MATHEMATICAL ANALYSIS OF MARKOV MODELS FOR SOCIAL PROCESSES

by

Joseph M. Whitmeyer

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2010

Approved by:

____

Dr. Stanislav A. Molchanov

____

Dr. Joseph Quinn

____

Dr. Volker Wihstutz

____

Dr. William P. Brandon

ABSTRACT

JOSEPH M. WHITMEYER. [Mathematical analysis of Markov models for social processes]. (Under the direction of Dr. STANISLAV MOLCHANOV.)

We present Markov models for two social processes: the spread of rumors and the change in the spatial distribution of a population over time. For the spread of rumors, we present two models. The first is for the situation in which all particles are identical but one initially knows the rumor. The second is for a situation in which there are two kinds of particles: spreaders, who can spread the rumor, and ordinary particles, who only can learn the rumor. We find that the limiting distribution for the first model is the convolution of two double exponential distributions and for the second model is a double exponential distribution.

The stochastic dynamics for our model of the change in the spatial distribution of a population over time include the four basic demographic processes: birth, death, migration, and immigration. We allow interaction between particles only inasmuch as the immigration rate can depend on the existing configuration of particles. We focus on the critical case of constant mean density, under the conditions of long jumps migration, immigration in which distant particles have a positive effect, or both. We prove, under these conditions, the existence of ergodic limiting behavior: the point process is stationary in space and time. Without the strong mixing due to these conditions, the population vanishes due to infinite clusterization.

ACKNOWLEDGMENTS

I am indebted to Dr. Molchanov for his direction and guidance. Without it this would scarcely have been possible. I am grateful also to Drs. Wihstutz, Quinn, and Brandon for their help and support. I thank Yaqin Feng for her close reading of the material in the second chapter. Finally, I am grateful to my wife, Rosemary Hopcroft, and my children, Mark and Sophie, who have encouraged this enterprise from the start and have been tolerant of its demands.

INTRODUCTION

We present here Markov models of two social processes: two simple models of the spread of rumors and a more complex model of the spatial distribution of a population determined by the basic demographic processes of birth, death, migration, and immigration. We focus on situations and parameter settings in which the outcomes are unclear in advance and the analysis more revealing. Our analysis of the rumors model, therefore, focuses on the outcome as the population becomes large. Concerning the demographic model, for some configurations of the birth, death, and immigration rates, a population will degenerate, i.e., disappear, and for others a population will explode. Our analysis concentrates accordingly on the model with critical settings of the parameters that produce a situation in between those extremes. In both models we are interested especially in asymptotic results: the outcomes that are approached as the connected population grows large and as time increases, respectively.

We may note that the situations on which we focus are precisely those in which mathematical models will be most useful and necessary. Policy makers, for example, are likely to be interested in demographic processes associated with a stable population, certainly more than in processes that lead to degeneration or explosion of the population! In addition, there are some arguments and evidence that there may even be a tendency for some social processes to evolve to a critical state (e.g., [Jensen (1998)], [Whitmeyer and Yeingst (2006)]), although this need not be true for the models still to be informative and useful.

The use of mathematical models in the social sciences goes back at least to the nineteenth century, and we touch on some of this history in Chapter 2. In recent years, however, the overwhelming emphasis in the social sciences (excluding economics) has been on the use of statistical models, commonly some variant of linear regression models. While these can be useful for some situations, there is reason to believe that they are inadequate and inappropriate for an extensive variety of social processes

[Whitmeyer (2009)]. A better alternative, often, is to create mathematical models that embody the key elements of the processes and, although simplifications of empirical processes and, correspondingly, abstract, illuminate the dynamics, outcomes, and critical aspects of the processes. This is the approach that motivates our work here.

In Chapter 1, we present two Markov models of the spreading of rumors. Specifically, we determine the limiting distribution as the population becomes large for the time to spreading of the rumor to the full population. For the first model, all particles are identical but one initially knows the rumor. The limiting distribution is the convolution of two double exponential distributions. For the second model, there are two kinds of individuals: spreaders, who can spread the rumor, and ordinary individuals, who can only learn the rumor. Here, the limiting distribution is simply a double exponential distribution.

In Chapter 2, we present a Markov model of a population of particles for which the stochastic dynamics include the basic demographic processes: birth, death, migration, and immigration. Some interaction between particles is allowed: the immigration process at a given location depends on the spatial configuration of existing particles. We focus on the critical case of constant mean density and prove, under appropriate conditions, the existence of ergodic limiting behavior: the point process is stationary in space and time. These conditions are those that promote strong mixing: immigration that is promoted by particles even a long distance from the immigration site, long jump migration, or both. Without such strong mixing the population vanishes due to infinite clusterization.

We begin with the background to the analysis. We note the Galton-Watson process as the historical origin of the study of branching processes and motivate the development of our model by noting that this model produces clusterization in stable populations, which we wish to avoid. We describe the four demographic processes mathematically and the summary process in infinitesimal time. For the two heavy tail processes, migration and immigration, we present evidence from empirical studies of these processes, especially for humans, supporting the plausibility of our assumptions.

We continue with the analysis of the moments of the demographic model. We use the forward equations to derive differential equations for the first three moments, and present a recurrence equation from which the differential equations for all remaining moments may be calculated. From the equations for the first two moments, we are able to show the stationarity of the model under the critical setting of the rate parameters. Finally, we derive the aysmptotic of the variance of the number of particles in a region as the size of the region increases.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: TWO MARKOV MODELS OF THE SPREAD OF RUMORS

1.1 Introduction

The spread of rumors from person to person over a large group of people is a socially important and mathematically interesting phenomenon. In times and places where mass communication does not exist or is not available, rumors can travel quickly with far-reaching consequences. Examples are the Great Fear that possessed much of the French countryside in 1789, during the French Revolution [Goodwin (1966)], and, more recently, the quick surge of public opposition to Communist governments in Central Europe in 1989 [Ash (1993)]. The spread of rumors is the spread of one kind of information or, even more generally, individual attribute. Thus, phenomena such as the spread of information in general, innovations, or fashion may be similar in their progression.

