.$$

There exist sequences c_k converging to 0 *but with the Ansatz from* (1.32) $h(k)/k\to 1$ *as $k\to\infty$, and consequently $f_\alpha(k)/k\to\alpha$.*

Corollary (Universality) *The relations* (1.31)–(1.33) *hold for all* $g\in\mathscr{F}_2(x(1-x))$. (Refer to (1.20) for the definition of $\mathscr{F}_2$ and for the conjecture following that relation.)

Remark. The first interesting consequence of the corollary to Theorem 5 is the fact that the properties of the process in which clusters grow is universal in a whole class of evolutions, namely for all $g\in\mathscr{F}_2(x(1-x))$, we obtain the behaviour displayed in (1.30)–(1.33). This *universality* is related to the fact that $x(1-x)$ is a *fixed point* under the map $\hat{F}$. This fact is what makes it interesting to study the questions, answered for 2-dimensional voter model by Cox and Griffeath [8] and the 1-dimensional voter model Arratia [2], here in this paper for a mean-field limit of a hierarchical model.

The second consequence is that we get the whole continuum of possible behaviour ranging from stability $c_k\geqq k(\log k)^s$, with $s>1$, certain type (see (1.33)) small clusters $c_k\sim k\,(\log k)^s$ with $s\leqq 1$; diffusive clustering $c_k\approx$ const, large clusters $c_k=k^{-\beta}(\beta>0)$ and self-similarity $c_k=c_0\left(\frac{c_0}{1+c_0}\right)^{-k}$. This raises the question as to how to prove the analogous picture for the voter model considering general kernels $p(x,y)$ for picking the neighbour whose opinion one adopts.

(d) Hierarchical systems for fixed N

(i) In this section we shall address the question as to how closely the mean-field prediction describes the hierarchical system for *fixed N*. In systems with $g(x)=x(1-x)$ this investigation can be based on the duality relations involving systems of coalescing walks on the hierarchical structure. However in view of recent results of Cox and Greven [7] the duality relations are not essential for the conclusion we are after here.

When we do not let $N\to\infty$, serious problems arise when considering the multiple time-scale behaviour and the fixed-point analysis of Theorem 1, respectively Theorem 2. We shall not investigate the first more involved problem in this paper but focus on the analog of Theorem 3 on stability versus clustering and sketch some ideas on the fixed-point problem. We shall see in this subsection that the mean-field limit prediction of Theorem 3 is *very* good. The fixed-point property of Fisher-Wright systems is discussed in part (ii) of this subsection.

We shall need the two transformed sequences (C_k), $\tilde{c}_k$ derived from (c_k):

$$\tilde{c}_m = \sum_{k=m}^{\infty} c_k N^{-2(k-m)}. \tag{1.35}$$

$$C_m = \sum_{k=m}^{\infty} \tilde{c}_k N^{-(k-m)}. \tag{1.36}$$

One important aspect of the proof of the theorem below is the fact that we can apply Fourier analysis to study the random walk on Ω^N due to the group structure of this hierarchical set. This idea has previously been exploited by Sawyer and Felsenstein [13].

Theorem 6 (Behaviour for N fixed) (a) *The system $\{x_{\xi,0}(t)\}_{\xi\in\Omega^N}$ has a set of extremal invariant measures which are homogeneous (under the group action induced by addtion in Ω^N) and which are given as the weak limit of $\mathscr{L}((x_{\xi,0}(t))_\xi | x_{\xi,0}(0)=\theta,\ \xi\in\Omega^N)$.*

Case 1 $\sum_m \frac{1}{C_m} < \infty$.

$$\{\nu_\theta\}_{\theta\in[0,1]} \quad \textit{with}\ E^{\nu_\theta}(x_{\xi,0})=\theta,\ \nu_\theta \textit{ is mixing.} \tag{1.37}$$

Case 2 $\sum_m \frac{1}{C_m} = +\infty$

$$\{\delta_{\{x_{\xi,0}\equiv 1\}}, \delta_{\{x_{\xi,0}\equiv 0\}}\}. \tag{1.38}$$

(b) *For every homogeneous initial distribution μ which is shift-ergodic and has the property $E^\mu(x_{\xi,0})=\theta$ we have*

$$\mathscr{L}((x_{\xi,0}(t))_{\xi\in\Omega^N}) \underset{t\to\infty}{\Rightarrow} \begin{cases} \nu_\theta & \textit{if } \sum C_m^{-1} < \infty \\ \theta\,\delta_{\{x_{\xi,0}\equiv 1\}} + (1-\theta)\,\delta_{\{x_{\xi,0}\equiv 0\}} & \textit{if } \sum_m C_m^{-1} = \infty. \end{cases} \tag{1.39}$$

We should now like to compare the relation between the conditions

$$\sum_k \frac{1}{C_k} \begin{cases} <\infty \\ =\infty \end{cases} \quad \text{and} \quad \sum_k \frac{1}{c_k} \begin{cases} <\infty \\ =\infty. \end{cases}$$

By explicit calculation one checks that for c_k such that

$$\overline{\lim_{k\to\infty}} \left(\sqrt[k]{c_k}\right) < N$$

we have

$$\sum_k \frac{1}{c_k}\begin{cases} <\infty \\ =\infty \end{cases} \Leftrightarrow \sum_k \frac{1}{C_k}\begin{cases} <\infty \\ =\infty. \end{cases} \tag{1.40}$$

Therefore in *all* cases of interest the mean-field prediction gives the correct answer. Only in the case where $\overline{\lim\limits_{k\to\infty}}(\sqrt[k]{c_k}) \geqq N$ (but $\sum c_k N^{-k} < \infty$ by assumption on the model!) we may have different behaviour.

Obviously the critical range of growth is $c_k \sim k(\log k)^s$ since for $s>1$ the sum over c_k^{-1} exists and for $s \leqq 1$ diverges.

Therefore we have

Corollary (of Theorems 3 and 6) (a) *If* $\liminf\limits_{k\to\infty} c_k/k(\log k)^s > 0$ *for some* $s>1$, *then the system is stable in the mean-field limit and for the hierarchical system with parameter N we are in case* 1 (1.37).
(b) *If* $\limsup\limits_{k\to\infty} c_k/k(\log k) < \infty$, *the system clusters in the mean-field limit and also for every hierarchical system with parameter N we are in case* 2 (1.38).

(ii) The above discussion raises the question whether the whole scenario presented in the introduction can also be verified in the case of N fixed. The first point is to explain the universality via the existence of the multiple time scale picture and the corresponding fixed-point property of the systems of interacting Fisher-Wright diffusions. The second point is to express the dichotomy stability versus clustering, and the pattern of cluster formation, via the interaction chain. For that purpose we modify the system (1.5) in a way analogous to passing from interacting systems with components indexed by $\mathbf{Z}^d$ to those indexed by the d-dimensional torus (compare [6]).

Define

$$\Omega^{N,k} = \{\xi \in \Omega^N \mid d(\xi, \tilde{\xi}) \leq k\}$$

and modify the definition of $x_{\xi,j}$ in (1.5) for $j>k$ by the convention $x_{\xi,j} = x_{\xi,k}$. Then consider the system of stochastic differential equations given by (1.5) but replacing the index set Ω^N by $\Omega^{N,k}$ and using above convention on the $x_{\xi,j}$ for $j>k$. We denote the resulting system by $\{(\bar{x}_{\xi,0}(t)), \xi \in \Omega^{N,k}\}$ that is we supress the dependence of the dynamic on k!

To define the fixed-point property we introduce the process

$$\theta^k(t) := \frac{1}{N^k}\Big(\sum_{\xi;\, d(\xi,\tilde{\xi}) \leq k} \bar{x}_{\xi,0}(t\tilde{\beta}_k(N))\Big), \tag{1.41}$$

for some scale function $\tilde{\beta}_k$ to be specified later on.

Suppose now that the following two relations hold:

$$\mathscr{L}((\theta^k(t))_{t\in\mathbf{R}^+}) \underset{k\to\infty}{\Rightarrow} \mathscr{L}((\theta_t)_{t\in\mathbf{R}^+}) \tag{1.42}$$

where (θ_t) is a martingale and a diffusion on $[0,1]$ with generator

$$u_g(x)\left(\frac{\partial}{\partial x}\right)^2. \tag{1.43}$$

It is not within the scope of the methods of this paper to verify the above relations, which amounts to proving that the *finite system scheme* in the sense of Cox and Greven [6] holds for hierarchical models. This will be treated in a forthcoming paper of Cox, Greven and Shiga.

Theorem 7 (Fixed-point property for N fixed) *Assume that N is fixed and $\geqq 2$ and that $g(x)=x(1-x)$.*

If (1.42) *and* (1.43) *hold, then we have*

$$u_{x(1-x)}(y)=\text{const}\; y(1-y). \tag{1.44}$$

Remark. This means that even in the case in which we do not pass to the mean-field limit $N\to\infty$ the function $g(x)=x(1-x)$ does have a fixed-point property. The time scales are given simply by $\beta_k(N)$ if $\sum\limits_m C_m^{-1}<\infty$. In the case $\sum\limits_n C_m^{-1}=\infty$ thing are more subtle due to the fact that the asymptotics of hitting times for random walks on the set $\Omega_N^k\subseteq\Omega_N$ where all sequences became 0 after the index k does depend on the decay of the c_k, in the latter case.

So far we have shown that the mean-field prediction is good as far as the stability-clustering dichotomy goes. Naturally we would like to show that Theorems 2–5 can be proved as well for a wide class of diffusion coefficients g, namely those which are in the domain of attraction of the Fisher-Wright interacting diffusions, which is also a "fixed-point" for the transformation described in (1.41)–(1.43) without passing to the mean-field limit. It seems very well within reach of the present methodology to prove that these results hold for $g(x)=x(1-x)$, but the problem is to show the stability of the fixed-point and to prove then, that this implies that the results hold for all g in the domain of attraction. However hierarchical models are the best candidates to try such a program.

2 Proof of Theorem 1

For the most part this section will consist in reducing everything to the theorems proved in [9] about hierarchical one or two level systems, that is, systems with a *finite* number of components in the limit of the number of components going to infinity.

