Weierstraß-Institut

für Angewandte Analysis und Stochastik

Leibniz-Institut im Forschungsverbund Berlin e. V.

Moment asymptotics for multitype branching random walks in random environment

Onur Gün[1], Wolfgang König[1,2], Ozren Sekulović[3]

submitted: October 30, 2013

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: onur.guen@wias-berlin.de
wolfgang.koenig@wias-berlin.de

[2] Institut für Mathematik
TU Berlin
Str. des 17. Juni 136
10623 Berlin
Germany

[3] University of Montenegro
Cetinjska 2
81 000 Podgorica
Montenegro
E-Mail: ozrens@t-com.me

ABSTRACT. We study a discrete time multitype branching random walk on a finite space with finite set of types. Particles follow a Markov chain on the spatial space whereas offspring distributions are given by a random field that is fixed throughout the evolution of the particles. Our main interest lies in the averaged (annealed) expectation of the population size, and its long-time asymptotics. We first derive, for fixed time, a formula for the expected population size with fixed offspring distributions, which is reminiscent of a Feynman-Kac formula. We choose Weibull-type distributions with parameter $1/\rho_{ij}$ for the upper tail of the mean number of j type particles produced by an i type particle. We derive the first two terms of the long-time asymptotics, which are written as two coupled variational formulas, and interpret them in terms of the typical behavior of the system.

1. INTRODUCTION.

Branching processes and their applications have been studied for a long time in the communities of mathematics, physics and biology. Additional structure and features like (1) spatiality (random migration of the particles), (2) several types of particles, or (3) dependence of the branching rates on the space and on an additional independent random input lead to variants called (1) *branching random walks*, (2) *multitype branching processes* and (3) *branching processes in random environment*, respectively. They have attracted researchers because of the additional mathematical richness that they bring into the model and also because of a greater degree of applicability to real-world phenomena. For example, multitype branching processes, where particles are of several types and can give birth to particles of a type other than their own, are motivated by a rich class of applications in physics and biology, such as cosmic ray cascades, bacterial populations and cancer research (see the books [M71], [J75] and [KA02] for various examples).

In this paper, we study a model that exhibits all the three features mentioned above; a *multitype branching random walk in random environment*, where particles move independently following a Markov chain on a finite state space and multitype branching takes place according to random site-dependent distributions, which are static, i.e., fixed throughout the evolution. The particles can have a finite number of types, and particles of each type can give birth to new particles of any type. To the best of our knowledge, this combination of features has not yet been considered in the mathematical literature.

One of the most fundamental objects to study is the total number of particles in the system, the global population size, at a given time n in the limit as $n \to \infty$. The two main goals of the present paper are (1) a formula for its expectation (taken over branching/killing and migration, but not over random offspring distributions) for fixed n, and (2) the description of the large-n asymptotics of its expectation over the branching probabilities in terms of two coupled variational formulas and their interpretation in terms of a pathwise behavior of the branching process. Hence, we are concerned with the *annealed* setting. Note that the corresponding quenched setting is not interesting, since we decided (to keep the technical difficulties low) to work on an arbitrary finite state space, and therefore there is only a fixed number of random branching probabilities; no effect from infinitely many random inputs is present.

Fixing n, for multitype processes without spatiality, it is standard knowledge [AN72] that the expectation of the population size over the branching/killing for fixed branching probabilities can be described in terms of the n-th power of a certain characteristic matrix, the *mean-offspring matrix*. For

branching random walks, it can be described by the discrete-time version of a Feynman-Kac-type formula. But the question is what structure arises in the case of a multitype branching random walk, i.e., when these two features are combined. Our first result gives the answer: it can be represented in both ways, but on an enlarged space, the product of the state space and the type space. However, there are some subtle differences, one of which is the way in which the branching probabilities are attached to the two spaces: actually, they are attached to the *sites* of the state space and to the *edges* of the type space, at least in the particular model that we study.

Turning now to the description of the large-n asymptotics of the annealed expectation of the global population size, we therefore have the choice between working on the expectation of the n-th power of a certain random matrix or on the expectation of a Feynman-Kac-like formula. Each of the two settings gives rise to an interesting proof. We decided to work on the latter and to comment on the former only in an informal manner, since it appears to us as if making this line of arguments rigorous would be technically much more involved. We choose the distribution of the branching probabilities as a *Weibull-type distribution*, as this distribution exhibits an interesting competition between the migration and the branching probabilities, such that an appealing picture arises. Working on the expectation of the Feynman-Kac-like formula requires the application of a standard large-deviation principle for the empirical pair measures of the underlying Markov chain and goes through quite smoothly. The arising two coupled variational formulas give rise to a deeper understanding of the main branching process trees, i.e., of those that give the main contribution to the expected population size.

Let us give some comments on the existing literature on branching processes in a random environment. One of the models that have been studied before is multitype branching processes with environments varying in time. To name a few, in [T81], various classification results depending on the long time behavior of the multiplications of mean offspring matrices have been proved, and a much finer analysis in a very general set up has been done in [BCN99] using harmonic functions of multiplication of mean matrices. Let us proceed with spatial branching processes in random environment. Branching discrete-time random walks on $\mathbb{Z}^d$ with time-space i.i.d. offspring distributions were studied in the context of survival properties, global/local growth rates and diffusivity; and their connections with directed polymers in random environment, see e.g. [BGK05, Y08, CY11]. Detailed analyses of recurrence/transience properties of discrete-time branching Markov chains with only space-dependent environment, which does not exhibit in general the the usual dichotomy valid for irreducible Markov chains, were carried out in [CMP98, MP00, MP03, CP07, M08, BGK09, GMPV10], to mention some. The main techniques in these studies relate these models to the better-known random walk in random environments, using the spectral properties of underlying Markov process and studying the embedded Galton-Watson processes in random environment.

The remainder of Section 1 is organized as follows. In Section 1.1 we introduce the branching process in a fixed medium and give representations of the main object of our interest, the global particle number, in Section 1.2. The random environment is introduced in Section 1.3, and our main result is presented in Section 1.4. We comment and interpret it in Section 1.5, where we in particular analyse the main quantities appearing in the main result. The special case where migration is dropped leads to even more explicit formulas, which we present in Section 1.6. Finally, in Section 1.7 we phenomenologically discuss and compare another approach to the main result in terms of Frobenius eigenvalue

theory for a random matrix, which gives some interesting insights. The proofs of all our results are in Section 2.

1.1 The Model.

The model we will study is a multitype branching Markov chain on a finite state space in discrete time in a fixed environment of branching probabilities. (This environment will be taken random in Section 1.3 below.)

Let $(X_n)_{n\in\mathbb{N}_0}$ be an irreducible Markov chain on a finite state space $\mathcal{X}$ with transition matrix $P = (P_{xy})_{x,y\in\mathcal{X}}$. Let $\mathcal{T}$ be a finite set, the set of types. We equip $\mathcal{T}$ with a set $\mathcal{A}$ of directed edges $(i,j) \in \mathcal{T} \times \mathcal{T}$ and obtain a directed finite graph $\mathcal{G} = (\mathcal{T}, \mathcal{A})$. We assume that each directed edge appears at most once in $\mathcal{A}$, and for each $i \in \mathcal{T}$, there is at least one $j \in \mathcal{T}$ such that $(i,j) \in \mathcal{A}$. Self-edges (i,i) may appear in $\mathcal{A}$. Finally, we assume $\mathcal{G}$ is a connected graph.