In mathematics, a sizable literature exists on the spread of rumors. Deterministic models exist from the 1950s (e.g., [Rapoport and Rebhun (1952.)]). It is recognized, however, that stochastic models are preferable [Pearce (2000)], especially as they are more accurate near the absorption state, and the focus has been on stochastic models since the work of Daley and Kendall [Daley and Kendall (1965)]. The Daley-Kendall model has spawned many refinements and variations (e.g., [Watson (1988)], [Pittel (1990)], [Maki and Thompson (1973.)], [Sudbury (1985)], [Lefèvre and Picard (1994)] and [Pearce (2000)]). We should note that there are similarities between the spread of rumors and the even more extensively studied phenomenon of epidemics but, as Pearce [Pearce (2000)] points out, the processes are sufficiently different that the models differ as well.

The Daley-Kendall model involves three kinds of individuals: susceptibles, who do not yet know the rumor; spreaders, who can pass the rumor on to others; and immunes, who are former spreaders who can no longer pass along the rumor. The processes involving immunes, who are produced by encounters between two spreaders or a spreader and an immune, mean that the models do not have an exact solution. Here we present a pair of models that do not allow for immunes and, thus, are simpler than the Daley-Kendall model and its successors. Our models have the virtues of being completely solvable and generating previously unobtained results.

1.2 The Models

Following Daley and Kendall, we call an individual that does not yet know the rumor a susceptible and an individual that can pass the rumor on to others a spreader. In our models, every susceptible is capable of learning the rumor. Our first model, in fact, has only one kind of individual. We begin with one individual knowing the rumor. Every other individual begins as a susceptible, and once an individual hears the rumor it becomes a spreader. Our second model has two kinds of individuals, spreaders and ordinary individuals. Only the spreaders can spread the rumor. Ordinary individuals begin as susceptibles and can learn the rumor but cannot spread it. We call an ordinary individual that has learned the rumor informed. To keep the second model simple, we assume a fixed number m of spreaders. This might correspond to a situation in which professional agents pass information to their clients, who are not motivated or expert enough to pass on the information themselves. Examples are physicians telling their patients about new medical results or treatments, sales agents informing potential customers about a product, or political agents talking to people they meet about some candidate or policy.

Our main interest is the full spreading time, the time until the rumor has reached all susceptible individuals. In Model 1, we have $N + 1$ individuals and our initial situation is that one individual knows the rumor and N are susceptibles. Again, in this model when a susceptible learns the rumor it becomes a spreader. Model 2 has initial populations of m spreaders and N susceptibles. We denote the full spreading time τ_N for Model 1 and T_N for Model 2. We want to study the asymptotic behavior of τ_N and T_N.

We take the spread of rumors to be a Markov process. Using continuous time, if there are k informed individuals the time to informing one additional individual (denoted $\tau_{k,k+1}$ and $T_{k,k+1}$) has an exponential distribution with parameter λ_k. That is, in Model 1 the $\tau_{k,k+1}$ and in Model 2 the $T_{k,k+1}$ are independent random variables with distribution $\exp(\lambda_k)$. The generating matrix for this process is, therefore, $[a_{jk}]$ where $a_{jk} = -\lambda_j$ for $j = k$, $a_{jk} = \lambda_j$ for $k = j + 1$, and $a_{jk} = 0$ otherwise.

We derive the parameters λ_k as follows. Following Daley and Kendall [Daley and Kendall (1965)] and Pierce [Pearce (2000)], we assume homogeneous mixing of the population and random encounters between individuals. Let λ denote the basic rate of rumor spreading, which may be thought of as the rate of spreading when there is one spreader and any individual the spreader encounters is susceptible. Then the basic rate is multiplied by the number of spreaders as well as by the probability that an individual encountered by a spreader is susceptible. In our models this probability is taken to be simply the proportion of susceptibles that are not already infected. The number of spreaders in the first model is k, the number of informed individuals. In the second model, the number of spreaders is fixed at m. This yields:

Model 1:

$$\lambda_k = \frac{N+1-k}{N} k\lambda$$

.

Model 2:

$$\lambda_k = \frac{N-k}{N} m\lambda$$

.

Because the $\tau_{k,k+1}$ and the $T_{k,k+1}$ are independent, we can easily calculate the mean and variance of the full spreading time. Using the fact that the sum of the harmonic series through the Nth term for large N equals $\ln N + \gamma + o(1)$ (Euler's $\gamma \approx .5771$) we obtain for the means:

Model 1:

$$E\tau_N = \sum_{k=1}^{N} \tau_{k,k+1} = \sum_{k=1}^{N} \lambda_k^{-1} = \frac{N}{\lambda(N+1)} \left(\sum_{k=1}^{N} \frac{1}{k} + \sum_{k=1}^{N} \frac{1}{N+1-k} \right) = \frac{2}{\lambda}(\ln N+\gamma)+o(1)$$

Model 2:

$$ET_N = \sum_{k=0}^{N-1} T_{k,k+1} = \sum_{k=0}^{N-1} \lambda_k^{-1} = \frac{N}{m\lambda} \sum_{k=1}^{N} k^{-1} = \frac{1}{m\lambda} N(\ln N + \gamma) + o(N)$$

.

The variances are more interesting:

Model 1:

$$\begin{aligned} \mathrm{Var}(\tau_N) &= \sum_{k=1}^{N} \mathrm{Var}(\tau_{k,k+1}) = \sum_{k=1}^{N} \lambda_k^{-2} = \sum_{k=1}^{N} \frac{N^2}{(N+1-k)^2 k^2 \lambda^2} \\ &= \frac{2N^2}{\lambda^2 (N+1)^2} \left(\sum_{k=1}^{N} \frac{1}{k^2} + \frac{2}{\lambda N + 1} \sum_{k=1}^{N} \frac{1}{k} \right) = \frac{2}{\lambda^2} \left(\frac{\pi^2}{6} + o(1) \right) = \frac{\pi^2}{3\lambda^2} + o(1). \end{aligned}$$

Model 2:

$$\mathrm{Var}(T_N) = \sum_{k=0}^{N-1} \mathrm{Var}(T_{k,k+1}) = \sum_{k=0}^{N-1} \lambda_k^{-2} = \frac{N^2}{m^2\lambda^2} \sum_{k=0}^{N-1} \frac{1}{(N-k)^2} = \frac{N^2}{m^2\lambda^2} \left(\frac{\pi^2}{6} + o(1) \right).$$

We used, here, the well-known fact that $\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$, i.e., $\sum_{k=1}^{N} \frac{1}{k^2} = \frac{\pi^2}{6} + o\left(\frac{1}{N}\right)$.