(i) The first step in the proof consists in showing (1.11), that is, in the time scale $\beta_j(N)$ the average on the levels $l>j$ remain constant and are equal to θ'. We begin by rewriting the equations defining the system $\{x_{\xi,0}(t),\,\xi\in\Omega^N\}$. For this purpose, recall that the dependence on N is *not* displayed in the notation for the components of $X^N(t)$. Note that $\mathscr{L}(\{w_\xi(\lambda t),\,\xi\in\Omega^N\})=\mathscr{L}(\{\sqrt{\lambda}\,w_\xi(t),\,\xi\in\Omega^N\})$. Hence the *law* of $\{x_{\xi,l}(t)\}_\xi$ is given by the following system of stochastic differential equations, which is derived from (1.5) and where we write as $l=j+k$ with $k\geqq 1$:

$$(2.1)\qquad dx_{\xi,j+k}(t\beta_j(N)) \\ =\frac{1}{N^k}\left(\sum_{m=k+1}^{\infty} c_{j+m-1}\frac{1}{N^{m-k-1}}(x_{\xi,j+m}(t\beta_j(N))-x_{\xi,j+k}(t\beta_j(N)))\right)dt \\ +\frac{1}{\sqrt{N^k}}\left(\frac{1}{\sqrt{N^{j+k}}}\sum_{\xi'\in\xi(j+k)}\sqrt{2g(x_{\xi',0}(t\beta_j(N))}\,dw_{\xi'}(t)\right).$$

We analyze first the drift term in the expression above. By the summability assumption on the (c_k) we have that there exists a $K=K(j,k)$ so that for all $N\in\mathbf{N}$:

$$\left|\sum_{m=k+1}^{\infty} c_{j+m-1}\frac{1}{N^{m-k-1}}(x_{\xi,j+m}(t\beta_j(N))-x_{\xi,j+k}(t\beta_j(N)))\right|\leqq K<\infty.$$

Furthermore we can bound the martingale term on the rhs of (2.1) by observing that

$$\left(\frac{1}{\sqrt{N^{j+k}}}\sum_{\xi'\in\xi(j+k)}\sqrt{2g(x_{\xi',0}(t\beta_j(N))}\,dw_{\xi'}(t)\right)_{t\in\mathbf{R}^+}$$

is a martingale with mean quadratic variation bounded by $Ct+\text{const}$, independent of N.

Combining the above two facts and using Chebychev's inequality and a standard martingale inequality [10, (2.56)] we obtain for every $k\geqq 1$ and fixed $T\in\mathbf{R}^+$:

$$(2.2)\qquad \operatorname{Prob}(\sup_{t\leqq T}|x_{\xi,j+k}(t\beta_j(N))-x_{\xi,j+k}(0)|\geqq\varepsilon)\leqq\frac{1}{\varepsilon}\frac{\text{Const}}{\sqrt{N}}.$$

On the other hand by the law of large numbers we know that

$$x_{\xi,j+k}(0)\xrightarrow[N\to\infty]{}\theta',\qquad \mu-\text{a.s.}$$

Therefore (2.2) implies

$$\operatorname{Prob}(\sup_{t\leqq T}|x_{\xi,j+k}(t\beta_j(N))-\theta'|\geqq\varepsilon)\xrightarrow[N\to\infty]{}0\qquad\forall\varepsilon>0.$$

Therefore we have proved that for all $k\geqq 1$

$$(2.3)\qquad \mathscr{L}(\{x_{\xi,j+k}(t\beta_j(N))\}_{t\in R^+})\underset{N\to\infty}{\Rightarrow}\delta_{\{Z\equiv\theta'\}},$$

which completes the proof of (1.11).

(ii) The next step is to study $x_{\xi,j}(t\beta_j(N))$ and prove (1.10). The basic tool here is to reduce things to the situation in which we can apply the results obtained in [9] on finite systems when the system size tends to infinity. We introduce the abbreviation $\bar{x}_{\xi,m}(t)=x_{\xi,m}(t\beta_j(N))$. Then we obtain, using again the scaling property of brownian motion, that the law of $\{\bar{x}_{\xi,m}(t),\ \xi\in\Omega^N\}$ is determined by the following system of differential equations:

$$(2.4)\qquad d\bar{x}_{\xi,j}(t)=c_j(\bar{x}_{\xi,j+1}(t)-\bar{x}_{\xi,j}(t))+\frac{1}{\sqrt{N^j}}\sum_{\xi'\in\xi(j)}\sqrt{2g(\bar{x}_{\xi',0}(t))}\,dw_{\xi'}(t)$$
$$+\frac{1}{N}\left[\left(\sum_{m=j+2}^{\infty} c_{m-1}\frac{1}{N^{m-(j+2)}}(\bar{x}_{\xi,m}(t)-\bar{x}_{\xi,j}(t))\right)dt\right].$$

The next observation is that by (2.3) we can replace $\bar{x}_{\xi,j+1}(t)$ uniformly in $t\leqq T$ by θ'. The third term in (2.4) is $O(N^{-1})$. Thus for $N\to\infty$ we can replace this equation uniformly in $t\leqq T$ by the system of equations giving in the limit the same law as $\mathscr{L}(\{\bar{x}_{\xi,j}\,|\,\xi\in\Omega^N\})$

$$(2.5)\qquad d\bar{\bar{x}}_{\xi,j}(t)=c_j(\theta'-\bar{\bar{x}}_{\xi,j}(t))\,dt+\frac{1}{\sqrt{N^j}}\Big(\sum_{\xi'\in\xi(j)}\sqrt{2g(\bar{\bar{x}}_{\xi',0}(t))}\,dw_{\xi'}(t)\Big).$$

The next goal is to replace the second term on the right side of (2.5) by $\sqrt{2F_j(\bar{\bar{x}}_{\xi,j}(t))}\,dw_\xi(t)$. This again is understood in the sense of equality in law of the solution for the resulting equation in the limit $N\to\infty$, with the given equation. The proof proceeds by induction over j using Theorem 4 in [9] which dealt with systems with a hierarchy of two levels instead of the countably many we have here.

Introduce for that purpose the map $\xi\to\xi'$ which consists of dropping the coordinates $\xi_0,\xi_1,\ldots,\xi_{j-2}$. Define

$$\tilde{x}_{\xi',0}(t)=\bar{\bar{x}}_{\xi,j-1}(tN^{j-1}).$$

By the induction hypotheses about the system (2.5) we can replace the system of stochastic differential equations for $\{\bar{\bar{x}}_{\xi,j-1}(t)\}_\xi$ in the time scale tN^{j-1} in the limit $N\to\infty$ by

$$d\tilde{x}_{\xi',0}(t)=c_j(\theta'-\tilde{x}_{\xi',0}(t))\,dt+\sqrt{F_{j-1}(\tilde{x}_{\xi',0}(t))}\,dw_{\xi'}(t).$$

Therefore instead of $(\bar{\bar{x}}_{\xi,j}(tN^j))_\xi)$ the object in question is now $\{\tilde{x}_{\xi',1}(tN)\}_{\xi'}$. To the latter system we apply Proposition 2 in [9] to obtain the assertion for j. This theorem also contains the case $j=1$, so that we can start the induction.

Combining the last steps implies that in the limit $N\to\infty$ our Eq. (2.5) can be replaced by (in the sense of weak convergence of processes):

$$(2.6)\qquad \begin{aligned} &d\tilde{x}_{\xi,j}(t)=c_j(\theta'-\tilde{x}_{\xi,j}(t))\,dt+\sqrt{2F_j(\tilde{x}_{\xi,j}(t))}\,dw_\xi(t)\\ &\mathscr{L}(\tilde{x}_{\xi,j}(0))=\delta_{\theta'}\end{aligned}$$

which proves (1.10).

(iii) It remains to prove relations (1.12) and (1.13). The proof of (1.12) will be based on (1.13). The relation (1.13) however is, using the technique in (2.4), an immediate consequence of Theorem 4 (1.27) in [9].

The relation (1.12) follows for $k=j-1$ from combining (1.13) with (1.10) and (1.11). For $j-2, j-3$ etc. simply observe that according to (1.11), $x_{\xi,m}(t)$

does not change in time scale $\beta_k(N)$ if $m>k$ once $N\to\infty$, and that furthermore $\Gamma_\theta^m(\cdot)$ is continuous in θ for every $m\in\mathbf{N}$. Then use (1.13) successively, from the bottom to the top.

3 Proof of Theorem 2

The proof of Theorem 2 and its corollary proceeds in two steps. First we show the uniqueness of the fixed point for F^* and the convergence result then in a second step we investigate local properties of $\hat{F}$ around that fixed point.

(i) We want to prove $F^*(cg)=L(c)g\ \forall c\in\mathbf{R}^+$ for some function L implies $g(x)=\text{const}\,x(1-x)$. The convergence result will be a by product. In order derive this uniqueness of the fixed point, we work with the following relation between the $(F_k)(c_k)$ and $E(Z_j^j)^2$:

$$E(Z_j^j)^2=\left(\sum_{k=0}^{j} c_k^{-1}\right)F_{j+1}(\theta)+\theta^2. \tag{3.1}$$

We shall prove this equation below and now continue with the proof.