To each $y \in \mathcal{X}$ we attach a matrix $F_y = (F_y^{(i,j)})_{(i,j)\in\mathcal{A}}$ of probability distributions on $\mathbb{N}_0$, the environment. Given $F = (F_y)_{y\in\mathcal{X}}$, we define a discrete-time Markov process $(\eta_n)_{n\in\mathbb{N}_0}$ on $\mathbb{N}_0^{\mathcal{T}\times\mathcal{X}}$, where $\eta_n(i,x)$ is the number of particles of type i at site x at time n. The environment F does not depend on time and is fixed throughout the evolution of particles. We specify the transition mechanism of $(\eta_n)_{n\in\mathbb{N}_0}$ as follows: given that the configuration is equal to η at time n, during the time interval $(n, n+1)$,

(i) a particle of type i located at site $y \in \mathcal{X}$ produces, independently for $j \in \mathcal{T}$ such that $(i,j) \in \mathcal{A}$, precisely k particles of type j at the same site y with probability $F_y^{(i,j)}(k)$, for any $k \in \mathbb{N}_0$. All offspring productions are independent over all the particles in $\mathcal{X}$ and over time $n \in \mathbb{N}_0$,

(ii) immediately after creation, each new particle at x chooses a site y with probability P_{xy} and moves there. All jumps are independent over all the particles in $\mathcal{X}$ and over time $n \in \mathbb{N}_0$.

The resulting particle configuration is η_{n+1}. Note that, unlike in the most general set up of multitype branching processes, we assume that a particle of type i produces particles of type j independently in j, that is, the offspring distribution coming from an i type particle is in a product form. Finally, one can immerse the spatial movement into a multitype setting by simply adopting the spatial points as types and rewrite offspring distributions, this time involving terms from the transition matrix P. We will comment more on this connection later.

For definiteness, we consider localized initial conditions in $\mathcal{X}$ and $\mathcal{T}$. To this end, fix a site $y \in \mathcal{X}$ and type $j \in \mathcal{T}$. We start the Markov chain $(\eta_n)_n$ with the initial configuration $\eta_0(i,x) = \delta_j(i)\delta_y(x)$, and by $\mathrm{P}_{j,y}$ and $\mathrm{E}_{j,y}$ we denote its distribution and expectation, respectively. Note that they depend on the realization of the environment F.

We are interested in the expectation of the global population size, $|\eta_n| := \sum_{i\in\mathcal{T},x\in\mathcal{X}} \eta_n(i,x)$,

$$u_n(i,x) := \mathrm{E}_{i,x}[|\eta_n|], \qquad n \in \mathbb{N}_0, x \in \mathcal{X}, i \in \mathcal{T}. \tag{1.1}$$

Note that the expectation is taken only on the migration, and the branching/killing, but the environment F is kept fixed.

1.2 Representations of the expected particle number.

Our analysis of the population size is based on a description that is reminiscent of the Feynman-Kac representation of the solution to the heat equation with additive potential. To formulate this, we need to introduce a Markov chain $T = (T_n)_{n \in \mathbb{N}_0}$ on the type space $\mathcal{T}$ with transition probabilities

$$p_{ij} = \frac{\mathbb{1}\{(i,j) \in \mathcal{A}\}}{\deg^+(i)}, \qquad i, j \in \mathcal{T}, \tag{1.2}$$

where $\deg^+(i) = |\{k \in \mathcal{T} : (i,k) \in \mathcal{A}\}|$ is the outdegree of i. We define T and X independently on a common probability space and write $\mathbb{P}_{i,x}^{(T,X)}$ and $\mathbb{E}_{i,x}^{(T,X)}$ for probability and expectation, respectively, where T starts from i and X from x. We denote by $m_{ij}(y) = \sum_{k \in \mathbb{N}_0} k F_y^{(i,j)}(k)$ the expectation of $F_y^{(i,j)}$ (the mean number of j type particles at site y that an i type particle at site y produces in one generation) where we set $m_{ij}(y) = 0$ for $(i,j) \notin \mathcal{A}$.

The first of the two following representations for u_n is the announced Feynman-Kac-type formulation, which we will use for our proofs in Section 2.2, and the second one is in terms of the n-th power of a particular matrix, which we will use in our heuristic explanations in Section 1.7.

Proposition 1.1. *For any $i \in \mathcal{T}$ and any $x \in \mathcal{X}$ and any $n \in \mathbb{N}_0$,*

$$u_n(i,x) = \mathbb{E}_{i,x}^{(T,X)} \Big[\prod_{l=1}^n \big(m_{T_{l-1}T_l}(X_{l-1}) \deg^+(T_{l-1}) \big) \Big], \tag{1.3}$$

$$= \sum_{j \in \mathcal{T}, y \in \mathcal{X}} B_{(i,x),(j,y)}^n, \tag{1.4}$$

where B^n is the nth power the $(\mathcal{T} \times \mathcal{X}) \times (\mathcal{T} \times \mathcal{X})$ matrix B with coefficients

$$B_{(i,x),(j,y)} = m_{ij}(x) P_{xy} \mathbb{1}\{(i,j) \in \mathcal{A}\}. \tag{1.5}$$

As mentioned in the previous section, one can enlarge the type space to $\mathcal{T} \times \mathcal{X}$, that is a particle of type $i \in \mathcal{T}$ at site $x \in \mathcal{X}$ can be viewed as a particle of type (i,x). Hence, our model can be seen as a multitype branching process on the finite type space $\mathcal{T} \times \mathcal{X}$. For a general multitype branching process with finitely many types it is very well-known that the expected number of particles at generation n can be described in terms of the entries of the n-th power of the mean matrix of one generation. In our model, first a particle produces offsprings and then they migrate along the spatial space. As a result, in the enlarged type space of $\mathcal{T} \times \mathcal{X}$ the mean number of (j,y) offsprings produced by an (i,x) type particle is simply $m_{ij}(x) P_{xy} \mathbb{1}\{(i,j) \in \mathcal{A}\}$ which leads to formula in (1.4) with the mean matrix B given as in (1.5).

An interpretation of formula (1.3) is that it is a change of the order of integration. More precisely, one can write the expectation $u_n(i,x)$ as a summation over paths $(x = x_0, i = t_0), (x_1, t_1), \ldots, (x_n, t_n)$ in $T \times \mathcal{X}$ of length n, the expectation of the number of such paths one can find in the n-level branching tree of the process. Clearly, this expectation is given by the term $\prod_{l=1}^n m_{t_{l-1}, t_l}(x_{l-1}) P_{x_{l-1} x_l}$. Finally, by adding the terms $\deg^+(t_{l-1})$, one can make turn the summation into an expectation over $\mathbb{P}_{i,x}^{(T,X)}$.

Writing $\prod_{l=1}^n \cdots$ as $\exp(\sum_{l=1}^n \log \cdots)$ in the expectation in (1.3), we encounter a discrete-time version of a Feynman-Kac formula for the Markov chain (T, X) on $\mathcal{T} \times \mathcal{X}$, however with an interesting

difference: the potential $\log m_{i,j}(x)$ depends on the *vertices* of the space $\mathcal{X}$, but on the *edges* of the type space $\mathcal{T}$. A further (more or less negligible) difference is the appearance of the degree term, which accounts for the missing probability structure that we had *a priori* on $\mathcal{T}$ and artificially inserted; this term will drop out in the end. Finally, note that Proposition 1.1 reveals that our model can be seen as a version of a branching random walk on the enlarged spatial space $\mathcal{T} \times \mathcal{X}$.