The central limit theorem applies to neither model. This is indicated in the first model by the finite variance. In the second model, the third moment (not shown) increases too quickly relative to the second moment, and thus the central limit theorem fails to apply here as well.

We can expect, however, that for each model the full spreading time approaches a limiting distribution. That is, we seek the asymptotic distributions of random variables $\zeta^{(1)}$ and $\zeta^{(2)}$, where:

$$\tau_N - A_N \xrightarrow{\text{law}} \zeta^{(1)}, \ \frac{T_N - \tilde{A}_N}{B_N} \xrightarrow{\text{law}} \zeta^{(2)},$$

where $A_N \to \infty$, $\tilde{A}_N \to \infty$, and $B_N \to \infty$ are appropriate normalization factors.

Theorem 1.1. *Model 1. For $N \to \infty$, $\lambda\tau_N - 2\ln N \overset{\text{law}}{=} \zeta_1 + \zeta_2$, where ζ_1 and ζ_2 are i.i.d. random variables with a double exponential distribution.*

Model 2. For $N \to \infty$, $\frac{m\lambda}{N}T_N - \ln N \overset{\text{law}}{=} \zeta$, where ζ is a random variable with a double exponential distribution.

To prove this, we use the following two lemmas.

Lemma 1.2. *(From Feller ([Feller (1971)], Ch. 1). Let $X_1, X_2, ..., X_n$ be i.i.d. random variables, $X_i \sim \text{Exp}(1)$. Order them: $X_{(1)}, X_{(2)}, ..., X_{(n)}$. Then, $X_{(1)}, X_{(2)} - X_{(1)}, ..., X_{(n)} - X_{(n-1)}$ are independent, exponentially distributed random variables, with $X_{(1)} \sim \text{Exp}(n)$ and $X_{(i+1)} - X_{(i)} \sim \text{Exp}(n-i)$.*

We do not give the proof here but offer the following rationale. The underlying explanation is the Markov property that the process is memoryless. We have that $X_{(1)} = \min(X_1, X_2, ..., X_n)$, thus, $P(X_{(1)} > a) = P(X_1 > a, X_2 > a, ..., X_n > a) = (\exp(-a))^n = e^{-na}$. It follows that $X_{(1)} \sim \text{Exp}(n) = \eta_1/n$, where $\eta_1 \sim \text{Exp}(1)$.

We have, next, that $X_{(2)} - X_{(1)} = \min(X_1 - X_{(1)}, X_2 - X_{(1)}, ..., X_n - X_{(1)})$ (with $X_{(1)} - X_{(1)}$ removed from this list). Because the process is memoryless, $P(X_{(2)} - X_{(1)} > a) = P(X_1 - X_{(1)} > a, X_2 - X_{(1)} > a, ..., X_n - X_{(1)} > a) = (\exp(-a))^{n-1} = e^{-(n-1)a}$. Thus, $X_{(2)} - X_{(1)} \sim \text{Exp}(n-1) = \eta_2/(n-1)$, where $\eta_2 \sim \text{Exp}(1)$. The distributions for the remaining random variables follow in the same fashion.

Lemma 1.3. *Let η_1, η_2, η_n be i.i.d. random variables, $\eta_i \sim \text{Exp}(1)$. Then, $\zeta :=$*

$\sum_{i=1}^{\infty} \frac{\eta_i - 1}{i} + \gamma$ *has a double exponential distribution, characterized by the distribution function* $F(x) = e^{-e^{-x}}$.

Proof. Let $X_1, X_2, , X_n$ be i.i.d. random variables, $X_i \sim \text{Exp}(1)$. Then, by Lemma 1.2, for i.i.d. random variables η_i with $\eta_i \sim \text{Exp}(1)$:

$$X_{(n)} = \max_{i \leq n}(X_i) \stackrel{\text{law}}{=} \sum_{i=1}^{n} \frac{1}{i} + \sum_{i=1}^{n} \frac{\eta_i - 1}{i}.$$

Let $X_1, X_2, , X_n$ be i.i.d. random variables, $X_i \sim \text{Exp}(1)$. Then,

$$X_{(n)} - \ln n \xrightarrow{\text{law}} \zeta = \sum_{i=1}^{n} \frac{\eta_i - 1}{i} + \gamma < \infty(\text{p} - \text{a.s.}). \tag{1.1}$$

We also have, however, that:

$$\begin{aligned} P(X_{(n)} - \ln n < x) &= P(X_{(n)} < \ln n + x) = (P(X_1 < \ln n + x))^n \\ &= \left(1 - \frac{e^{-x}}{n}\right)^n \xrightarrow[n \to \infty]{} e^{-e^{-x}}. \end{aligned} \tag{1.2}$$

Together, (1.1) and (1.2) imply that $\zeta = \sum_{i=1}^{\infty} \frac{\eta_i - 1}{i} + \gamma$ has what we will call the canonical double exponential distribution (with $E(\zeta) = \gamma$). We will use this fact in the form: $\sum_{i=1}^{N} \frac{\eta_i}{i} = \ln N + \zeta_N$, $\zeta_N \xrightarrow{\text{law}} \zeta_\infty$ with the canonical double exponential distribution. □

We now give the proof of Theorem 1.1.