Suppose we have a function $g\in H^+$ such that the following holds:

$$\int \Gamma_\theta^{c,g}(dx)\,g(x)=L(c)\,g(\theta) \qquad \text{for some } L(c)\in\mathbf{R}^+, \qquad \forall c>0, \tag{3.2}$$

where $\Gamma_\theta^{c,g}(\cdot)$ is the equilibrium distribution of the Markov process x_t defined by

$$dx_t=c(\theta-x_t)\,dt+\sqrt{2g(x_t)}\,dw_t, \qquad w(t) \text{ is standard brownian motion.} \tag{3.3}$$

Define numbers $L^{(j)}$ recursively $L^{(j)}g=L(c_j/L^{(j-1)})g$, $L^{(0)}g=L(c_0)\,g$. We can then rewrite (3.1) as follows:

$$E(Z_j^j)^2-\theta^2=\left(\sum_{k=0}^{j} c_k^{-1}\right)L^{(j)}\,g(\theta). \tag{3.4}$$

In Sect. 4 we will show that $\sum_k c_k^{-1}=+\infty$, implies that $\mathscr{L}(Z_j^j)\underset{j\to\infty}{\Rightarrow}\theta\delta_1+(1-\theta)\,\delta_0$. Then for such a sequence c_k this together with (3.4) implies that:

$$\theta(1-\theta)=\left(\sum_{k=0}^{j} c_k^{-1}\right)L^{(j)}\,g(\theta)+o(1) \quad \text{as} \quad j\to\infty. \tag{3.5}$$

Hence (3.5) implies that $\left(\sum_{k=0}^{j} c_k^{-1}\right)L^{(j)}$ converges as $j\to\infty$ to say $c^*>0$. But now (3.5) becomes

$$\theta(1-\theta)=c^*\,g(\theta)+o(1) \quad \text{as} \quad j\to\infty \tag{3.6}$$

or in other words

$$g(\theta)=(c^*)^{-1}\,\theta(1-\theta). \quad \text{qed} \tag{3.7}$$

Applying this to the sequence $c_k = c_0 > 0\ \forall k$ implies in particular that (1.18′) holds.

Finally we prove the relation (3.1). Consider the following stochastic differential equation

$$dx_t = c(\theta - x_t)\,dt + \sqrt{2g(x_t)}\,dw(t) \tag{3.8}$$

whose equilibrium distribution $\Gamma_\theta^{c,g}(\cdot)$ defines the transitions of the chain $(Z_k^j)_{k=1,\dots,j}$ (c, g run through c_k and F_k, compare Sect. 2).

By Ito's formula we can get a differential equation for EX_t^2 and derive

$$\int x^2\,\Gamma_\theta^{c,g}(dx) = \int \frac{g(x)}{c}\,\Gamma_\theta^{c,g}(dx) + \theta^2. \tag{3.9}$$

By the definition of F_k this means (recall (1.8)):

$$\int x^2\,\Gamma_\theta^{c_0,g}(dx) = \frac{F_1(\theta)}{c_0} + \theta^2. \tag{3.10}$$

Hence iteration gives the claimed identity (3.1).

(ii) The second step is now to analyse the local properties of $\hat{F}$ around the fixed-point. The map $\hat{F}$ can be written in the form $\hat{F}(g) = F^*(g)\,G(g)$ with $F^*(g)(\theta) = E^{\Gamma_\theta}(g(\cdot))$, $G(g)(\theta) \equiv \left(\dfrac{c\,\|g\|}{c + \|g\|}\right)^{-1}$. We start by simply representing the Frechet derivative of F^* denoted $D_{F^*}(x(1-x))(\cdot)$ in terms of a kernel acting on functions. The norm we use on H, suggests writing functions $h \in H$ in the form $h(x) = x(1-x)f(x)$ (recall (1.16) here).

Lemma 3.1 *Let $h(x) = x(1-x)f(x)$ with $h \in H$ and let $p = \dfrac{c}{1+c}$. Furthermore, $\dfrac{1-p}{p}$ is abbreviated by γ and $B(a, b)$ is the Beta-function (see* [1] *).*

$$\begin{aligned} (3.11)\quad pD_{F^*}(x(1-x))_{(h)} &= [B(\theta/\gamma, (1-\theta)/\gamma)]^{-2} \\ &\cdot \left\{ \left[-B(\theta/\gamma, (1-\theta)/\gamma) \int x^{\theta/\gamma}(1-x)^{(1-\theta)/\gamma} \int_0^x \frac{(\theta - y) f(y)}{\gamma y(1-y)}\,dy \right] \right. \\ &+ \left[B(\theta/\gamma, (1-\theta)/\gamma) \int x^{(\theta/\gamma)-1}(1-x)^{((1-\theta)/\gamma)-1} \int_0^x \frac{(\theta - y) f(y)}{\gamma (y(1-y))}\,dy \right] \\ &+ \left. [B((\theta/\gamma)+1, ((1-\theta)/\gamma)+1) \int f(x)\,x^{(\theta/\gamma)-1}(1-x)^{((1-\theta)/\gamma)-1}\,dx] \right\}. \end{aligned}$$

The second and important step is now to show that the right hand side of (3.11) can be simplified quite a bit. $D_{F^*}(x(1-x))_{(h)}$ is a linear operator on H, which we write as $D_{F^*}(x(1-x))_{(h)}[\cdot]$. For the function G we can find a tangential (nonlinear) map in h, of a very simple form which we call $D_G(x(1-x))_{(h)}$. (See (1.17) for terminology).

Lemma 3.2 *With the same notation as in Lemma* 3.1 *we have:*

$$(3.12)\qquad D_{F^*}(x(1-x))_{(h)}(\theta)=\frac{\theta(1-\theta)}{1+\gamma}\int\frac{x^{\theta/\gamma}(1-x)^{(1-\theta)/\gamma}}{B(\theta/\gamma+1,((1-\theta)/\gamma)+1)}f(x)\,dx$$
$$D_G(x(1-x))_{(f)}(\theta)\equiv -1\,\|f\|_\infty.$$

From this point it is easy to finish the proof of the Theorem 2. Note that by the definition of the Beta-function,

$$K(\theta,dx)=\frac{x^{\theta/\gamma}(1-x)^{(1-\theta)/\gamma}}{B(\theta/\gamma+1,((1-\theta)/\gamma)+1)}\,dx$$

is a probability transition kernel. Define $D_{\hat F}=D_{F^*}G+FD_G$, then $\hat F$ and $D_{\hat F}$ are tangential. Therefore (3.2) implies after some calculation that for $h(x)=x(1-x)f(x)$ with $h\in H$:

$$(3.13)\qquad \|D_{\hat F}(x(1-x))_{(h)}[\,\cdot\,]\|\leqq\frac{1}{\gamma+1}\|f\|_\infty=\frac{1}{\gamma+1}\|h\|.$$

For $A\colon H\to H$ let $\|A\|$ denote the operator norm, namely, $\|A\|=\sup\limits_{x\neq 0}\frac{\|A(x)\|}{\|x\|}$. Then (3.13) can be rewritten as

$$(3.14)\qquad \|D_{\hat F}(x(1-x))\|\leqq\frac{1}{1+\gamma}<1.\quad\text{qed}$$

It therefore remains to verify the relations (3.11) and (3.12) to finish the proof of Theorem 2.

Proof of Lemma 3.1 and 3.2 In order to show that the operator F is Frechet-differentiable at the point $x(1-x)$ it suffices to show that the Gateaux-derivative exists in a neighbourhood of this point and is continuous there. The continuity is obvious from the expression we get for the Gateaux-derivative (which then equals of course the Frechet derivative), compare (3.11).

Define

$$(3.15)\qquad \Gamma_\gamma^\varepsilon(\theta,dx)=\frac{c(\varepsilon,\theta)}{x(1-x)[1+\varepsilon f(x)]}\exp\left(\int_0^x\frac{\theta-y}{\gamma(y(1-y))[1+\varepsilon f(y)]}\,dy\right)dx$$

where $c(\varepsilon,\theta)$ is the normalising factor turning $\Gamma_\gamma^\varepsilon(\theta,\cdot)$ into a probability measure. Then we have to compute

$$\left(\frac{d}{d\varepsilon}\int x(1-x)[1+\varepsilon f(x)]\,\Gamma_\gamma^\varepsilon(\theta,dx)\right)\bigg|_{\varepsilon=0}.$$

This expression equals

$$[B(\theta/\gamma,(1-\theta)/\gamma)]^{-2}\left\{\left[-B(\theta/\gamma,(1-\theta)/\gamma)\int x^{\theta/\gamma}(1-x)^{(1-\theta)/\gamma}\int_0^x\frac{(\theta-y)f(y)}{\gamma y(1-y)}\,dy\right]\right.$$
$$+\left[B((\theta/\gamma)+1,((1-\theta)/\gamma)+1)\int x^{(\theta/\gamma)-1}(1-x)^{((1-\theta)/\gamma)-1}\int_0^x\frac{(\theta-y)f(y)}{\gamma y(1-y)}\,dy\right]$$
$$\left.+\left[B((\theta/\gamma)+1,((1-\theta)/\gamma)+1)\int f(x)\,x^{(\theta/\gamma)-1}(1-x)^{((1-\theta)/\gamma)-1}\,dx\right]\right\},$$

which proves Lemma 3.1.

Remark. If we put $F(x)=\int_0^x \frac{(\theta-y)f(y)}{\gamma y(1-y)}\,dy$ then $F(x)+\text{const}$ gives the *same* result so that we can use $-\int_x^1 \frac{(\theta-y)f(y)}{\gamma y(1-y)}\,dy$.

Let

$$\begin{aligned}B_x(\theta/\gamma,(1-\theta)/\gamma)&=\int_0^x y^{\theta/\gamma-1}(1-y)^{((1-\theta)/\gamma)-1}\,dy\\&=I_x(\theta/\gamma,(1-\theta)/\gamma)\,B(\theta/\gamma,(1-\theta)/\gamma).\end{aligned}$$

This means that we can eliminate the end terms in the integration by parts. Therefore

$$\begin{aligned}&G(\theta)/(1+\gamma)=-[B(\theta/\gamma,(1-\theta)/\gamma)]^{-1}\\&\cdot\left\{\left[\int\left(\frac{\theta(1-\theta)}{1+\gamma}B_x(\theta/\gamma,(1-\theta)/\gamma)-B_x(\theta/\gamma+1,(1-\theta)/\gamma+1)\right)\frac{(\theta-x)f(x)}{\gamma x(1-x)}\,dx\right]\right.\\&\left.-\left[\frac{\theta(1-\theta)}{1+\gamma}\int f(x)\,x^{(\theta/\gamma)-1}(1-x)^{((1-\theta)/\gamma)-1}\,dx\right]\right\}.\end{aligned}$$

We get

$$G(\theta)=\int\frac{x^{\frac{\theta}{\gamma}}(1-x)^{\frac{1-\theta}{\gamma}}}{B(\theta/\gamma,(1-\theta)/\gamma)}f(x)\,dx. \tag{3.16}$$

The derivation uses the following facts (compare [1]):

$$\begin{aligned}B(z+1,w+1)&=\frac{zw}{(z+w)(z+w+1)}B(z,w)\\I_x(a+1,b)&=I_x(a,b)-\frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}x^a(1-x)^b\\I_x(a,b+1)&=\frac{(a+b)\,I_x(a,b)-a I_x(a+1,b)}{b}\\I_x(a+1,b+1)&=\frac{(a+1+b)}{b}\left\{I_x(a,b)-\frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}x^a(1-x)^b\right\}\\&\quad-\frac{a+1}{b}\left\{I_x(a,b)-\frac{\Gamma(a+b)}{\Gamma(a+1)\,\Gamma(b)}x^a(1-x)^b\right.\\&\quad\left.-\frac{\Gamma(a+1-b)}{\Gamma(a+2)\,\Gamma(b)}x^{a+1}(1-x)^b\right\}.\end{aligned}$$

This finishes the proof of Lemma 3.2.