1.3 The random environment.

Let us describe our assumptions on the random environment. We assume that the collection of all distributions $F_y^{(i,j)}$ with $y \in \mathcal{X}$ and $(i, j) \in \mathcal{A}$ is independent. Their distribution depends on (i, j), but not on y. We call $F = (F_y)_{y \in \mathcal{X}}$ the *random environment* and denote by Prob and $\langle \cdot \rangle$ probability and expectation with respect to F, respectively. Note that, $m_{ij}(y)$ is a random variables whose distribution is induced by $F_y^{(i,j)}$. Since we are here interested only in the expectation of the global particle number, we will make our assumptions on the environment only in terms of the quantities $m_{ij}(y)$. In particular, we assume that the collection of the $m_{ij}(y)$ is independent in $y \in \mathcal{X}$ and $i, j \in \mathcal{T}$. Then, it is our goal to find the first two terms of the large-n asymptotics of the expectation of $u_n(i, x)$, i.e., the annealed moments of the global population size when the system starts from one i type particle at $x \in \mathcal{X}$.

One can already guess from (1.3) that the large-time asymptotics of the branching process does not depend on characteristics like expectation or variance of the offspring expectation, but predominantly on their upper tails, since u_n is basically a product of high powers of them. We will study the case where $m_{ij}(y)$ lies, in terms of upper tails, in the vicinity of the *Weibull distribution* with parameter $1/\rho_{ij} \in (0, \infty)$, i.e.,

$$\mathrm{Prob}(m_{ij}(y) > r) \approx \exp\{-r^{1/\rho_{ij}}\}, \qquad r \to \infty. \tag{1.6}$$

In the language of [GM98], $\log m_{ij}(y)$ lies in the vicinity of the double-exponential distribution, which is nothing but a reflected Gumbel distribution. The precise assumption on $m_{ij}(y)$ can be written down in terms of the logarithmic moment generating functions given by

$$H_{ij}(t) := \log\langle m_{ij}(y)^t \rangle, \qquad t > 0, \quad i, j \in \mathcal{T}. \tag{1.7}$$

Via Tauberian theorems, the upper tails of $m_{ij}(x)$ stand in a one-to-one relation with the regularity of the moment generating function at infinity. Therefore, we sharpen the assumption (1.6) by requiring the following:

Assumption 1.2. *For any* $(i, j) \in \mathcal{A}$ *there exists* $\rho_{ij} \in (0, \infty)$ *such that*

$$\lim_{t \to \infty} \frac{H_{ij}(ct) - cH_{ij}(t)}{t} = \rho_{ij} c \log c, \qquad c \in (0, 1). \tag{1.8}$$

For $(i, j) \notin \mathcal{A}$, we put $\rho_{ij} = 0$. Hence, our environment distribution is characterized by the matrix-valued parameter $\rho = (\rho_{ij})_{i,j \in \mathcal{T}}$. The larger ρ_{ij} is, the thicker the tails of $m_{ij}(y)$ are, i.e., the easier it is for $m_{ij}(y)$ to achieve extremely high values.

Assumption 1.2 was used in a number of papers on the parabolic Anderson model on $\mathbb{Z}^d$ (see [GM98], e.g.) in a similar context. By the virtue of (1.8), in our main result we will see *two* explicit terms of the asymptotics of the expectation of u_n, both of which describe interesting aspects of the long-time behavior of the branching process. The fact that the parameter t appears both in the arguments of

H in the numerator and in the denominator is the reason that this distribution class is particularly amenable to an interesting asymptotics of the annealed asymptotics, since it leads to a match of the large deviation scales of the probability with respect to the Markov chain and the second order term coming from the random medium. For any other regularity assumption on H, it would not be possible to match these two scales. This is why we found the Weibull distribution particularly suitable. Of course, this would change if we would work on $\mathbb{Z}^d$ or $\mathbb{R}^d$ instead of a fixed finite state space, and many interesting additional questions would arise from the unboundedness of the space, but this is not the focus in the present paper.

1.4 The main result.

We are now heading towards a formulation of our main result on the asymptotics of the annealed moments of the global particle number. For any discrete set S, we denote by $\mathcal{M}_1(S)$ the set of probability measures on S and by $\mathcal{M}_1^{(s)}(S^2)$ the set of probability measures on S^2 with equal marginals. The first quantity of interest is

$$\lambda(\rho) = \sup\left\{\langle \mu, \rho \rangle : \mu \in \mathcal{M}_1^{(s)}(\mathcal{T}^2)\right\}, \qquad \text{where } \langle \mu, \rho \rangle = \sum_{(i,j)\in\mathcal{A}} \mu(i,j)\rho_{ij}, \qquad (1.9)$$

and the set of the corresponding maximizers:

$$\Lambda(\rho) := \left\{\mu \in \mathcal{M}_1^{(s)}(\mathcal{T}^2) : \langle \mu, \rho \rangle = \lambda(\rho)\right\}. \qquad (1.10)$$

We introduce some notation. Each measure $\nu \in \mathcal{M}_1^{(s)}((\mathcal{T} \times \mathcal{X})^2)$ has a number of marginal measures that are defined on different spaces, but in order to keep the notation simple, we denote by $\overline{\nu}$ all these marginals, namely,

$$\overline{\nu}(i,j,x) = \sum_{y\in\mathcal{X}} \nu((i,x),(j,y)), \qquad \overline{\nu}(i,x) = \sum_{j\in\mathcal{T}} \overline{\nu}(i,j,x),$$

$$\overline{\nu}(i,j) = \sum_{x\in\mathcal{X}} \overline{\nu}(i,j,x), \qquad \overline{\nu}(i) = \sum_{j\in\mathcal{T}} \overline{\nu}(i,j). \qquad (1.11)$$

To describe the second term in the asymptotics, we need to introduce two functionals on measures $\nu \in \mathcal{M}_1^{(s)}((\mathcal{T} \times \mathcal{X})^2)$, an energy functional $\mathcal{S}$ and an entropy functional $\mathcal{I}$. Indeed, define

$$\mathcal{S}(\nu) \quad := \quad \sum_{(i,j)\in\mathcal{A}} \rho_{ij} \sum_{x\in\mathcal{X}} \overline{\nu}(i,j,x) \log \overline{\nu}(i,j,x) + \sum_{(i,j)\in\mathcal{A}} \overline{\nu}(i,j)\rho_{ij} \log \rho_{ij}, \qquad (1.12)$$

$$\mathcal{I}(\nu) \quad := \quad \sum_{i,j\in\mathcal{T}} \sum_{x,y\in\mathcal{X}} \nu((i,x),(j,y)) \log \frac{\nu((i,x),(j,y))}{\overline{\nu}(i,x)P_{xy}\mathbb{1}\{(i,j)\in\mathcal{A}\}}. \qquad (1.13)$$

We set $\mathcal{I}(\nu) = \infty$ if ν is not absolutely continuous with respect to the measure $((i,x),(j,y)) \mapsto \overline{\nu}(i,x)P_{xy}\mathbb{1}\{(i,j)\in\mathcal{A}\}$. Then $\mathcal{I}(\nu)$ is equal to the entropy of ν with respect to this measure; note that it is not normalized, but has mass equal to $\sum_{i\in\mathcal{T}} \overline{\nu}(i)\deg^+(i)$.