Proof. Model 1:

For i.i.d. random variables η_k with $\eta_k \sim \text{Exp}(1)$ we can write:

$$\tau_N = \sum_{k=1}^{N} \tau_{k,k+1} = \sum_{k=1}^{N} \frac{N}{(N+1-k)k\lambda} \eta_k.$$

Then,

$$\begin{aligned} \tau_N &= \frac{N}{\lambda(N+1)} \left(\sum_{k=1}^{N} \frac{1}{k} \eta_k + \sum_{k=1}^{N} \frac{1}{(N+1-k)} \eta_k \right) \\ &= \frac{N}{\lambda(N+1)} \left(2 \sum_{k=1}^{N} \frac{1}{k} + \sum_{k=1}^{N} \frac{\eta_k - 1}{k} + \sum_{k=1}^{N} \frac{\eta_k - 1}{(N+1-k)} \right). \end{aligned}$$

Because $\sum_{k=1}^{\infty} \frac{\eta_k - 1}{k} < \infty(\mathrm{p-a.s.})$, we have, as $N \to \infty$:

$$\sum_{k=1}^{N} \frac{\eta_k - 1}{N+1-k} = \sum_{k=\frac{N}{2}+1}^{N} \frac{\eta_k - 1}{N+1-k} + o(1) = \sum_{k=1}^{\frac{N}{2}} \frac{\tilde{\eta}_k - 1}{k} + o(1) = \sum_{k=1}^{N} \frac{\tilde{\eta}_k - 1}{k} + o(1),$$

where the $\tilde{\eta}_k$ are i.i.d. random variables with $\tilde{\eta}_k \sim \mathrm{Exp}(1)$.

Thus, as $N \to \infty$:

$$\begin{aligned} \lambda\tau_N - 2\ln N &= \frac{N}{N+1}\left(2\sum_{k=1}^{N}\frac{1}{k} + \sum_{k=1}^{N}\frac{\eta_k - 1}{k} + \sum_{k=1}^{N}\frac{\tilde{\eta}_k - 1}{k} + o(1)\right) - 2\ln N \\ &= \frac{N}{N+1}\left(\sum_{k=1}^{N}\frac{\eta_k - 1}{k} + \sum_{k=1}^{N}\frac{\tilde{\eta}_k - 1}{k} + 2\gamma + o(1)\right) \overset{\text{law}}{=} \zeta_1 + \zeta_2, \end{aligned}$$

where ζ_1 and ζ_2 have canonical double exponential distributions.

Model 2:

For i.i.d. random variables η_k with $\eta_k \sim \mathrm{Exp}(1)$ we can write:

$$T_N = \sum_{k=0}^{N-1} T_{k,k+1} = \sum_{k=0}^{N-1} \frac{N}{(N-k)m\lambda}\eta_k = \frac{N}{m\lambda}\sum_{k=0}^{N-1}\frac{\eta_k}{N-k} = \frac{N}{m\lambda}\sum_{k=1}^{N}\frac{\tilde{\eta}_k}{k},$$

where the $\tilde{\eta}_k$ are i.i.d. random variables with $\tilde{\eta}_k \sim \mathrm{Exp}(1)$. Then, as $N \to \infty$:

$$\frac{m\lambda}{N}T_N = \sum_{k=1}^{N}\frac{\tilde{\eta}_k - 1}{k} + \sum_{k=1}^{N}\frac{1}{k} \overset{\text{law}}{=} \zeta + \ln N,$$

where ζ has the canonical double exponential distribution.

$\square$

1.3 Prediction Intervals

The results can be used to generate prediction intervals for the full spreading time given the population of susceptibles. By prediction interval we mean the shortest interval of time such that the probability that it contains the full spreading time is a given percent. We illustrate the generation of prediction intervals here.

Finding prediction intervals is complicated by the fact, shown in Figure 1, that the double exponential distribution (labeled "One") and the convolution of two double exponential distributions (labeled "Two") are asymmetric. This means that for a given error, the bounds for the narrowest prediction interval may not be equidistant from the mean.

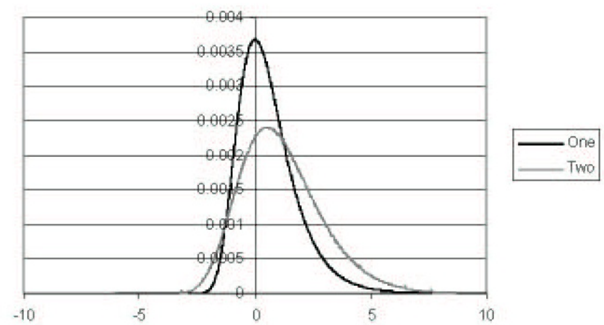

Figure 1.1: Density Functions for Double Exponential Distributions.

It is easy to show, e.g., using the method of Lagrange multipliers, that the prediction interval of narrowest width will occur when the heights of the density function at the two bounds are equal. I.e., letting x_1 and x_2 denote the lower and upper bounds of a prediction interval with a given error and letting $f(x)$ denote the density function, the width of the prediction interval $x_2 - x_1$ will be minimal when $f(x_1) = f(x_2)$.

This can be used to determine the prediction interval for Model 2, in which the normalized distribution of the full spreading time approaches simply a double exponential distribution. There is no simple equation for the two bounds of a prediction interval when the distribution is double exponential, but the constraint of the error term plus the fact that $f(x_1) = f(x_2)$ can be combined to find these points numerically. For example, for a 95 percent prediction interval for the canonical double exponential distribution, $x_1 = -1.56$ and $x_2 = 3.16$. The partition of the error is strongly asymmetrical: .0085 on the left and .0415 on the right.

For the convolution of two double exponential distribution, the limiting distribution in Model 1, the prediction interval is not as easy to determine. The convolution of two double exponential distributions must be calculated numerically, and a search procedure, guided by the stipulation that $f(x_1) = f(x_2)$, can find the prediction interval of minimum width. For a 95 percent prediction interval for the convolution of two canonical double exponential distributions, $x_1 = -2.06$ and $x_2 = 4.81$. The partition of the error is asymmetrical here as well: .012 on the left and .038 on the right.

Consider a numerical example. Suppose the population of susceptibles $N = 10^6$ and suppose the expected number of contacts per person per unit of time is 1, i.e., $\frac{1}{\lambda} = 1$. For model 1, then, the 95 percent prediction interval for the full spreading time is between 25.57 and 32.44 time units. For model 2, with m the number of spreaders, the 95 percent prediction interval for the full spreading time is between $12.25 \cdot 10^6/m$ and $16.98 \cdot 10^6/m$ time units.