4 Proof of Theorem 3

(a) The starting point is the following.

Lemma 4.1 *The sequence $\mu_\theta^{R,0}(\cdot)$ of measures converges weakly for $R\to\infty$ for every $\theta\in[0,1]$, to a probability measure $\mu_\theta^\infty(\cdot)$ on $[0,1]$.*

Proof. There are two cases to distinguish, namely $\sum c_k^{-1}=\infty$, $<\infty$. We first consider the case $\sum c_k^{-1}<\infty$. By the relation (3.1)

$$\int x^2 \mu_\theta^{j,j-k}(dx)=F_{j+1}(\theta)\left(\sum_{l=k}^{j} c_l^{-1}\right)+\theta^2$$

and consequently

$$\operatorname{Var}(Z_{j-k}^j)=F_{j+1}(\theta)\sum_{l=k}^{j} c_l^{-1}\leqq \|g\|_\infty \sum_{l=k}^{j} c_l^{-1}\leqq \|g\|_\infty \sum_{l=k}^{\infty} c_l^{-1}.$$

Hence

$$\sup_{j>k}(\operatorname{Var} Z_{j-k}^j)\xrightarrow[k\to\infty]{} 0. \tag{4.1}$$

Note that by construction for every $S>R$, with $R, S\in\mathbb{N}$, $\theta\in[0,1]$:

$$\mathscr{L}((Z_j^R)_{j=0,\dots,R}\,|\,Z_{-1}^R=\theta)=\mathscr{L}((Z_j^S)_{j=S-R,S-R+1,\dots,S}\,|\,Z_{S-R-1}^S=\theta).$$

Furthermore, given any Lipschitz function f on $[0,1]$, $\theta\to\int_{[0,1]} f(y)\,\Gamma_\theta^j(dy)$ also has Lipschitz constant at most $\operatorname{Lip}(f)$ (cf. [9, (2.58)]). Therefore for $j>k$ we have

$$\left|\int_{[0,1]} f(y)\,\mu_\theta^{j,0}(dy)-\int_{[0,1]} f(y)\,\mu_\theta^{k,0}(dy)\right|\leqq \operatorname{Lip}(f)\left(\|g\|_\infty\sum_{l=k}^{j} c_l^{-1}\right)^{\frac{1}{2}}. \tag{4.2}$$

We can conclude from (4.2) that we have for every $\tilde\theta\in[0,1]$,

$$\mu_{\tilde\theta}^{j,0}=\mathscr{L}(Z_j^j)\text{ converges weakly as } j\to\infty. \tag{4.3}$$

This finishes the proof in the case $\sum c_k^{-1}<\infty$, the case $\sum c_k^{-1}=+\infty$ is treated in (4.22)–(4.24).

It will turn out below that there is a simple test to determine whether the system is stable or clusters in the mean-field limit.

Lemma 4.2 *The system is stable, respectively clusters, iff*

$$\int \mu_{\theta'}^\infty(d\theta)\,\theta(1-\theta)\ \begin{matrix} >0 & \text{for } \theta'\in(0,1)\\ =0 & \text{for } \theta'\in(0,1). \end{matrix} \tag{4.4}$$

To see this, note that $E^{\mu_\theta^\infty}(x)=\theta$ implies that in the second case not only $\mu_\theta^\infty(\cdot)$ has support on $\{0,1\}$ but also that $\mu_\theta^\infty(\cdot)=\theta\delta_1+(1-\theta)\delta_0$ by conservation of the mean, finishing case 2. Furthermore in the first case we know that $\mu_\theta^\infty((0,1))>0$ if $\theta\in(0,1)$. It suffices therefore to show that $\mu_\theta^\infty(\{0,1\})$ is either 0 or 1.

This last fact is now proved. Consider for every k the inhomogeneous Markov chain $(X^{(k)}_{-k}, X^{(k)}_{-k+1}, \ldots, X^{(k)}_0)$ with transition kernels $\Gamma^k_\theta(\cdot)$, $\Gamma^{k-1}_\theta(\cdot), \ldots, \Gamma^0_\theta(\cdot)$ starting in θ' at time $-k-1$. Since the transition kernels for the last, say m, time steps are the same for the processes $X^{(k)}$ with $k=m+1, m+2, \ldots$, the processes will converge if some marginal at some time point converges as $k\to\infty$. However, for each k the chain is a martingale since $E^{\Gamma^k_\theta}(x)=\theta$ for every k. Then by the backward martingale convergence theorem and by the previous observation we see that this reversed martingale $(X^{(k)}_{-j})_{j=0,\ldots,k}$ converges in distribution to a reversed martingale $(X^{(\infty)}_{-k})_{k\in\mathbf{N}}$. It remains to show that $\int \mu^\infty_{\theta'}(d\theta)\,\theta(1-\theta)>0$ implies $\mathrm{Prob}(X^\infty_0\in\{0,1\})=0$.

First observe to this end that since $\{0,1\}$ are absorbing points and since $\Gamma^k_\theta((0,1))=1\ \forall k,\ \theta\in(0,1)$ we have

$$\mathrm{Prob}(X^\infty_0\in\{0,1\})=\mathrm{Prob}(X^\infty_{-k}\in\{0,1\}) \qquad \forall k\in\mathbf{N}.$$

However for every $0<\delta<\min(1-\theta',\theta')$ we can find $\varepsilon>0$ with $\varepsilon\leqq|\theta'-\delta|\wedge|1-(\theta'+\delta)|$. Then with $Z^j_{-1}=\theta'$

$$\begin{aligned}\mathrm{Prob}(Z^j_{j-k}\in[1-\varepsilon,1]\cup[0,\varepsilon])&\leqq\mathrm{Prob}(|Z^j_{j-k}-\theta'|\geqq\delta)\\ &\leqq\frac{1}{\delta^2}\mathrm{Var}(Z^j_{j-k})\leqq\frac{1}{\delta^2}\sup_{j\geqq k}\mathrm{Var}(Z^j_{j-k}), \qquad \forall j\geqq k.\end{aligned}$$

Hence in order to conclude that $\mathrm{Prob}(X^\infty_0\in\{0,1\})=0$ it suffices to prove that $\sup_{j\geqq k}\mathrm{Var}(Z^j_{j-k})$ converges to 0 as $k\to\infty$.

From (3.1) we know

$$\int x^2\mu^{j,j-k}_\theta(dx)=F_{j+1}(\theta)\left(\sum_{\ell=k}^{j}c_\ell^{-1}\right)+\theta^2$$

or rephrased

$$\mathrm{Var}(Z^j_{j-k})=F_{j+1}(\theta)\left(\sum_{\ell=k}^{j}c_\ell^{-1}\right)\leqq\|g\|_\infty\sum_{\ell=k}^{j}c_\ell^{-1}\leqq\|g\|_\infty\sum_{\ell=k}^{\infty}c_\ell^{-1}.$$

If $\int\mu^\infty_{\theta'}(d\theta)\,\theta(1-\theta)>0$, then $F_j(\theta')=E(g(Z^j_0))$ cannot converge to 0 hence $\sum_k^\infty c_k^{-1}<\infty$ due to $\mathrm{Var}(Z^j_{j-k})\leqq 1\ \forall j,k$.

If $\sum_0^\infty c_\ell^{-1}<\infty$, then apparently above inequality gives the wanted

$$\sup_{j\geqq k}\mathrm{Var}(Z^j_{j-k})\xrightarrow[k\to\infty]{}0.$$

(b) In order to decide which of the cases in (4.4) occurs for a given sequence (c_k) we shall first consider the case $g(x)=x(1-x)$ where we can perform an explicit calculation giving us a quite enlightening direct proof. (The general case is obtained via indirect proof.)

Assume that $g(x)=x(1-x)$ and introduce

$$a_R(\theta')=\int\mu^{R,0}_{\theta'}(d\theta)\,(\theta(1-\theta)). \tag{4.5}$$

Note that $(1/4-\theta(1-\theta))$ is continuous and bounded so that by Lemma 4.1, $a_R(\theta')$ converges as $R\to\infty$ for every $\theta'\in[0,1]$. Therefore (4.4) can be rephrased as follows

(4.6) $$a_R(\theta)\xrightarrow[R\to\infty]{}0 \quad \forall\theta'\in[0,1] \Leftrightarrow \text{clustering},$$
$$a_R(\theta)\xrightarrow[R\to\infty]{}G(\theta)>0 \quad \forall\theta\in(0,1) \Leftrightarrow \text{stability}.$$

In order to calculate $a_{R+1}(\theta)$ from $a_R(\theta)$ we use the fact proved in [9], that for $F_k(x)=d_k(x(1-x))$

(4.7) $$E^{\Gamma^k_\theta}(x(1-x))=\left(\frac{1}{1+\frac{d_k}{c_k}}\right)\theta(1-\theta).$$

This means (by iteration) in combination with the recursion formula for d_k (see [9, Theorem 2]):

(4.8) $$a_R(\theta')=\prod_1^R\left(\frac{1}{1+\frac{d_k}{c_k}}\right),\qquad d_{k+1}=\frac{c_k\,d_k}{d_k+c_k}.$$

Combining (4.8) with (4.6), taking logarithms in (4.8) and then using $\log(x+1)\leqq x$ respectively $\log(1+x)\approx x$ for $|x|\to 0$ allows us to conclude that:

Lemma 4.3

(4.9) $$\textit{clustering in the mean-field limit}\Leftrightarrow\sum_k\left(\frac{d_k}{c_k}\right)=+\infty$$

(4.10) $$\textit{stability in the mean-field limit}\Leftrightarrow\sum_k\left(\frac{d_k}{c_k}\right)<\infty.$$

We will also prove the following.