Now we can state our main result:

Theorem 1.3. *Under Assumption 1.2 for any $i \in \mathcal{T}$ and $x \in \mathcal{X}$, as $n \to \infty$,*

$$\langle u_n(i,x) \rangle = (n!)^{\lambda(\rho)} \mathrm{e}^{-n\chi(\rho)} \mathrm{e}^{o(n)} = \exp\left(\lambda(\rho)n \log \frac{n}{\mathrm{e}} - n\chi(\rho) + o(n)\right), \qquad (1.14)$$

where

$$\chi(\rho) = \inf\left\{\mathcal{I}(\nu) - \mathcal{S}(\nu)\colon \nu \in \mathcal{M}_1^{(s)}\big((\mathcal{T} \times \mathcal{X})^2\big), \overline{\nu} \in \Lambda(\rho)\right\}. \tag{1.15}$$

The proof of Theorem 1.3 is in Section 2.2. We proceed with some comments on this proof and the interpretation of the formula. Starting from the representation in (1.3) in Proposition 1.1, we follow the patterns of [GM98], however with some notable changes. The main step is rewriting the Feynman-Kac representation in terms of the *empirical pair measure*

$$\nu_n = \frac{1}{n}\sum_{l=1}^n \delta_{((T_{l-1},X_{l-1}),(T_l,X_l))}, \tag{1.16}$$

which is the central object in this approach. In terms of the space-type random walk (X,T), the number $n\nu_n((i,x),(j,y))$ plays the role of the number of j type offspring of any i type particle located at x by time n that makes a step to y right after creation. Hence, ν_n stands for the union of all n-step paths $((X_0,T_0),\ldots,(X_n,T_n))$ that make precisely $n\nu_n((i,x),(j,y))$ steps $(i,x) \to (j,y)$ for every $i,j \in \mathcal{T}$ and every $x,y \in \mathcal{X}$. The term $\mathcal{I}(\nu)$ is the negative exponential rate of the probability of this union under the Markov chain X, together with the combinatorial complexity of the trajectories of types, and $\mathcal{S}(\nu)$, together with the leading term $\lambda(\rho)$, is the one under the expectation w.r.t. the random environment under Assumption 1.2.

1.5 Discussion of the variational formulas.

Theorem 1.3 in particular shows that the main contribution to the annealed moments of the particle numbers, $\lambda(\rho)$, comes from those n-step branching process subtrees which produce, for some $\mu \in \Lambda(\rho)$, at approximately $n\mu(i,j)$ of the n steps a number of j type particles from one or more i type particles, for any $i,j \in \mathcal{T}$. Then the value $\langle \mu, \rho \rangle$ gives the leading contribution on the scale $n \log \frac{n}{e}$. It is interesting to note that the optimality of the leading term has nothing to do with the spatial part of the branching process, but only with the creation of particles. The reason is that all the probabilities of spatial actions, i.e., of the random walk X, are on the scale n, but the values of the offspring expectations $m_{ij}(x)$, averaged over the environment are typically on the scale $n^{O(1)}$ under Assumption 1.2.

The interpretation of the second-order term is that, for any maximizer ν of $\mathcal{S} - \mathcal{I}$ satisfying $\overline{\nu} = \mu$, the main contribution comes from those n-step branching process trees that place all the births of j type particles from i type particles in such a way on $\mathcal{X}$ that approximately $n\nu((i,x),(j,y))$ such births take place at x, and the newly created particle immediately jumps to y, for any $i,j \in \mathcal{T}$ and any $x,y \in \mathcal{X}$.

In this light, let us analyse the leading term $\lambda(\rho)$ a bit more closely. A *simple cycle* on $\mathcal{G}$ is a path $\gamma = (i_1,\ldots,i_l,i_{l+1})$ in $\mathcal{T}$, with steps (i_m,i_{m+1}) in $\mathcal{A}$, that begins and ends at the same vertex $i_1 = i_{l+1}$, but otherwise has no repeated vertices or edges. We write $(i,j) \in \gamma$ if the directed edge (i,j) belongs to γ, that is, if $(i,j) = (i_m,i_{m+1})$ for some $m \in \{1,\ldots,l\}$. We call $|\gamma| = l$ its *length*. We denote by Γ_l the set of all simple cycles of length l and by Γ the set of all simple cycles. We define

$$\mu_\gamma(i,j) = \begin{cases} 1/|\gamma| & \text{if } (i,j) \in \gamma, \\ 0 & \text{otherwise.} \end{cases} \tag{1.17}$$

It is clear that $\mu_\gamma \in \mathcal{M}_1^{(s)}(\mathcal{T}^2)$ for any $\gamma \in \Gamma$. Simple cycles are important for the asymptotics of the annealed moments because the set of extremes of $\mathcal{M}_1^{(s)}(\mathcal{T}^2)$ consists exactly of the simple cycles of the graph $\mathcal{G}$. Since we could not find a proper reference for this fact, we formulate it as a lemma and prove it in Section 2.3.

Lemma 1.4. *The set of extremes of the convex set $\mathcal{M}_1^{(s)}(\mathcal{T}^2)$ is equal to $\{\mu_\gamma \colon \gamma \in \Gamma\}$.*

Since the optimization problem in (1.9) is a linear optimization problem on the convex, compact set $\mathcal{M}_1^{(s)}(\mathcal{T}^2)$, the Krein-Milman theorem and Lemma 1.4 imply the following characterization of the leading term in (1.14):

Lemma 1.5.

$$\lambda(\rho) = \max\left\{\langle\mu_\gamma, \rho\rangle; \gamma \in \Gamma\right\} = \max\left\{\frac{1}{|\gamma|}\sum_{m=1}^{|\gamma|} \rho_{i_m i_{m+1}}; (i_1, \ldots, i_{|\gamma|+1}) \in \Gamma\right\}. \tag{1.18}$$

The interpretation of Lemma 1.5 is that the leading contribution to the expected population size comes from optimal cycles $(i_1, \ldots, i_{|\gamma|}) \in \Gamma_{|\gamma|}$ in the sense that already all those n-step branching process trees contribute alone optimally which produce only i_{m+1} type particles from i_m type particles for any $m \in \{1, \ldots, |\gamma|\}$ (with $i_{|\gamma|+1} = i_1$), but essentially no other offspring.