1.4 The Future: Model 3

Our next step, which we do not carry out here, will be to develop a third model by removing the assumption in the second model that the number of spreaders is fixed. The third model again has two kinds of individuals, spreaders and ordinary individuals. In Model 3, however, spreaders, like ordinary individuals, can be either susceptible or informed. The process begins with a small number of informed spreaders who then spread the rumor to susceptible spreaders and ordinary individuals. Once a susceptible spreader hears the rumor, it becomes informed and can spread the rumor itself. Only an informed spreader can spread the rumor.

Mathematically, this means the third model concerns a two-dimensional Markov process. There are two basic rates of rumor spreading, one (λ) for ordinary individuals and one (μ) for spreaders. For ordinary individuals, therefore, the $\tau_{k,k+1}$, the times to inform one more person, are independent random variables with distribution $\text{Exp}(\lambda_{k,m})$ and for the spreaders the times to inform one more person, say, the $\nu_{m,m+1}$, are independent random variables with distribution $\text{Exp}(\mu_m)$. Letting N denote the total number of ordinary individuals and M the total number of spreaders and letting

k denote the number of informed ordinary individuals and m the number of informed spreaders, these parameters are given by:

Ordinary individuals:

$$\lambda_{k,m} = \left(\frac{N-k}{N+M}\right) m\lambda$$

Spreaders:

$$\mu_m = \left(\frac{M-m}{N+M}\right) m\lambda.$$

As can be seen from the formulas, the process for ordinary individuals does not affect the process for the spreaders. The converse is not true, as the spread of rumors among spreaders affects the spread of rumors among ordinary individuals, indicated by the m in the formula for $\lambda_{k,m}$. The third model, therefore, is a step up in complexity from the first two models.

1.5 Conclusion

In this study we used simple models of the spread of rumors to find the limiting distributions for the full spreading time, the time for the rumors to spread to the entire population. We treated two scenarios: when everyone in the population spreads the rumor and when only a fixed set of individuals spread the rumor. In the first scenario the full spreading time is $2\ln N + \zeta_N$ and the limiting distribution for ζ_N is a convolution of two double exponential distributions. In the second scenario the full spreading time is much greater, $\frac{n}{m}(\ln N + \tilde{\zeta}_N)$, and the limiting distribution for $\tilde{\zeta}_N)$ is simply a double exponential distribution.

To put these results more fully in context: if we have a classical branching process with no deaths with continuous time, then the expected time until N particles exist is $\frac{\ln N}{\lambda}$. In model 1, the full spreading time is twice as slow, an effect due mainly to the extremes of the process. That is, the parameter, $\lambda_k = \left(\frac{N+1-k}{n}\right) k\lambda$, becomes small when the number of spreaders, k, is small or when it is close to N. The slowing down effect is much stronger in model 2. In this scenario, as can be seen from the parameter $\lambda_k = \left(\frac{N-k}{n}\right) m\lambda$, the spreading process slows down at the extreme when there are few susceptibles, i.e., k is close to N. These contrasts are illuminated by the numerical example given above for a population of $N = 10^6$ ($\ln N \approx 13.8$).

CHAPTER 2: CRITICAL BRANCHING PROCESS ON $\mathbb{Z}^2$

2.1 Introduction

In 1873, Francis Galton posed a problem in the *Educational Times* [Galton (1 April 1873)] concerning the extinction of surnames, i.e., the extinction of male lines of descendants. He wanted to know, given the probability of a given number of male offspring per male, what proportion of surnames would disappear and how many people would hold a surname that survived. When he received no satisfactory solution to his problem, he persuaded the Reverend Henry William Watson, mathematician, clergyman, and alpinist, to take it up, and in 1874, they published the first mathematical treatment of what has become known as the Galton-Watson process [Galton and Watson (1874)]. After that point, progress was made on the problem most notably by J. B. S. Haldane [Haldane (1927)] and J. F. Steffenson [Steffenson (1930)], and the final solution was determined by 1950 with contributions by D. Hawkins and S. Ulam [Hawkins and Ulam (1944)], T. E. Harris [Harris (1963)], and A. M. Yaglom [Yaglom (1947)]. A. J. Lotka [Lotka (1931)] applied this model to the U.S. population, using data from the 1920 Census.

The Galton-Watson process is a simple example of a branching process [Kolmogorov and Dmitriev (1947)], a term for stochastic processes arising from incorporating probability theory into population processes [Kendall (1966.)]. In the continuous time version of the Galton-Watson process, an individual or *particle*, the term we will use, in an infinitesimal period of time dt produces one offspring with probability βdt and disappears (dies) with probability μdt. If it produced an offspring, then, there are two particles, each of which can produce an offspring or die, and the process continues in the same fashion. It is well known that the entire population, encompassing all lines, becomes extinct with probability 1 for $\mu \geq \beta$. Only when $\beta > \mu$ (the supercritical case) is there a positive probability that extinction does not occur. In fact, in this case the population follows the predictions of the Reverend Malthus [Malthus (1826.)] and grows exponentially: $En(t) = N_0 e^{(\beta-\mu)t}$, where $n(t)$ denotes the population at time t and N_0 is the initial population [Harris (1963)].

Current questions in what we might call mathematical ecology, of course, have moved on from Galton's original interest. In particular, we are seeking models that can generate and thereby provide possible explanations for two phenomena that have been observed empirically. First, biopopulations, including human populations, may exhibit stationarity in space and time. Roughly, this means that the stochastic process in question depends neither on the time we begin observing it nor on the place where we observe it. Mathematically, we will take this to mean that the mean and the variance of the number of particles at a given location do not depend on either the location or the time.

Although stationarity in space and time is seen among some populations of organisms, clearly it has not been a feature of human populations for most of recorded history. It may, however, hold for some human populations before the invention of agriculture in the Neolithic period and some contemporary developed countries. More importantly, some modern societies may have stationarity, at least in time, as a goal. They may seek to maintain current population levels without any more population increase. One question that our model addresses is what the ensuing spatial distribution of such a population will be.

This leads to the second empirically observed phenomenon. The spatial distributions of many species deviate strongly from a Poissonian point field, or more generally, patches, meaning a pattern of mostly empty space with sporadic, isolated concentrations of population. A patches pattern is what would result either from the random assignment of particles independently at each point in space or from the simplest demographic processes, as we describe below. The absence of this pattern in many populations, including some human populations, is another condition on the model we seek.