Lemma 4.4

(4.11) $$\sum_k\frac{d_k}{c_k}<\infty\Leftrightarrow\sum_k\frac{1}{c_k}<\infty.$$

Assume Lemma 4.3 for the moment. The combination of (4.9), (4.10) with (4.11) would prove the assertion of the theorem for the case $g(x)=cx(1-x)$. Therefore the task remains to prove Lemma 4.4. We define $\gamma_k=d_k/c_k$. Then the relation $d_{k+1}=c_k\,d_k/(c_k+d_k)$ can be rewritten as

(4.12) $$\gamma_{k+1}=\frac{c_k}{c_{k+1}}\,\frac{\gamma_k}{\gamma_k+1}.$$

This implies in particular

(4.13)
$$\gamma_N \leqq \prod_{k=1}^{N} \frac{c_k}{c_{k+1}} = \frac{c_1}{c_{N+1}},$$

so that

(4.14)
$$\sum_{k=1}^{\infty} \gamma_k \leqq c_1 \sum_{k=1}^{\infty} \frac{1}{c_k}.$$

Next write $b_k = (c_k)^{-1}$ and rewrite (4.12):

(4.15)
$$\frac{\gamma_{k+1}}{\gamma_k}(1+\gamma_k) = \frac{b_{k+1}}{b_k}$$

or in other words

(4.16)
$$\begin{aligned} b_N &= \left[\prod_1^N \left(\frac{\gamma_{k+1}}{\gamma_k}\right)(1+\gamma_k)\right] b_1 \\ &= \left[\frac{\gamma_{N+1}}{\gamma_1} \prod_{k=1}^{N} (1+\gamma_k)\right] b_1 \leqq \frac{b_1}{\gamma_1} \gamma_{N+1} \left(\prod_{k=1}^{N} (1+\gamma_k)\right) \end{aligned}$$

so that using $\log(1+x) \leqq x$ for $x \geqq 0$ we obtain:

(4.17)
$$\sum_{k=1}^{\infty} \frac{1}{c_k} \leqq \frac{b_1}{\gamma_1} \exp\left(\sum_{k=1}^{\infty} \gamma_k\right) \sum_{k=1}^{\infty} \gamma_k.$$

Combining (4.14) and (4.17) proves immediately Lemma 4.4. This completes the proof of Theorem 3 for $g(x) = \text{const}\, x(1-x)$.

(c) In the case where $g(x)$ does not have the special form $g(x) = dx(1-x)$ we will use a different type of argument, namely proof by contradiction. The starting point of the proof is the following expression for the second moment of the interaction chain, which we derived in Sect. 3 Eq. (3.5):

(4.18)
$$E_\theta (Z_j^j)^2 = \left(\sum_{k=0}^{j} c_k^{-1}\right) F_{j+1}(\theta) + \theta^2.$$

First we show that $\sum_k c_k^{-1} < \infty$ implies stability. Here we proceed by contradiction. By Lemma 4.1 and Lemma 4.2, we then need to show that the following assumption leads to a contradiction:

(4.19)
$$\sum_k c_k^{-1} < \infty \quad \text{and} \quad \mu_\theta^{j,0} \underset{j \to \infty}{\Rightarrow} \theta \delta_1 + (1-\theta)\delta_0.$$

The relation (4.19) implies if we insert it in the lhs of (4.18), that

(4.20)
$$E_\theta (Z_j^j)^2 \xrightarrow[j \to \infty]{} \theta$$

and if we insert it in the rhs of (4.18)

(4.21) $$F_j(\theta)=E_\theta(g(Z_j^j))\xrightarrow[j\to\infty]{}0.$$

Combining (4.20) and (4.21) gives $\theta^2=\theta$ or $\theta=0$ or 1, which contradicts $\theta\in(0,1)$. Hence $\sum_k c_k^{-1}<\infty$ implies stability.

The second step of the proof is to show that $\sum_k c_k^{-1}=+\infty$ implies clustering. First note that (4.18) implies

(4.22) $$F_j(\theta)\xrightarrow[j\to\infty]{}0 \qquad \forall\theta\in(0,1).$$

On the other hand by the definition of $F_k(\cdot)$ it can be represented as

(4.23) $$F_k(\theta)=E_\theta(g(Z_k^k)).$$

Since $g(x)>0$ for $x\in(0,1)$, and $g(0)=g(1)=0$ we conclude by combining (4.22) and (4.23) and the fact that $(Z_j^k)_{j=1,\dots,k}$ is a bounded martingale, that

(4.27) $$\mathscr{L}(Z_k^k)\underset{k\to\infty}{\Rightarrow}\theta\delta_1+(1-\theta)\,\delta_0$$

or in other words $\mu_\theta^\infty((0,1))=0$ which contradicts our assumption. Hence if $\sum c_k^{-1}=+\infty$ the system clusters. This completes the proof for general g.

5 Proof of Theorem 4 and 5

(a) Proof of Theorem 4

In order to have self-similarity in the mean-field limit we need that $\Gamma_\theta^k=\Gamma_\theta^{k+1}$ $\forall k=0,\dots$. This is equivalent to ($g(x)=d_0\,x(1-x)$ here!)

(5.1) $$\frac{d_{k+1}}{c_{k+1}}=\frac{d_k}{c_k}:=\alpha \qquad \forall k\in\mathbf{N}.$$

Inserting again the recursion formula for d_{k+1} (see [9, formula (1.13)] we obtain

(5.2) $$\frac{d_k}{c_k}=\frac{1}{c_{k+1}}\,\frac{d_k\,c_k}{c_k+d_k}\Leftrightarrow\frac{c_k}{c_{k+1}}\,\frac{\alpha}{1+\alpha}=\alpha$$

so that

(5.3) $$\frac{c_k}{c_{k+1}}=1+\alpha$$

or

(5.4) $$c_k = c_0\left(\frac{1}{1+\alpha}\right)^{k-1} = c_0\left(\frac{1}{1+\frac{d_0}{c_0}}\right)^{k-1} = c_0\left(\frac{c_0}{d_0+c_0}\right)^{k-1},$$

which proves (1.28).

The next step is to determine the domain of attraction of self-similarity.

First observe that if $c_{k+1}/c_k = p \in (0, 1)$ for all $k \in N$, then formula (5.2) reads with $\gamma_k := d_k/c_k$

(5.5) $$\gamma_{k+1} = \frac{1}{p}\left(\frac{\gamma_k}{1+\gamma_k}\right).$$

The function $\mathbf{R}^+ \to \mathbf{R}^+$ defined by $x \to p^{-1}\dfrac{x}{1+x}$ is monotone, concave and has derivative p^{-1} in $x=0$ and tends to 1 as $x \to \infty$. Therefore the equation $x = p^{-1}x(1+x)^{-1}$ has a positive root x^* such that for $x > x^*$ we have $x > p^{-1}x(1+x)^{-1}$ and for $0 < x < x^*$ we have $x < p^{-1}x(1+x)^{-1}$. Since $\gamma_1 > 0$ it is straightforward to show

(5.6) $$\gamma_k \to x^* \quad \text{as} \quad k \to \infty, \qquad x^* = \frac{1}{p} - 1.$$

This proves that the (c_k) given in (1.29) with $b_{k+1}/b_k \equiv 1$ are in the domain of attraction of self-similarity. It is easy to adapt the argument to the case that $b_{k+1}/b_k \to 1$ so that "p" depends on k but converges as $k \to \infty$.

It remains to show that (1.29) gives all (c_k) which are in the domain of self-similarity. We have the relation $\gamma_{k+1} = \gamma_k(1+\gamma_k)^{-1} c_k/c_{k+1}$. Attraction to self-similarity implies that $\gamma_k \to \alpha$ as $k \to \infty$. This forces c_{k+1}/c_k to converge to $(1+\alpha)^{-1}$, which completes the proof.

(b) Proof of Theorem 5

Case 1 Let us assume for the moment $c_k \equiv c > 0$.

Define $\hat{Z}^k(\beta) = Z^k_{[\beta k]}$. We first establish in part (i) tightness of $\mathscr{L}((\hat{Z}^k(\beta))_{\beta \in [0,1]})$, second we show in part (ii) that every weak limit point is the Fisher-Wright diffusion and thereby finish the proof.

(i) The starting point is the following embedding of the processes $(Z^k_j)_{j=0,\dots,k}$ in a brownian motion.

Lemma 5.1 *There exists a brownian motion $W(t)$ (independent of everything else) and a sequence $\{S_k(\beta)\}_{\beta \in [0,1]}$ of nondecreasing processes such that $\{S_k(\beta) \leqq t\} \in \sigma(\{W(s)\}_{s \leqq t})$ and*

(5.7) $$\mathscr{L}((\hat{Z}^k(\beta))_{\beta \in [0,1]}) = \mathscr{L}((W(S_k(\beta)))_{\beta \in [0,1]}).$$

Proof. Since $(Z_n^j)_{n=-1,0,\ldots,j}$ is a martingale for every j, for every bounded convex function f the sequence $\langle \mu_\theta^{j,j-k}, f\rangle$ is nondecreasing in k. A theorem by H. Rost (see Chacon and Walsh [5]) applied to brownian motion on $[0,1]$ with absorption in 0 and 1 says the following:

Let ν, μ be two probability measures on $[0,1]$ such that

$$\langle \nu, f\rangle \leqq \langle \mu, f\rangle \qquad \text{for every } f \text{ bounded and subharmonic.}$$

Furthermore let $W^*(t)$ be brownian motion with initial distribution ν (and absorption in $\{0,1\}$). Then there exists a stopping time T of $W^*(t)$ such that

$$\mu = \mathscr{L}(W^*(T)).$$

It is easy to prove that for brownian motion on $[0,1]$ with absorption at 0 and 1:

$$f \text{ bounded and subharmonic} \Leftrightarrow f \text{ bounded and convex.}$$

By successively applying the above quoted theorem and using the fact that $W^*(t)$ is a strong Markov process with independent increments we obtain (5.7), since we can construct the desired process $S_k(\beta)$ as the sum of those stopping times T_j with $j \leqq [\beta k]$. This finishes the proof of Lemma 5.1.