1.6 Dropping the migration.

In this section we illustrate our result in Theorem 1.3 in the special case where migration is absent, i.e., a multitype branching model in a random environment of branching rates without any reference to a spatial component. Here we can give a more explicit description of the annealed asymptotics. Formally, we drop the spatial component of the model by picking a trivial Markov chain on $\{x\}$ with $P_{xx} = 1$ for some x and remove it from the notation. We analyse the decisive quantities $\lambda(\rho)$, $\Lambda(\rho)$ and $\chi(\rho)$ defined in (1.9) and (1.10). The functionals $\mathcal{S}$ and $\mathcal{I}$ are accordingly modified as follows. For $\nu \in \mathcal{M}_1^{(s)}(\mathcal{T}^2)$ let

$$\mathcal{S}(\nu) := \sum_{(i,j)\in\mathcal{A}} \rho_{ij}\nu(i,j)\log\nu(i,j) + \sum_{(i,j)\in\mathcal{A}} \nu(i,j)\rho_{ij}\log\rho_{ij}, \tag{1.19}$$

$$\mathcal{I}(\nu) := \sum_{(i,j)\in\mathcal{A}} \nu(i,j)\log\frac{\nu(i,j)}{\overline{\nu}(i)} \tag{1.20}$$

where $\overline{\nu}(i) = \sum_j \nu(i,j)$. Let

$$\Gamma(\rho) = \left\{\gamma \in \Gamma \colon \mu_\gamma \in \Lambda(\rho)\right\} = \left\{\gamma \in \Gamma \colon \langle\mu_\gamma, \rho\rangle = \lambda(\rho)\right\}. \tag{1.21}$$

We restrict to the case where $\rho_{ij} = \rho_i \geq 1$ for all i, j. We give an explicit solution to the variational formula in (1.15) that defines $\chi(\rho)$. In the even more restricted case where $\rho_{ij} = \rho \in (0, \infty)$ for all i, j, we have $\lambda(\rho) = \rho$ and $\Lambda(\rho) = \mathcal{M}_1^{(s)}(\mathcal{T}^2)$. Let $l_{\min}$ be the girth of the directed graph $\mathcal{G}$, that is, the length of a shortest simple cycle of the graph.

Lemma 1.6. *Let $\rho_{ij} = \rho_i \geq 1$ for all j. Then*

$$\chi(\rho) = \min \left\{ \lambda(\rho) \log |\gamma| - \sum_{m=1}^{|\gamma|} \frac{1}{|\gamma|} \rho_{i_m} \log \rho_{i_m} : \gamma \in \Gamma(\rho) \right\}. \tag{1.22}$$

More specifically, if $\rho_i = \rho \geq 1$ for all i, then $\chi(\rho) = \rho \log l_{\min} - \rho \log \rho$, and the set of minimizers in (1.22) is equal to $\Gamma_{l_{\min}}$.

1.7 Comparison to a Frobenius eigenvalue approach.

To analyse the annealed moment asymptotics of the particle number in the spatial multitype branching process in random environment, another approach is also very tempting. We want to roughly explain this briefly for the simpler case where migration is absent, i.e., in the setting of Section 1.6. (The general case is not much different, thanks to (1.4), but notationally more cumbersome.) In this case, (1.3) (or Lemma 2.1 below) simplifies to

$$u_n(i) = \sum_{j \in \mathcal{T}} \left(M^n \right)_{ij}, \qquad i \in \mathcal{T}, n \in \mathbb{N}_0, \tag{1.23}$$

where we recall that the left-hand side is the expected number (taken only over the branching/killing) of particles at time n when we started with just one i type at time 0, for fixed environment, and we write $M = (m_{ij})_{i,j \in \mathcal{T}}$ for the expectation matrix, where m_{ij} is the expectation of the number of j type offspring of a i type particle. For simplicity, we assume that $\mathcal{A} = \mathcal{T}^2$, i.e., that Assumption 1.2 holds for every $(i, j) \in \mathcal{T}^2$. Assume that, almost surely, M is irreducible, that is, for any $i, j \in \mathcal{T}$ there is an $n \in \mathbb{N}$ such that $(M^n)_{ij} > 0$. Then the high powers of M can be approached with the help of the *Frobenius eigenvalue* as follows. Define, for any positive irreducible matrix $A = (a_{i,j})_{i,j \in \mathcal{T}}$,

$$\mu(A) = \lim_{k \to \infty} \frac{1}{k} \log \sum_j (A^k)_{i,j}, \tag{1.24}$$

and note that this limit exists and does not depend on i [S06]. Furthermore, $\mathrm{e}^{\mu(A)}$ is equal to the largest eigenvalue of A, the Frobenius eigenvalue of A, which has also the characteristic property that it is a simple eigenvalue both algebraically and geometrically. Hence, we can approximate $\langle u_n(i) \rangle \approx \langle \mathrm{e}^{n\mu(M)} \rangle$. This means that we are faced with the question of a large-deviation principle for the random vector M.

We consult Assumption 1.2 and see that, for each $(i, j) \in \mathcal{T}^2$, the variable $\log m_{ij} - H_{ij}(n)/n$ satisfies an LDP with speed n and rate function $\mathbb{R} \ni m \mapsto \frac{\rho_{ij}}{\mathrm{e}} \mathrm{e}^{m/\rho_{ij}}$. This is easily calculated from (1.6) by using that $H_{ij}(n) = \rho_{ij} n \log(\rho_{ij} n) - \rho_{ij} n + o(n)$ for $n \to \infty$. By independence, the entire matrix $K_n = \log M - H(n)/n = (\log m_{ij} - H_{ij}(n)/n)_{i,j \in \mathcal{T}}$ satisfies an LDP on $\mathbb{R}^{\mathcal{T}^2}$ with speed n and rate function $m \mapsto \prod_{i,j} \frac{\rho_{ij}}{\mathrm{e}} \mathrm{e}^{m_{ij}/\rho_{ij}}$. We have $m_{ij} = \mathrm{e}^{K_n(i,j)} \mathrm{e}^{H_{ij}(n)/n}$ for each $i, j \in \mathcal{T}$. In order to identify the asymptotics of $\langle \mathrm{e}^{n\mu(M)} \rangle$, one needs to employ the LDP that K_n satisfies, but it appears difficult to write the Frobenius eigenvalue of M in terms of the one of K_n. For making decisive progress here, it seems as if one must apply some tools employed in the present paper to the formula (1.24) as follows: Explicitly write out the n-fold product of the matrix A, write it in terms of

the empirical pair measure of the resulting multi-index and use the well-known combinatorics for the number of strings that lead to a given empirical pair measure, to get that

$$\mu(A) = \sup_{\nu \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)} \left(\langle \nu, \log A \rangle - \sum_{i,j \in \mathcal{T}} \nu(i,j) \log \frac{\nu(i,j)}{\overline{\nu}(i)} \right), \tag{1.25}$$

where $\langle \nu, \log A \rangle = \sum_{i,j \in \mathcal{T}} \nu(i,j) \log a_{ij}$, and $\overline{\nu}$ denotes the marginal measure of ν. It appears also not easy to extract the precise dependence of the leading term $\mu((\rho n/\mathrm{e})^\rho)$ on n without this procedure. Summarizing, we believe that the method that we use in the present paper is essentially the only doable way.

2. PROOFS OF THE MAIN RESULTS.

2.1 Proof of Proposition 1.1.

We now write the expected number of offspring in terms of an expectation of the product of the expectation matrices $M(x) = (m_{ij}(x))_{i,j \in \mathcal{T}}$ along the path of the Markov chain $X = (X_n)_{n \in \mathbb{N}_0}$ on $\mathcal{X}$. This may be seen as a discrete-time version of a Feynman-Kac formula and is completely standard. By $\mathbb{P}_x^{(X)}$ and $\mathbb{E}_x^{(X)}$ we denote probability and expectation with respect to the random walk when started at $x \in \mathcal{X}$.