A model of population processes in space may be obtained by extending the Galton-Watson process by considering independent GW processes occurring in space. Specifically, we can consider a random point field $n(t,x)$ in the lattice $\mathbb{Z}^d$ (or in the space $\mathbb{R}^d$), with a critical GW process at each occupied point and no interaction or spatial dynamics. Assume, therefore, that $n(0,x)$ is the initial point field on $\mathbb{Z}^d$, given by the Bernoulli law: for any independent $x \in \mathbb{Z}^d$, $P\{n(0,x) = 1\} = \rho_0$, $P\{n(0,x) = 0\} = 1 - \rho_0$, where ρ_0 is the initial density of the particles. Assume now that each initial particle (located at x for $n(0,x) = 1$) generates its own family, concentrated at the same location $x \in \mathbb{Z}^d$. The result is a field $n(t,x)$ with independent values and constant density: $En(t,x) \equiv \rho_0$.

For large t, in this model, the majority of the cells $x \in \mathbb{Z}^d$ will be empty because $P\{n(t,x) = 0\} = \frac{\beta t}{1+\beta t} = 1 - \frac{1}{\beta t} + O(\frac{1}{t^2})$ (which gives the formula $P\{n(t,x) = 0 \mid n(0,x) = N_0\} \sim e^{-N_0/\beta t}$) [Gikhman and Skorokhod (1974)]. The populated points, moreover, are increasingly sparse (of order $\frac{1}{\beta t}$) and contain increasingly large families (of order βt). This is the phenomenon of clusterization: the population consists of large dense groups of particles separated by large distances (the distances must be of order $t^{1/\alpha}$!). As $t \to \infty$ the clusterization becomes stronger and stronger.

In order to avoid a patches pattern, as desired, the process must fill out empty space to compensate for the degenerating families. The simplest way to accomplish this is to add a simple random walk to nearest neighbors to the branching process. In other words, this model includes diffusion with generator $\kappa\Delta$, where κ is the rate of diffusion and Δ is the discrete or lattice Laplacian:

$$\Delta f(x) = \sum_{x':|x'-x|=1} (f(x') - f(x))$$

In high dimensions ($d \geq 3$) this simple random walk (diffusion) with generator $\kappa\Delta$ is sufficient to eliminate clusterization. It is a remarkable fact, however, that for $d =$

2, the most appropriate condition for demographic or most ecological applications, such local diffusion is not sufficient and the clusterization will increase infinitely. If, however, we modify the simple random walk to allow for long jumps with certain conditions then we can eliminate clusterization even in two dimensions. We call this modified random walk "migration," as "diffusion" is no longer appropriate. Such migration, with the addition of a similar immigration process, gives us the model we analyze here.

2.2 Description of the model

Our central goal is to introduce a discrete mathematical model describing two well-known empirical facts from ecology: the stationarity of particle fields in space and time and strong deviations from the classical Poissonian picture, i.e., spatial intermittency in the distribution of species (clusterization or "patches"). We are talking about an isolated population that is not involved in complex multispecies interaction (such as a predator-prey scheme). We exclude direct interaction between particles (the typical assumption in the theory of branching processes), but the birth-death mechanism will create a kind of mean field attractive potential.

Notation. Let $n(t, x)$ be the number of particles at the site $x \in \mathbb{Z}^d$ and at the moment $t > 0$ (time is continuous). We call $n(t, \cdot) : \mathbb{Z}^d \to \mathbb{Z}^d_+$ the configuration of the system at the moment t. The configurations will be locally finite in the following strong sense: $En^k(t, x) \leq c_0^k k!$ for all $x \in \mathbb{Z}^d$, all $k \geq 1$, and appropriate time independent constant c_0. The last (Carleman's) estimation will give us the possibility of constructing the field $n(t, \cdot)$ and studying its limiting behaviors $t \to \infty$ using the moments (correlation functions): $k_t(x_1, ..., x_l) = En(t, x_1)...n(t, x_l), l \geq 0, t > 0, x_1, ..., x_l \in \mathbb{Z}^{dl}$.

Assume that the initial configuration has a Poissonian structure, i.e., $n(t, x), x \in \mathbb{Z}^d$ are i.i.d. r.v.s with the Poissonian law and the parameter $\rho_0 = En(0, \cdot) > 0$ (the initial density of the population). Obviously, Carleman estimation is true for $n(0, x)$. The random dynamics of the point field includes four components:

a) The *death* of the particles has *rate* $\mu > 0$. That is, a particle, independently of others, dies during the time interval $(t, t + dt)$ with probability μdt.

b) The *birth* of the particles has *rate* $\beta > 0$. In this study we do not consider parameters of particles such as mass, size, etc., and changes in them in the process of the birth of a new particle or splitting of a particle, although such changes pose interesting problems. We also consider only binary splitting, i.e., the reaction P $\to$ P + P.

c) *Migration* of the particles. This process depends on the probability kernel $a(z), z \in \mathbb{Z}^d, z \neq 0, \sum_{z \neq 0} a(z) = 1$ and the rate of migration λ. Each particle, located at time t in some site $x \in \mathbb{Z}^d$ can jump to the point $(x + z) \in \mathbb{Z}^d$ with probability $\lambda a(z)dt$ (independently of the other particles).

d) *Immigration* depends on the local configuration, the probability kernel $q(z), z \in \mathbb{Z}^d, \sum_z q(z) = 1$ and the coefficient of intensity κ. If $n(t, x + z)$ is the configuration centered at $x \in \mathbb{Z}^d$, then during time interval $(t, t + dt)$ a new particle immigrates to the site x with probability $\kappa \sum_{z \in \mathbb{Z}^d} n(t, x + z)q(z)dt$.

We will see later that the condition $\mu = \beta + \kappa$ is necessary and sufficient for the criticality of the field $n(t, x)$, i.e., for the conservation of the mean density: $E(n(t, x)) \equiv \rho_0, t \geq 0$. Note that we need not consider emigration because its effect on our population will be indistinguishable from that of death.