In view of (5.7) we must now consider $(S_k(\beta))_{\beta\in[0,1]}$ and prove:

Lemma 5.2 *$\mathscr{L}((S_k(\beta))_{\beta\in[0,1]})$ is weakly relatively compact in the space of probability measures on $D_{[0,1]}([0,1])$.*

(For notation compare Ethier and Kurtz [10].)

Proof. First observe that with $H = \inf(t \mid W^*(t) \in \{0,1\})$ we have

$$S_k(\beta) \leqq H < \infty \qquad \text{for all } \beta\in[0,1] \quad \text{a.s.} \tag{5.8}$$

Therefore we can and shall supress the $*$. Furthermore for a stopping time T of a brownian motion with $\mathscr{L}(W(T)) = \mu$, $\mathscr{L}(W(0)) = \nu$ we have $E(T) = \int x^2\, d\mu(x) - \int x^2\, d\nu(x)$. Recall also that the successive stopping times $T_1, T_2, \ldots, T_k$ defining the embedding of $(Z_j^k)_{j=-1,\ldots,k}$ in the brownian motion $W(t)$ have the property that T_{j+1} does depend only on Z_j^k. By combining the last two facts it is a tedious but straightforward exercise to derive the following estimate: Let $(x(s)$ be a function $[0,1] \to \mathbf{R}$ and define

$$w'(x, \delta, T) = \sup_{(t_1, t, t_2)\in\mathscr{T}} (|x(t_2) - x(t)|) \wedge (|x(t_1) - x(t)|)$$

with

$$\mathscr{T} = \{(t_1, t, t_2) \mid t_1 \leqq t \leqq t_2,\ t_2 - t_1 \leqq \delta\}.$$

Then

$$P(w'(S_k(\beta)), \delta, T) \geqq x) \leqq x^{-2} (\sup_{\alpha,\theta'} \int |(\Gamma_{\theta'}^{k,k-[(\alpha+\delta)k]}(d\theta) - \Gamma_{\theta'}^{k,k-[\alpha k]}(d\theta))\, \theta^2|)^2. \tag{5.9}$$

To see this (recall that $x(\beta) = S_k(\beta)$ is nondecreasing) consider in $[t_1, t_2]$ the point t', the smallest t value in t_1, t_2 where $x(t) - x(t_1) \geqq x$. Then it suffices to study the event $\{|x(t') - x(t_1)| \wedge |x(t_2) - x(t_1)| \geqq x\}$, whose probability can now be estimated using the independence of the increments, and bounding $E|x(t') - x(t_1)| \leqq E|x(t_2)-x(t_1)|$ with Chebyshev's inequality.

Using the fact that $\hat{Z}_\beta$ is a martingale we can rewrite the relation (5.9) as follows (replacing θ^2 by $\theta(1-\theta)$):

$$(5.10)\qquad \begin{aligned} &P(w'(S_k(\beta)), \delta, T) \geqq x) \\ &\leqq x^{-2} (\sup_{\alpha, \theta'} \int (|\mu_{\theta'}^{k, k-[(\alpha+\delta)k]}(d\theta) - \mu_{\theta'}^{k, k-[(\alpha)k]}(d\theta))(\theta(1-\theta))|)^2. \end{aligned}$$

We shall prove below that the following monotone increasing functions of β

$$\{\int \mu_{\theta'}^{k, k-[\beta k]}(d\theta)\, \theta(1-\theta)\}_{k\in\mathbf{N}},$$

satisfy uniformly in θ':

$$(5.11)\qquad \begin{gathered} \int \mu_{\theta'}^{k, k-[\beta k]}(d\theta)\, (\theta(1-\theta)) \xrightarrow[k\to\infty]{} 1-\beta, \\ \int |\mu_{\theta'}^{k, k-[\beta k]} - \mu_{\theta'}^{k, k-[\beta' k]}|\, (d\theta)\, (\theta(1-\theta)) \to |\beta'-\beta|\, \theta'(1-\theta'). \end{gathered}$$

The relations (5.8), (5.10) and (5.11) can now be used to verify the assumptions in Theorem 15.6 in [4, p. 128] (with $\alpha=1$, $\gamma=1$, F linear) to prove that the family $\mathscr{L}((S_k(\beta))_{\beta\in[0,1]})$ is relatively weakly compact. The details are standard and omitted.

We shall abbreviate the sequence of functions on the lhs of (5.11) by $\{F_{\theta',k}(\beta)\}_{k\in\mathbf{N}}$. In order to verify (5.11) the next step is to calculate $F_{\theta',k}(\beta)$. The formula (5.2) becomes with the abbreviation $\gamma_k = d_k/c_k$ in the case where $c_k \equiv c_1$ simply $\gamma_{k+1} = \gamma_k(1+\gamma_k)^{-1}$. This recursive relation has the unique solution

$$(5.12)\qquad \gamma_k = \frac{1}{k} \quad \text{and consequently} \quad d_k = \frac{c_1}{k}.$$

This implies in particular (using (1.9) and (4.7)) that

$$(5.13)\qquad \int \mu_{\theta'}^{j,k}(d\theta)\, \theta(1-\theta) = \left(\prod_k^j \left(\frac{1}{1+\frac{1}{m}}\right)\right) \theta'(1-\theta').$$

Next observe that for $k=\beta_1 \ell, j=\beta_2 \ell$ for $\ell\to\infty$,

$$(5.14)\qquad \prod_k^j \left(\frac{1}{1+\frac{1}{m}}\right) \xrightarrow[\ell\to\infty]{} \beta_1/\beta_2$$

which is easily seen by taking logarithms and using $\log\left(1+\frac{1}{k}\right) \approx \frac{1}{k}$ as $k\to\infty$.

The relation (5.14), (5.13) allows us to verify the supposition (5.11). This completes the proof of Lemma 5.2.

We continue the proof of Theorem 5, in the case $g(x)=x(1-x)$, $c_k\equiv c$. Lemma 5.2 allows us to pick now a subsequence $(k_j) \subseteq \mathbf{N}$ such that $\mathscr{L}((W(t))_{t\geqq 0}, S_{k_j}(\beta)_{\beta\in[0,1]})$ converges weakly as $j\to\infty$. We denote the limit process by $((W(t))_{t\geq 0}, S_\infty(\beta)_{\beta\in[0,1]})$, and note that the process S_∞ has continuous paths by (5.11) and Theorem 12.4 in [4]. A weakly convergent sequence of càdlàg random processes converging to a continuous limit can be constructed on a common probability space converging there in the supremum norm a.s. Since $W(t)$ has continuous paths we obtain that

$$\mathscr{L}((W(S_{k_j}(\beta)))_{\beta\in[0,1]}) \underset{j\to\infty}{\Rightarrow} \mathscr{L}((W(S_\infty(\beta)))_{\beta\in[0,1]}). \tag{5.15}$$

Our task in the next subsection will be to prove that $W(S_\infty(\cdot))$ has properties which determine this process uniquely, so that as a consequence $\mathscr{L}((W(S_k(\beta)))_{\beta\in[0,1]})$ converges. Furthermore if we have shown that the time change of Fisher-Wright diffusion has these properties we have finished the proof of (1.30), (1.31) in the case $g(x)=x(1-x)$.

(ii) The main properties of $W(S_\infty(\cdot))$ are summarized in the Lemma 5.3 below.

Lemma 5.3 *The process $W(S_\infty(\cdot))$ has the following properties:*

$$W(S_\infty(\cdot)) \quad \textit{has continuous paths,} \tag{5.16}$$

$$W(S_\infty(\cdot)) \quad \textit{is a martingale,} \tag{5.17}$$

$$E^\theta(W(S_\infty(\beta))(1-W(S_\infty(\beta))))=\theta(1-\theta)(1-\beta). \tag{5.18}$$

Proof. Using the construction of S_∞ as the a.s. limit of S_k on a big probability space it can easily be verified that $S_\infty(\beta)_{\beta\in[0,1]}$ is an increasing family of stopping times with respect to a filtration $\mathscr{A}_t$ for which W is an $\mathscr{A}_t$ martingale. Therefore by the optional sampling theorem it follows that $W(S_\infty(\cdot))$ is a martingale. The continuity of paths follows from the continuity of $S_\infty(\cdot)$, (which followed easily from the estimates in (5.10) combined with the relation (5.11) and Proposition 10.3 in [10]). Altogether this proves (5.16) and (5.17).

Finally the relation (5.18) is an immediate consequence of the relations (5.14) and (5.13) once we write the rhs of (5.18) as $\theta(1-\theta)(1-\beta)$.

This finishes the proof of Lemma 5.3.

Note that the process $W(S_\infty(\beta))$ is not time-homogeneous as it stands and the question is whether this can be fixed by passing to a different time scale. The next step is therefore to introduce the following one to one transformation of the time $\beta\leftrightarrow s$ given by

$$\beta=1-e^{-s} \tag{5.19}$$

mapping $[0,1)$ onto $[0,\infty)$. In this time scale the structure of the process simplifies and we have:

Lemma 5.4 *Let $\hat{Z}^\infty(\beta)=W(S_\infty(\beta))$. Then the process $(\hat{Z}^\infty(1-e^{-s}))_{s\in[0,\infty)}$ is a continuous martingale which satisfies*

$$E^\theta(\hat{Z}^\infty(1-e^{-s})(1-\hat{Z}^\infty(1-e^{-s})))=e^{-s}\theta(1-\theta). \tag{5.20}$$

Moreover the process $(\hat{Z}^\infty(1-e^{-s}))_{s\in[0,\infty)}$ is a time homogeneous Markov process, namely, the Fisher-Wright diffusion which is generated by $\frac{1}{2}x(1-x)\left(\frac{\partial}{\partial x}\right)^2$.