Lemma 2.1. *For any $i, j \in \mathcal{T}$, $x, y \in \mathcal{X}$ and any time n,*

$$\mathrm{E}_{(i,x)}[\eta_n(j,y)] = \mathbb{E}_x^{(X)} \big[\big(M(X_0) \cdots M(X_{n-1}) \big)_{ij} \delta_y(X_n) \big]. \tag{2.1}$$

Proof. We proceed the proof by induction on n. For $n = 1$ it is easy to see that

$$\mathrm{E}_{(i,x)}[\eta_1(j,y)] = m_{ij}(x) P_{xy}, \tag{2.2}$$

from which immediately the assertion for $n = 1$ follows. Now assume that (2.1) holds for $n \in \mathbb{N}$. We have

$$\mathrm{E}_{(i,x)}[\eta_{n+1}(j,y)] = \sum_{k \in \mathcal{T}} \sum_{z \in \mathcal{X}} m_{kj}(z) P_{zy} \mathrm{E}_{(i,x)}[\eta_n(k,z)] \tag{2.3}$$

$$= \sum_{k \in \mathcal{T}} \sum_{z \in \mathcal{X}} m_{kj}(z) P_{zy} \mathbb{E}_x^{(X)} \big[\big(M(X_0) \cdots M(X_{n-1}) \big)_{ik} \delta_{X_n}(z)) \big] \tag{2.4}$$

$$= \mathbb{E}_x^{(X)} \big[\big(M(X_0) \cdots M(X_n) \big)_{ij} \delta_{X_{n+1}}(y)) \big] \tag{2.5}$$

where in the second equality we have used the induction step and in the third equality we have used the Markov property of X_n. Hence, by induction the proof of the lemma is finished. $\qquad \square$

Now, explicitly writing out the matrix product in (2.1) and summing over $j \in \mathcal{T}$, we get

$$u_n(i,x) = \sum_{k_1, \ldots, k_n \in \mathcal{T}} \mathbb{E}_x^{(X)} \Big[\prod_{l=1}^n m_{k_{l-1}k_l}(X_{l-1}) \Big], \qquad k_0 = i. \tag{2.6}$$

Now we absorb the transition probabilities given in (1.2) in the product and easily rewrite the right-hand side of (2.6) as the right-hand side of (1.3), which finishes the proof of Proposition 1.1.

2.2 Proof of Theorem 1.3.

We start from Proposition 1.1 and recall the normalized empirical pair measure of (T, X):

$$\nu_n = \frac{1}{n} \sum_{l=1}^{n} \delta_{((T_{l-1}, X_{l-1}), (T_l, X_l))},$$

which is a probability measure on $(\mathcal{T} \times \mathcal{X})^2$. Now fix $i \in \mathcal{T}$ and $x \in \mathcal{X}$ and rewrite the right-hand side of (1.3) in terms of ν_n:

$$u_n(i, x) = \mathbb{E}_{i,x}^{(T,X)} \Big[\Big(\prod_{k,j \in \mathcal{T}} \prod_{z \in \mathcal{X}} m_{kj}(z)^{n \sum_{y \in \mathcal{X}} \nu_n((k,z),(j,y))} \Big)$$
$$\Big(\prod_{k \in \mathcal{T}} \mathsf{deg}^+(k)^{n \sum_{j \in \mathcal{T}, z, y \in \mathcal{X} } \nu_n((k,z),(j,y))} \Big) \Big]. \tag{2.7}$$

Let us introduce the function $\mathcal{D}$ for $\nu \in \mathcal{M}_1((\mathcal{T} \times \mathcal{X})^2)$ by

$$\mathcal{D}(\nu) := \sum_{k \in \mathcal{T}} \overline{\nu}(k) \log(\mathsf{deg}^+(k)). \tag{2.8}$$

Now we take expectation with respect to the environment of the right-hand side of (2.7) and use that the $m_{kj}(z)$ with $z \in \mathcal{X}$ and $k, j \in \mathcal{T}$ are independent. Using the notation introduced in (1.11) and the logarithmic moment generating function $H_{kj}(t) = \log\langle m_{kj}(0)^t \rangle$ we get:

$$\langle u_n(i, x) \rangle = \Big\langle \mathbb{E}_{i,x}^{(T,X)} \Big[\exp \Big(\sum_{k,j,z} \big(\log m_{kj}(z) \big) n \overline{\nu}_n(k, j, z) + \sum_{k} \log(\mathsf{deg}^+(k)) n \overline{\nu}_n(k) \Big) \Big] \Big\rangle$$
$$= \mathbb{E}_{i,x}^{(T,X)} \Big[e^{n \mathcal{D}(\nu_n)} \prod_{k,j \in \mathcal{T}} e^{\sum_{z \in \mathcal{X}} H_{kj}(n \overline{\nu}_n(k, j, z))} \Big] \tag{2.9}$$
$$= \mathbb{E}_{i,x}^{(T,X)} \Big[e^{n \mathcal{D}(\nu_n)} \prod_{k,j \in \mathcal{T}} \exp \Big(n \sum_{z \in \mathcal{X}} \frac{H_{kj}(n \overline{\nu}_n(k, j, z)) - \overline{\nu}_n(k, j, z) H_{kj}(n)}{n} \Big)$$
$$\times \exp \Big(\sum_{k,j \in \mathcal{T}} \overline{\nu}_n(k, j) H_{kj}(n) \Big) \Big].$$

Recall Assumption (1.2) and note that H_{kj} satisfies the asymptotics $H_{kj}(n) = \rho_{kj} n \log(\rho_{kj} n) - \rho_{kj} n + o(n)$ as $n \to \infty$ [GM98]. Furthermore, we use the asymptotics in (1.8) for every $z \in \mathcal{X}$ to conclude that as $n \to \infty$

$$\langle u_n(i, x) \rangle = \mathbb{E}_{i,x}^{(T,X)} \Big[e^{n \mathcal{D}(\nu_n)} \exp \Big(n \sum_{k,j \in \mathcal{T}} \rho_{kj} \sum_{z \in \mathcal{X}} \overline{\nu}_n(k, j, z) \log \overline{\nu}_n(k, j, z)$$
$$+ \sum_{k,j \in \mathcal{T}} \overline{\nu}_n(k, j) \Big(\rho_{k,j} n \log(\rho_{kj} n) - \rho_{kj} n \Big) \Big) \Big] e^{o(n)} \tag{2.10}$$
$$= \mathbb{E}_{i,x}^{(T,X)} \Big[e^{n \mathcal{S}(\nu_n) + n \mathcal{D}(\nu_n)} \exp \Big(n \log \Big(\frac{n}{e} \Big) \langle \overline{\nu}_n, \rho \rangle \Big) \Big] e^{o(n)},$$

where we used the definition of the functional $\mathcal{S}$ in (1.12) and the notation from (1.9) for the last equality. Also note that in the above display we have $\langle \overline{\nu}, \rho \rangle = \sum_{i,j} \overline{\nu}(i, j) \rho_{ij}$.

Our main tool in the proof is that the pair empirical measure $(\nu_n)_n$ of the Markov chain $(T_n, X_n)_{n\in\mathbb{N}}$ satisfies a large deviation principle on the set of probability measures on $(\mathcal{T} \times \mathcal{X})^2$ with scale n and the good rate function I' given by

$$I'(\nu) = \sum_{(i,j)\in\mathcal{A}, x,y\in\mathcal{X}} \nu((i,x),(j,y)) \log \frac{\nu((i,x),(j,y))}{\overline{\nu}(i,x)P_{xy}/\deg^+(i)} = I(\nu) + \mathcal{D}(\nu) \qquad (2.11)$$

if ν satisfies the marginal condition (i.e., lies in $\mathcal{M}_1^{(s)}((\mathcal{T} \times \mathcal{X})^2)$), and $I'(\nu) = \infty$ otherwise.