In Section 2.3 we develop the moment theory for the critical case and, under additional assumptions (that stochastic dynamics are active enough for $d \leq 2$) we prove the existence of the limiting distribution. Roughly speaking, we assume that for $d = 2$ one or both of the densities $a(z)$ and $q(z)$ belong to the domain of attraction of a stable symmetric distribution with parameter $0 < \alpha < 2$ (note that symmetric for $d = 2$ does not mean isotropic). For $d = 1$ the density must be from the domain of attraction of a stable symmetric law with parameter $0 < \alpha < 1$. We prove that for $d = 2$ and *heavy tails spatial dynamics* (i.e., infinite second moment of the spatial distribution) the density of the second correlation function $k^{(2)}(t, x_1, x_2)$ has a nontrivial limit $k^{(2)}(\infty, x_1, x_2)$, $t \to \infty$. Together with the conservation of the first moment (density), $k^{(1)}(t, x) \equiv \rho_0$, it establishes the fundamental fact of *tightness* for the finite dimensional distributions of the point field $n(t, \cdot)$. In any limit theorem about the ergodicity (existence of the limit distribution) for the Markov process the proof of tightness is the first and most important step.

For $d \leq 2$ we also derive the asymptotic for the variance of particles in a region, with an eye toward establishing a central limit theorem, although we do not do that here.

The heavy tails assumption for the migration process warrants some discussion. It means that even if much population movement is to immediately proximate places (see [Ravenstein (1885)]), that is, occurs as diffusion, some population movement takes place over long distances, no matter how far. Our model assumes a distribution of population over a field that is infinite in all directions and assumes a positive probability of migration beyond any given distance. These assumptions obviously cannot hold for the environments of humans or other organisms, yet, as idealizations they reflect empirical situations. Within the continental United States, for example, there is considerable variation in migration distance: it is easy to calculate that the longest migrations in the continental United States are five or six orders of magnitude greater than the shortest. Relevant specifically to our assumption for the form of $a(z)$, according to Greenwood and Hunt [Greenwood and Hunt (2003)], the most popular modeling framework in the empirical analysis of geographic migration is the gravity model ([Lee (2006)], [Lowry (1966.)]), which models movement between two sites as inversely proportional to the square of the distance between the sites. There seems to have been little or no exploration of the fit of models in which the distance is raised to a power higher than two.

It is true, also, that many of the features that affect human population movement in the United States or any country, including geographic features, the distribution of economic opportunities, kinship ties, and various other push and pull factors (see [Lee (2006)]) create an asymmetric and heterogeneous migration environment that is absent from the migration process in our model. Similar heterogeneities exist in the environments of other organisms as well. The coarse features of such influences, however, probably can be accommodated with the variation that occurs in the spatial distribution $q(z)$, which, note, we assume to be symmetric but not isotropic. We also note support from other sources for such effects of heavy tail distributions. Biologists, recently, have become interested in the effects of long distance migration and have found that it may be able to make a qualitative difference in the global characteristics of the population and its development. For example, long distance dispersal can affect the spatial genetic structure, reducing genetic drift and the loss of genetic diversity [Fayard, Klein and Lefevre (2007)]. A number of social scientists, including the social psychologist George Herbert Mead, the Russian historian Lev Gumilev, and social movement researchers such as Timothy Wickham-Crowley, have argued for the importance of a few mobile, active individuals in stimulating change and altering global patterns.

Immigration consists of the appearance of a new particle at a given location, which may be populated already or may be empty. We assume that this process depends positively on the presence of other particles, as a probabilistic function of the number of particles at different distances from the location. This introduces some interaction between particles, namely, in their combined influence on the appearance of a new particle. This distinguishes immigration from birth, death, and migration, which occur for each particle independently of the other particles. Although the negative of the presence of particles on immigration is plausible–particles may avoid moving to locations that are overly crowded, for example–we assume here only a positive effect.

The process we call immigration differs from components labeled immigration in earlier models of branching processes. Several studies have added a random immigration process to the Galton-Watson process or other branching process without incorporating space (e.g., [Kawazu and Watanabe (1971)], [Pakes (1971b) Journal of the Australian Mathematical Society], [Pakes (1986) Advances in Applied Probability], [Li (2006)]). Other models have made immigration state-dependent, still without a spatial aspect (e.g., [Foster (1971)], [Pakes (1971a) Advances in Applied Probability],[Yamazato (1975)]). In these models, obviously, immigration cannot be affected by the spatial configuration of particles. Ivanov [Ivanov (1980)] and Milos [Milos (2009)] incorporate immigration into a process that includes both branching and movement that is uniformly stochastically continuous in $\mathbb{R}^d$. Immigration, however, occurs in time and space randomly according to a homogenous Poisson random field and, again, is not dependent on the spatial configuration of particles. Birkner [Birkner (2003)] considers immigration in a system consisting of independent Markov chains on a lattice. This immigration occurs at a constant rate at a single point x_0, and is locally dependent in that if x_0 is occupied at the instant that immigration is to take place then no immigration occurs. In contrast, we allow immigration at any point in the lattice and it is never blocked.

2.3 Derivation of the moment equations

To implement the heavy tails assumption for migration and immigration, we assume that $a(z)$ and $q(z)$ take the forms:

$$a(z) = \frac{h_1(\theta)}{|z|^{2+\alpha}} \left(1 + O(|z|^{-2})\right), \; z \neq 0$$

$$q(z) = \frac{h_2(\theta)}{|z|^{2+\alpha}} \left(1 + O(|z|^{-2})\right), \; z \neq 0$$

with $0 < \alpha < 2$, $\theta = \arg \frac{z}{|z|} \in (-\pi, \pi] = T^1$, $h_1, h_2 \in C^2(T^1)$, $h_1, h_2 > 0$. The second moments of the spatial distributions $a(z)$, $q(z)$ are infinite; that is, these distributions have heavy tails. This is easily seen. Using the l_1 norm for z (with, therefore, $4n$ locations at distance $|z| = n$), for some constant $c > 0$, noting also that h_1 and h_2 are bounded, and letting $f(z)$ represent either $a(z)$ or $q(z)$:

$$\sum_{z \in Z^2} |z|^2 f(z) = \frac{h(\theta)}{|z|^\alpha} \left(1 + O(|z|^{-2})\right) \geq c \sum_{z \in Z^2} |z|^{-\alpha} \left(1 + O(|z|^{-2})\right)$$
$$= 4c \sum_{n=1}^{\infty} n^{1-\alpha} \left(1 + O(|z|^{-2})\right) = \infty.$$

The stipulation that $\sum_{z \neq 0} a(z) = 1$ and $\sum_z q(z) = 1$ may be met by appropriate scaling of the bounded functions h_1 and h_2. This is because, again letting $f(z)$ represent either $a(z)$ or $q(z)$, the sum $\sum_{z \neq 0} f(z) = 4 \sum_{n=1}^{\infty} \frac{h(\theta)}{n^{1+\alpha}} (1 + O(n^{-2})) \sim \zeta(1 + \alpha)$, where $\zeta(s)$ is Riemann's zeta function, which converges for $s > 1$.