Proof. The first statement follows immediately from Lemma 5.3. In order to prove that the transformed process is a Fisher-Wright diffusion we will show that it satisfies the appropriate martingale problem and then use the standard result that this martingale problem has a unique solution.

We begin by considering the mean square displacement of $\hat{Z}^k(\beta)$ between time α_1 and α_2 given that $\hat{Z}^k(\alpha_1)=\theta$. For that purpose observe first that (using (5.12))

$$(5.21)\qquad \sum_{[\alpha_1 k]}^{[\alpha_2 k]} \frac{d_j}{c_j} = \sum_{[\alpha_1 k]}^{[\alpha_2 k]} \frac{1}{j} \underset{k\to\infty}{\sim} (\log \alpha_2 k - \log \alpha_1 k) = \left(\log \frac{\alpha_2}{\alpha_1}\right).$$

Set $\alpha_2=\alpha_1+\Delta\alpha$, then the rhs becomes $\log\left(1+\frac{\Delta\alpha}{\alpha_1}\right)$. Therefore for $\alpha_2 = e^{s+\Delta s}$, $\alpha_1=e^s$ this reads $\log(1+e^{\Delta s}-1)=\Delta s$.

An elementary calculation using (5.13) combined with (5.20) and the fact above proves that the infinitesimal square displacement in s of $W(S_\infty(1-e^{-s}))$ is independent of s, and furthermore that on the event $W(S_\infty(1-e^{-s}))=\theta$ the infinitesimal square displacement is equal to

$$(5.22)\qquad \theta(1-\theta).$$

Therefore combining (5.16) and (5.22) we have the two facts:

$$(\hat{Z}^\infty(1-e^{-s}))^2 - \int_0^s \hat{Z}^\infty(1-e^{-t})(1-\hat{Z}^\infty(1-e^{-t}))\,dt \quad \text{is a martingale,}$$

$$\hat{Z}^\infty(1-e^{-s}) \quad \text{is a martingale}$$

which together comprise the martingale problem. But it is well-known (cf. Ethier and Kurtz [10, Chap. 4, Problem 24]), that this martingale problem has a unique solution, namely, the Fisher-Wright diffusion. Note that this process is also given by the unique solution of the stochastic differential equation

$$(5.23)\qquad dZ(s)=(Z(s)(1-Z(s)))^{\frac{1}{2}}\,dW(s)$$

where W is a standard Brownian motion.

Cases 2, 3 Literally the same arguments apply in these cases once we have replaced (5.12)–(5.14) and (5.21). But note that $h(k)$ has been chosen exactly in such a way that the rest of the calculation carries through.

(c) Proof of corollary to Theorem 5

The first step is to show that the results hold for all $g\in\mathscr{F}_2(x(1-x))$. For that purpose note that $\sup_x|F_k(x)/d_k\, x(1-x)-1|=\varepsilon_k$ with $\varepsilon_n=o\left(\frac{1}{n}\right)$ by assumption.

Since we have an explicit formula for $\tilde{\Gamma}_\theta^k(\cdot)$ (the $\tilde{}$ indicates that we use a specific g not equal to $x(1-x)$), namely (see [9], formula 2.3):

$$\tilde{\Gamma}_\theta^k(dx)=\frac{\text{const}}{F_k(x)}\exp\left(\int_\theta^x \frac{\theta-y}{F_k(y)}\,dy\right)dx, \tag{5.24}$$

we are able to explicitly bound the effects, on the expectation of $x(1-x)$, caused by replacing $F_k(x)$ by $d_k\, x(1-x)$ in (5.24).

Denote by $\tilde{\tilde{\Gamma}}_\theta^k(\cdot)$ the equilibrium for $g(x)=x(1-x)$ respectively with $\tilde{\tilde{\mu}}^{j,k}$ the corresponding marginals then we shall prove below that

$$\sum_{k=j_1}^{j_2} \left|\int (\tilde{\mu}_\theta^{n,k}(dx)-\tilde{\tilde{\mu}}_\theta^{n,k}(dx))\, x(1-x)\right| \Big/ \sum_{k=j_1}^{j_2} \int \tilde{\tilde{\mu}}_\theta^{n,k}(dx)\, x(1-x) \xrightarrow[n\to\infty]{} 0, \tag{5.25}$$

$$j_1=\alpha_1\, h(n) \quad j_2=\alpha_2\, h(n), \quad j_1, j_2\in[0,n]$$

as long as $h(n)\to\infty$ as $n\to\infty$. This then proves the results for general diffusion coefficient, since we can use (5.25) to repeat for any $g\in\mathscr{F}_2$ the proof we have outlined in the case $\gamma_n=1/n$ for the diffusion coefficient $g(x)=x(1-x)$. (See (5.9)–(5.11)).

In order to obtain (5.25) above, we exploit the property $g\in\mathscr{F}_2(x(1-x))$. Abbreviate

$$\tilde{a}_R(\theta'):=\int \tilde{\mu}_{\theta'}^{R,0}(d\theta)\,\theta(1-\theta) \tag{5.26}$$

$$\tilde{\tilde{a}}_R(\theta'):=\int \tilde{\tilde{\mu}}_{\theta'}^{R,0}(d\theta)\,\theta(1-\theta). \tag{5.27}$$

We shall establish below the inequalities

$$(1-\varepsilon_R)\,\tilde{\tilde{a}}_R(\theta')\leqq \tilde{a}_R(\theta')\leqq(1+\varepsilon_R)\,\tilde{\tilde{a}}_R(\theta) \tag{5.28}$$

with

$$\varepsilon_n:=\|\hat{F}^n(g)_{(x)}-x(1-x)\|.$$

Since $g\in\mathscr{F}_2(x(1-x))$ we know that $\varepsilon_n=o(1)$ and therefore (5.25) holds.

The proof of (5.28) uses again that (recall (3.1))

$$\int_0^1 x^2\,\mu_\theta^{k,0}(dx)=F_{k+1}(\theta)\left(\sum_{j=0}^k c_j^{-1}\right)+\theta^2, \tag{5.29}$$

which implies

$$\int_0^1 x(1-x)\,\mu_\theta^{k,0}(dx)=\theta(1-\theta)-F_{k+1}(\theta)\left(\sum_{j=0}^k c_k^{-1}\right). \tag{5.30}$$

This gives

$$(5.31)\quad \left|\int_0^1 x(1-x)\,(\tilde{\mu}_\theta^{k,0}(dx)-\tilde{\tilde{\mu}}_\theta^{k,0}(dx))\right| = |\tilde{F}_k(\theta)-\tilde{\tilde{F}}_k(\theta)| \sum_{j=0}^{k} c_j^{-1}$$

$$= |\tilde{\tilde{F}}_k(\theta) - d_k\,\theta(1-\theta)| \sum_{j=0}^{k} c_j^{-1}$$

$$= |\tilde{\tilde{F}}_k(\theta)/d_k - \theta(1-\theta)| \left(\sum_{j=0}^{k} c_j^{-1}\right) d_k$$

$$\leqq \varepsilon_k \left(\sum_{j=0}^{k} c_j^{-1}\right) d_k$$

$$= \varepsilon_k \int_0^1 x(1-x)\,\tilde{\tilde{\mu}}_\theta^{k,0}(dx).$$

The last relation implies now (5.28) via the monotonicity in k.

6 Proof of Theorem 6 and 7

(a) Proof of Theorem 6

The basic fact behind Theorem 6 is, that the criterion for stability or clustering of the process $X^N(t)$ on Ω^N for N fixed can be formulated in terms of the dichotomy transience versus recurrence of a particular random walk.

We shall start by introducing this random walk $(\tilde{Y}_n)_{n\in\mathbf{N}}$. First let Y_t be a continuous time random walk on the group Ω^N with jump rates

$$(6.1)\qquad \xi \to \xi':\ \sum_{k=j}^{\infty} \left(\frac{c_k}{N^k}\right)\frac{1}{N^k} \quad \text{where} \quad j=d(\xi,\xi').$$

It is convenient to write this rate in the form

$$(6.2)\qquad \frac{1}{N^{-2j}} \sum_{k=j}^{\infty} c_k\, N^{-2(k-j)}$$

and to introduce

$$(6.3)\qquad \tilde{c}_j = \sum_{k=j}^{\infty} c_k\, N^{-2(k-j)}.$$

Now let $(\tilde{Y}_n)_{n\in\mathbf{N}}$ denote the jump chain of the process $(Y_t)_{t\in\mathbf{R}^+}$. Then $\tilde{Y}_n$ is a discrete time random walk on the group Ω^N. Denote its transition kernel by $a(\xi,\xi')$ and note that this is symmetric.

Then the following holds: (See Cox and Greven [7], Theorem 1 and Theorem 2).

Proposition 6.1 *If $(\tilde{Y}_n)_{n\in\mathbb{N}}$ is transient then there exists a set $\{\nu_\theta\}_{\theta\in[0,1]}$ of invariant measures with the properties* (1.37), *while in the case where $(\tilde{Y}_n)_{n\in\mathbb{N}}$ is recurrent the only extremal homogeneous invariant measures are given by $\delta_{\{x_{\xi,0}\equiv 1\}}$ and $\delta_{\{x_{\xi,0}\equiv 0\}}$. In both cases relation* (1.39) *holds.*

We are therefore left with the task of determining the transience, respectively recurrence, properties in terms of the coefficients $(c_k)_{k\in N}$ alone. For this purpose we use Fourier-analysis and a result by Sawyer and Felsenstein [13] for symmetric random walks on the group Ω^N.