Upper bound: Since we are on a finite space, it is clear that we can restrict ourselves to $\nu \in \overline{\mathcal{M}_1^{(s)}((\mathcal{T} \times \mathcal{X})^2)}$. We define

$$\Lambda_\delta(\rho) := \Big\{ \mu \in \mathcal{M}_1^{(s)}(\mathcal{T}^2) \colon \langle \mu, \rho \rangle \geq \lambda(\rho) - \delta \Big\}, \qquad \delta > 0. \qquad (2.12)$$

We split the expectation on the right-hand side of (2.10) into the contribution from the events $\{\overline{\nu}_n \in \Lambda_\delta(\rho)\}$ and its complement. For the first order term in the first part we use the upper bound $\lambda(\rho)$ and, in the second part the first order term is bounded above by $\mathrm{e}^{n\log(n/\mathrm{e})(\lambda(\rho)-\delta/2)}$. Hence, we obtain, for all sufficiently large n,

$$\langle u_n(i,x) \rangle \leq \mathrm{e}^{n\log(n/\mathrm{e})\lambda(\rho)} \mathbb{E}_{i,x}^{(T,X)} \Big[\mathrm{e}^{n\mathcal{S}(\nu_n)+n\mathcal{D}(\nu_n)} \mathbb{1}\big\{\overline{\nu}_n \in \Lambda_\delta(\rho)\big\} \Big] \mathrm{e}^{o(n)}$$
$$+ \mathrm{e}^{n\log(n/\mathrm{e})(\lambda(\rho)-\delta/2)}. \qquad (2.13)$$

Since $\Lambda_\delta(\rho)$ is closed and $\mathcal{S}$ and $\mathcal{D}$ are continuous and bounded, with the same reasoning as in the proof of the upper bound in Varadhan's Lemma (see e.g. pages 32-34 in [H00]) , we can conclude that as $n \to \infty$

$$\mathbb{E}_{i,x}^{(T,X)} \Big[\mathrm{e}^{n\mathcal{S}(\nu_n)+n\mathcal{D}(\nu_n)} \mathbb{1}\big\{\overline{\nu}_n \in \Lambda_\delta(\rho)\big\} \Big] \leq \mathrm{e}^{o(n)} \exp\Big(-n \inf_{\Lambda_\delta(\rho)} (I - \mathcal{S}) \Big). \qquad (2.14)$$

In the above we also used that $I' = I + D$. Since I and $\mathcal{S}$ are continuous and bounded functionals, $\Lambda_\delta(\rho)$ is compact for any $\delta \geq 0$ and $\Lambda_\delta(\rho) \downarrow \Lambda(\rho)$ as $\delta \downarrow \rho$, we can conclude that

$$\lim_{\delta\downarrow 0} \inf_{\Lambda_\delta(\rho)} (I - \mathcal{S}) \geq \inf_{\Lambda(\rho)} (I - \mathcal{S}). \qquad (2.15)$$

This finishes the proof of the upper bound in (1.14).

Lower Bound: Now we prove the lower bound in (1.14). For any $\nu \in \Lambda(\rho)$ let

$$B_n(\nu) = \{\nu' \in \mathcal{M}_1^{(s)}((\mathcal{T} \times \mathcal{X})^2) : d(\nu',\nu) < n^{-1/2}\} \qquad (2.16)$$

where $d(\cdot,\cdot)$ is the usual total variation distance on probability distributions. We can find a constant $c \in \mathbb{R}$ that depends only on ρ such that for any $\nu \in \Lambda(\rho)$, for any n and $\nu' \in B_n(\nu)$

$$\langle \nu', \rho \rangle \leq \lambda(\rho) + cn^{-1/2}. \qquad (2.17)$$

Then, for any $\nu \in \Lambda(\rho)$, from (2.10) we have the lower bound

$$\langle u_n(i,x) \rangle \geq \mathrm{e}^{o(n)} \exp\Big(n\log\Big(\frac{n}{\mathrm{e}}\Big)(\lambda(\rho) - cn^{-1/2})\Big)$$
$$\mathbb{E}_{i,x}^{(T,X)} \Big[\mathbb{1}\{\nu_n \in B_n(\nu)\} \mathrm{e}^{n\mathcal{S}(\nu_n)+n\mathcal{D}(\nu_n)} \Big]. \qquad (2.18)$$

Using the explicit combinatorics (see the proof of Theorem II.8 in [H00]) we have for any $\nu \in \mathcal{M}_1^{(\mathrm{s})}((\mathcal{T} \times \mathcal{X})^2)$

$$\liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}_{i,x}^{(T,X)}\big(\nu_n \in B_n(\nu)\big) \geq -I'(\nu). \tag{2.19}$$

Using the continuity of $\mathcal{S}$ and $\mathcal{D}$ we get for any $\nu \in \Lambda(\rho)$

$$\langle u_n(i,x) \rangle \geq \mathrm{e}^{o(n)} \exp\left(n \log\left(\frac{n}{\mathrm{e}}\right)\lambda(\rho)\right)\mathrm{e}^{n\mathcal{S}(\nu)+n\mathcal{D}(\nu)}\mathrm{e}^{-nI'(\nu)}. \tag{2.20}$$

Recalling that $I' = I + D$ and taking the supremum over $\Lambda(\rho)$ finishes the proof of the lower bound.

2.3 Proof of Lemma 1.4.

We first prove that for any $\gamma \in \Gamma$ the measure μ_γ is an extremal point of $\mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$. We need to show that if $\mu_\gamma = a\mu_1 + (1-a)\mu_2$ for some $a \in (0,1)$ and $\mu_1, \mu_2 \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$, then it must be the case that $\mu_1 = \mu_2$. Since $a \in (0,1)$, both supports of μ_1 and μ_2 are contained in the support of μ_γ. Without loss of generality, suppose that there exists an edge (i,j) such that $(i,j) \in \mathrm{supp}\,\mu_\gamma$ with $\mu_1(i,j) = 0$. Since γ is a simple cycle, $\mathrm{supp}\,(\mu_1) \subsetneq \mathrm{supp}\,\mu_\gamma$ and $\mu_1 \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$, we get that $\mu_1 \equiv 0$. Thus, the supports of μ_1 and μ_2 are both equal to the support of μ_γ, which is equal to the set of the edges of the cycle γ. The only measure in $\mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$ having this support is μ_γ. Hence, $\mu_\gamma = \mu_1 = \mu_2$.

Now we show that if $\mu \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$ is not of form μ_γ for some $\gamma \in \Gamma$, then it is not an extremal point of $\mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$. By the marginal property, $\mathrm{supp}\,(\mu)$ contains some simple cycle γ_1. On the other hand, $\mathrm{supp}\,(\mu)$ is not a simple cycle. As a result, $\mathrm{supp}\,(\mu)$ contains a second simple cycle $\gamma_2 \in \Gamma$ such that $\gamma_2 \neq \gamma_1$. Now we choose $\varepsilon > 0$ small enough so that $\mu(i,j) \geq 4\varepsilon$ for any $(i,j) \in \gamma_1 \cup \gamma_2$. Then we have $\mu - \varepsilon|\gamma_1|\mu_{\gamma_1} - \varepsilon|\gamma_2|\mu_{\gamma_2} \geq 2\varepsilon$ and $1 - \varepsilon|\gamma_1| - \varepsilon|\gamma_2| \in (1/2, 1)$. Then the measure

$$\mu' := \frac{1}{1 - \varepsilon|\gamma_1| - \varepsilon|\gamma_2|}\Big(\mu - \varepsilon|\gamma_1|\mu_{\gamma_1} - \varepsilon|\gamma_2|\mu_{\gamma_2}\Big) \tag{2.21}$$

belongs to $\mathcal{M}_1(\mathcal{T}^2)$. Also since $\mu, \mu_{\gamma_1}, \mu_{\gamma_2}$ satisfy marginal property, so does μ' and hence, $\mu' \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$. Finally, we can write μ as

$$\mu = \varepsilon|\gamma_1|\mu_{\gamma_1} + \varepsilon|\gamma_2|\mu_{\gamma_2} + (1 - \varepsilon|\gamma_1| - \varepsilon|\gamma_2|)\mu' \tag{2.22}$$

and, as noted above $\varepsilon|\gamma_1|, \varepsilon|\gamma_2|, 1 - \varepsilon|\gamma_1| - \varepsilon|\gamma_2| \in (0,1)$ and $\mu_{\gamma_1}, \mu_{\gamma_2}, \mu' \in \mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$. Therefore, μ is not an extremal point of $\mathcal{M}_1^{(\mathrm{s})}(\mathcal{T}^2)$.

2.4 Proof of Lemma 1.6.

We rewrite $\mathcal{I} - \mathcal{S} = \widetilde{G} + \widetilde{F}$, where

$$\begin{aligned}
\widetilde{G}(\nu) &:= \sum_{i,j \in \mathcal{T}} (1 - \rho_i)\nu(i,j) \log \frac{\nu(i,j)}{\overline{\nu}(i)}, \\
\widetilde{F}(\nu) &:= -\sum_{i \in \mathcal{T}} \rho_i \overline{\nu}(i) \log \overline{\nu}(i) - \sum_{i \in \mathcal{T}} \overline{\nu}(i)\rho_i \log \rho_i.
\end{aligned} \tag{2.23}$$

Using this and Lemmas 1.4 and 1.5, we get that $\Lambda(\rho)$ is the convex hull of $\{\mu_\gamma : \gamma \in \Gamma(\rho)\}$. This and the fact that $\widetilde{F}$ is strictly concave gives

$$\widetilde{F}(\nu) \geq \min \left\{ \widetilde{F}(\mu_\gamma) \colon \gamma \in \Gamma(\rho) \right\}. \tag{2.24}$$

We have $\widetilde{G}(\nu) \geq 0$ for any $\nu \in \mathcal{M}_1((\mathcal{T} \times \mathcal{X})^2)$, since $\rho_i \geq 1$ for all $i \in \mathcal{T}$. Moreover, for any cycle γ, $\widetilde{G}(\nu_\gamma) = 0$. Hence, by (2.24) we get that

$$\chi(\rho) = \min \left\{ \widetilde{F}(\mu_\gamma) \colon \gamma \in \Gamma(\rho) \right\}. \tag{2.25}$$

Finally, note that for any simple cycle $\gamma = (i_1, \ldots, i_{|\gamma|})$ with $\gamma \in \Gamma(\rho)$

$$\widetilde{F}(\mu_\gamma) = \lambda(\rho) \log |\gamma| - \sum_{m=1}^{\gamma} \frac{1}{|\gamma|} \rho_{i_m} \log \rho_{i_m}. \tag{2.26}$$

This finishes the first part of the proof. If $\rho_i = \rho$ for all i, then $\lambda(\rho) = \rho$, $\Gamma(\rho) = \Gamma$ and $\widetilde{F}(\mu_\gamma) = \rho \log |\gamma| - \rho \log \rho$. Finally, noting that $\widetilde{F}$ is strictly concave finishes the proof of the second part.

BIBLIOGRAPHY

[AN72] K.B. ATHREYA and P.E. NEY, *Branching Processes.* Springer-Verlag (1972).

[BGK09] C. BARTSCH, N. GANTERT and M. KOCHLER, Survival and growth of a branching random walk in random environment. *Markov Proc. Relat. Fields* **15:4**, 525–548 (2009).

[BCN99] J.D. BIGGINS, H. COHN and O. NERMAN, Multi-type branching in varying environment. *Stoch. Proc. Appl.* **83**, 357–400 (1999).

[BGK05] M. BIRKNER, J. GEIGER and G. KERSTING, Branching processes in random environment – a view on critical and subcritical cases. In: J.-D. Deuschel and A. Greven (Eds.), Interacting Stochastic Systems, pp. 269–291, Springer (2005).

[CMP98] F. COMETS, M.V. MENSHIKOV and S. POPOV, One-dimensional branching random walk in random environment: A classification. *Markov Proc. Relat. Fields* **4**, 465–477 (1998).

[CP07] F. COMETS and S. POPOV, Shape and local growth for multidimensional branching random walks in random environment. *ALEA Lat. Am. J. Probab. Math. Stat.* **3**, 273–299 (2007).

[CY11] F. COMETS and N. YOSHIDA, Branching random walks in space-time random environment: survival probability, global and local growth rates. *J. Theor. Probab.* **24**, 657–687 (2011).

[GMPV10] N. GANTERT, S. MÜLLER, S. POPOV and M. VACHKOVSKAIA, Survival of branching random walks in random environment. *J. Theoret. Probab.*, **4:23**, 1002–1014 (2010).

[H00] F. DEN HOLLANDER, Large deviations, Fields Institute Monographs, 14, American Mathematical Society, Providence, RI (2000).

[GM98] J. GÄRTNER and S. MOLCHANOV, Parabolic problems for the Anderson model II. Second-order asymptotics and structure of high peaks. *Probab. Theory Relat. Fields* **111**, 17–55 (1998).

[J75] P. JAGERS, *Branching Processes with Biological Applications.* Wiley, London (1975).

[KA02] M. KIMMEL and D.E. AXELROD, *Branching Processes in Biology.* Springer-Verlag (2012).

[MP00] F.P. MACHADO and S. POPOV, One-dimensional branching random walk in a Markovian random environment. *J. Appl. Probab.* **77**, 1156–1163 (2000).

[MP03]   F.P. MACHADO and S. POPOV, Branching random walk in random environment on trees. *Stoch. Proc. Appl.* **106**, 95–106 (2003).

[M71]   C.J. MODE, *Multitype Branching Processes*. American Elsevier, New York (1971).

[M08]   S. MÜLLER, Recurrence and transience for branching random walks in an i.i.d. random environment. *Markov Proc. Relat. Fields* **14:1**, 115–130 (2008).

[S06]   E. SENETA, *Non-negative Matrices and Markov Chains*, Springer Series in Statistics No. 21 (2006).

[T81]   D. TANNY, On multitype branching processes in random environment. *Adv. Appl. Probab.* **13**, 464–497 (1981).

[Y08]   N. YOSHIDA, Central limit theorem for branching random walk in random environment. *Ann. Appl. Probab.* **18**, 1619–1635 (2008).