Lemma 6.1 *The random walk $(\tilde{Y}_n)_{n\in\mathbb{N}}$ is*

$$\text{(6.4)} \qquad \textit{transient if} \quad \sum_{j=1}^{\infty} \frac{1}{C_j} < \infty$$

$$\text{(6.5)} \qquad \textit{recurrent if} \quad \sum_{j=1}^{\infty} \frac{1}{C_j} = +\infty,$$

where

$$\text{(6.6)} \qquad C_j = \sum_{k=j}^{\infty} \tilde{c}_k N^{-(k-j)}.$$

Proof. Note that the number R_k of ξ' with $d(\xi,\xi')=k$ is $N^k - N(1-\delta_k(0))$ for every $\xi\in\Omega^N$. Therefore the jump rate of (Y_t) can be written as Const $R_k^{-1}(\tilde{c}_k/N^k)$ and consequently the jump chain has a transition kernel of the form

$$\text{(6.7)} \qquad \text{Const}\left(\frac{1}{R_k}\frac{\tilde{c}_k}{N^k}\right) \quad \text{for the transition } \xi\to\xi', \quad \text{if } d(\xi,\xi')=k.$$

In order to apply Sawyer's result on recurrence, respectively transience, for random-walks on Ω^N, we translate as follows:

$$\text{(6.8)} \qquad \begin{aligned} r &= N \\ R_k &= N^k - N(1-\delta_k(0)) \\ \mathrm{p}_\mathrm{k} &= \frac{\tilde{\mathrm{c}}_\mathrm{k}}{\mathrm{N}^\mathrm{k}} \\ \mathrm{f}_\mathrm{k} &= \mathrm{p}_0 + p_1 + \dots + p_{k-1} - (p_k/r - 1). \end{aligned}$$

Then recurrence (transience) of $\tilde{Y}_n$ is equivalent to (Sect. (3.4) in Sawyer and Felsenstein [13]):

$$\text{(6.9)} \qquad \sum_{1}^{\infty} \frac{r}{r^j} f_j^2/(1+f_j)(1-f_j) = \infty \; (<\infty).$$

Rewriting this gives the simpler condition ($f_j \to 1$ as $j\to\infty$)

$$\text{(6.10)} \qquad \sum_{j=1}^{\infty} r^{-(j-1)} \frac{1}{1-f_j} = \infty \; (<\infty).$$

Since

$$1-f_k=(p_{k+1}+\dots)+p_k/r-1 \tag{6.11}$$
$$=\frac{1}{N^{k+1}}\left(\sum_{j=k+1}^{\infty}\tilde{c}_j N^{-(j-k-1)}\right)+\left(\frac{\tilde{c}_k}{N^k(N-1)}\right)$$

we have

$$1-f_k \geqq \frac{1}{N^{k+1}}\left(\sum_{j=k+1}^{\infty}\tilde{c}_j N^{-(j-k-1)}\right)$$
$$1-f_k \leqq \frac{1}{N^k}\left(\sum_{j=k}^{\infty}\tilde{c}_j N^{-(j-k)}\right).$$

We rewrite (6.10) using the estimates above to get:

$$\sum_{j=1}^{\infty} r^{j-1}\frac{1}{1-f_j} \leqq N^2\left(\sum_{j=2}^{\infty}\frac{1}{C_j}\right),\quad \sum_{j=1}^{\infty} r^{j-1}\frac{1}{1-f_j} \geqq N\sum_{j=1}^{\infty}\frac{1}{C_j}. \tag{6.12}$$

This means the criterion (6.10) boils down to the condition

$$\sum_{m=1}^{\infty}\frac{1}{C_m}=\infty\,(<\infty), \tag{6.13}$$

which proves the Lemma 6.1.

Clearly Proposition 6.1 and Lemma 6.1 combined prove the Theorem 6(a), (b).

(b) Proof of Theorem 7

In order to prove Theorem 7, we will need the following duality relation between interacting Fisher-Wright diffusions and coalescing random walk system η_t.

Here are the duality relations (see Shiga [14]):

$$E_X\Big(\prod_{\xi'\in A} x_{\xi',0}(t)\Big)=E_A\Big(\prod_{\xi'\in\Omega^N} x_{\xi',0}(0)\Big)^{\eta_{\xi'}(t)}\Big) \tag{6.14}$$

where $\eta(t)=\{\eta_\xi(t)\}_{\xi\in\Omega^N}$ has the evolution (here $\eta_\xi(t)$ is interpreted as the number of particles at ξ at time t) given by the following two rules:

(i) Each particle moves indepencent of every other particle according to the random walk with jump rate for the jump $\xi\to\xi'$ given by

$$\sum_{k=j}^{\infty}\frac{c_k}{N^{2k}}\quad \text{with } j=d(\xi,\xi').$$

(ii) Each pair of particles at the same site has an exponential clock with rate d_0 and when the clock rings the two particles coalesce.

In the proof of Lemma 5.4 (see in particular (5.22)–(5.23)) we saw that it suffices to prove that the mean-square displacement of $(\theta_t)_{t\in R^+}$ is given by const$(\theta(1-\theta))$ if we are at the point θ. This is true by explicit calculation. Use the abbreviation $\tilde{t}=t+\Delta s\beta(k)$ to calculate

$$(6.15)\quad E((\theta^k(t+(\Delta s)\,\beta(k)))^2\,|\theta_t^k=\theta)-[E(\theta^k(t+\Delta s\,\beta(k))\,|\theta_t^k=\theta)]^2$$

$$=\left(\frac{1}{N^{2k}}\sum_{\xi,\,\xi'\in M_k}\hat{E}(x_{\xi,0}(\tilde{t})\,x_{\xi',0}(\tilde{t}))\right)-\theta^2$$

with $\tilde{\xi}=(0,0,\ldots)$, $M_k=\{\xi\in\Omega_N\,|\,d(\tilde{\xi},\xi)\leqq k\}$, $\hat{E}=E(\cdot\,|\theta_t^k=\theta)$

$$=\left(\frac{1}{N^{2k}}\sum_{\substack{\xi,\,\xi'\in M_k\\ d(\xi,\,\xi')=k}}\hat{E}(x_{\xi,0}(\tilde{t})\,x_{\xi',0}(\tilde{t}))\right)+o(1)-\theta^2\quad\text{as }k\to\infty.$$

To continue we shall use the duality relation with coalescing random walk. Define (see sequel of (6.14) for notation)

$$q_{\tilde{t}}^k(\xi,\xi')=\operatorname{Prob}(|\eta(\tilde{t})|=1\,|\,\eta(0)=\delta_\xi(\cdot)+\delta_{\tilde{\xi}}(\cdot)).$$

If we prove that for ξ, ξ' with $d(\xi,\xi')=k$ and $\tilde{t}=\Delta s\beta(k)$ we have

$$(6.16)\quad q_{\tilde{t}}(\xi,\xi')\xrightarrow[k\to\infty]{}q_{\Delta s}\quad\text{with }q_{\Delta s}=q\Delta s+o(\Delta s),$$

then we can continue (6.15) by calculating the terms in the sum via (6.14):

(6.17)

$$E((\theta^k(t+(\Delta s)\,\beta(k)))^2|\,\theta_t^k=\theta)-[E(\theta^k(t+\Delta s\,\beta(k))\,|\theta_t^k=\theta)]^2\xrightarrow[k\to\infty]{}q_{\Delta s}\,\theta(1-\theta).$$

Next using $q_{\Delta s}=q\Delta s+o(\Delta s)$ we obtain, that the infinitesimal mean square displacement of θ_t is given by

$$\text{const}\cdot\theta(1-\theta),$$

which finishes the proof.

It remains to verify (6.16). The key to this result is the fact that $\Omega^{N,k}$ forms an abelian group. Another way to view $\Omega^{N,k}$ is as a k-dimensional torus of size N intersected with $\mathbf{Z}^k$. The task is then to study the hitting times of points of a random walk on this object. Since rather different tools are used for this we refer to a forthcoming study of hierarchical systems without taking meanfield limits in Fleischmann and Greven, where hitting times for random walks on Ω^N are studied. The required result (6.16) is established there.

Acknowledgement. We are indebted to F. den Hollander pointing out an error in the calculation leading to the uniqueness of the fixed point. The presentation benefits from the comments of an anonymous referee.

References

1. Abramowitz, M., Stegun, I.A.: Handbook of mathematical functions. London: Dover 1965
2. Arratia, R.: Coalescing Brownian motion and the voter model on **Z**. Ph.D. dissertation. University of Wisconsin, MA (1979)

3. Baillon, J., Clement, P., Greven, A., den Hollander, F.: On the attracting orbit of a nonlinear transformation arising from renormalization of hierarchically interacting diffusions. (Manuscript 1993)
4. Billingsley, P.: Convergence of probability measures. New York: Wiley 1968
5. Chacon, R.V., Walsh, J.B.: One-dimensional potential embedding. In: Meyer, P.A. (ed.) Sém. de. Prob. X. (Lect. Notes Math., vol. 511, pp. 19–23) Berlin Heidelberg New York: Springer 1976
6. Cox, J.T., Greven, A.: On the long term behaviour of some finite particle systems. Probab. Theory Relat. Fields **85**, 195–237 (1990)
7. Cox, J.T., Greven, A.: Ergodic theorems for infinite systems of locally interacting diffusions. Ann. Probab. (to appear, 1994)
8. Cox, J.T., Griffeath, D.: Diffusive clustering in the two dimensional voter model. Ann. Probab. **14**, 347–370 (1986)
9. Dawson, D.A., Greven, A.: Multiple scale analysis of interacting diffusions. Probab. Theory Relat. Fields (to appear, 1993)
10. Ethier, S., Kurtz, T.: Markov processes, characterisation and convergence. New York: Wiley 1986
11. Lipcer, R.S., Shiryayev, A.N.: Statistics of random processes. I. General theory. (Appl. Math., vol. 5) Berlin Heidelberg New York: Springer 1977
12. Liggett, T.: Infinite particle systems. (Grundlehren Math. Wiss.) Berlin Heidelberg New York: Springer 1985
13. Sawyer, S., Felsenstein, J.: Isolation by distance in a hierarchically clustered population. J. Appl. Probab. **20**, 1–10 (1983)
14. Shiga, T.: An interacting system in population genetics. J. Math. Kyoto Univ. **20**(2), 213–242 (1980)
15. Stroock, D., Varadhan, S.: Multidimensional diffusion processes. Berlin Heidelberg New York: Springer 1979