We summarize the process through the following expression:

$$n(t + dt, x) = n(t, x) + \xi_{dt}(t, x)$$

where the r.v. ξ is defined:

$$\xi_{dt}(t,x) = \begin{cases} +1 & \beta n(t,x)dt + \lambda \sum_{z \neq 0} a(z) n(t, x+z)dt + \kappa \sum_z q(z) n(t, x+z)dt \\ -1 & \mu n(t,x)dt + \lambda \sum_{z \neq 0} a(z) n(t,x)dt \\ 0 & 1 - (\beta + \mu) n(t,x)dt - \lambda \sum_{z \neq 0} a(z) n(t, x+z)dt - \lambda \sum_{z \neq 0} a(z) n(t,x)dt \\ & \quad -\kappa \sum_z q(z) n(t, x+z)dt \end{cases} \tag{2.1}$$

In words, in an infinitesimal time interval dt, $n(t, x)$, the number of particles at time t and at location x in the two-dimensional lattice, may increase by 1 due to one of the particles giving birth or due to migration of a particle from some point or due to immigration. It may decrease by 1 due to one of the $n(t, x)$ particles dying or due to one of them migrating elsewhere on the lattice. Last, it will remain the same if none of the listed changes occur. Note that the probability of more than one change occurring in the infinitesimal interval dt will be $O(dt^2)$ and can be ignored.

The generators for the migration process, $\lambda\mathcal{L}_a$, and for the immigration process, $\kappa\mathcal{L}_q$, are generalizations of the discrete Laplacian. The operators, $\mathcal{L}_a$ and $\mathcal{L}_q$, are defined:

$$\mathcal{L}_a f(x) := \sum_{z\neq 0} a(z)(f(x+z) - f(x))$$
$$\mathcal{L}_q f(x) := \sum_{z} q(z)(f(x+z) - f(x))$$

2.3.1 First moment

The differential equation and initial condition for the first moment $k_t^{(1)}(x)$ are:

$$\frac{\partial k_t^{(1)}(x)}{\partial t} = (\lambda\mathcal{L}_a + \kappa\mathcal{L}_q)k_t^{(1)}(x)) + (\beta - \mu + \kappa)k_t^{(1)}(x) \tag{2.2}$$
$$k_0^{(1)}(x) = \rho_0$$

Because of translation invariance, we have:

$$\frac{\partial k_t^{(1)}(x)}{\partial t} = (\beta - \mu + \kappa)k_t^{(1)}(x)$$

which has the solution:

$$k_t^{(1)}(x) = \rho_0 e^{(\beta-\mu+\kappa)t}$$

The critical case, therefore, is when $\mu = \beta + \kappa$, giving a model stationary in space and time for the first moment:

$$k_t^{(1)}(x) = \rho_0$$

The differential equation is derived as follows. Let $\mathcal{F}_{\leq t}$ denote the σ-algebra in the probability space corresponding to time t. Using the Kolmogorov forward equations with ξ_{dt} defined as in (2.1):

$$\begin{aligned}
k^{(1)}_{t+dt}(x) \quad &= E[E[n(t,x)+\xi_{dt}|\mathcal{F}_{\leq t}]] \\
&= E[k^{(1)}_t(x) + (\beta-\mu)n(t,x)dt + \lambda\sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt \\
&+\kappa\sum_z q(z)n(t,x+z)dt + O(dt^2)] \\
&= k^{(1)}_t(x) + (\beta+\kappa-\mu)k^{(1)}_t(x)dt + \lambda\sum_{z\neq 0} a(z)(k^{(1)}_t(x+z)-k^{(1)}_t(x))dt \\
&+\kappa\sum_z q(z)(k^{(1)}_t(x+z)-k^{(1)}_t(x))dt + O(dt^2)]
\end{aligned}$$

This gives differential equation (2.2).

2.3.2 Second moment

The differential equation and initial condition for the second moment $k^{(2)}_t(x,y)$ are:

$$\begin{aligned}
\frac{\partial k^{(2)}_t(x,y)}{\partial t} &= [\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay}) + \kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy})]k^{(2)}_t(x,y)) \\
&+ 2(\beta-\mu+\kappa)k^{(2)}_t(x,y) + s_2
\end{aligned} \tag{2.3}$$

$$k^{(2)}_0(x,y) = \rho_0^2 + \delta_0(x-y)\rho_0$$

where s_2, a source, is a function of the first moment:

$$\begin{aligned}
s_2 = \delta_0(x-y)[(\beta+\mu+\kappa+2\lambda)k^{(1)}_t(x) + (\lambda\mathcal{L}_a+\kappa\mathcal{L}_q)k^{(1)}_t(x)) \\
- a(y-x)\lambda(k^{(1)}_t(x)+k^{(1)}_t(y))
\end{aligned}$$

defining $a(0)=0$.

Substituting for the first moment and using translation invariance:

$$s_2 = [\delta_0(x-y)(\beta+\mu+\kappa+2\lambda) - 2a(y-x)\lambda]\rho_0 e^{(\beta-\mu+\kappa)t}$$

In the critical case ($\mu = \beta+\kappa$), therefore:

$$\frac{\partial k_t^{(2)}(x,y)}{\partial t} = [\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay})+\kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy})]k_t^{(2)}(x,y)) + s_2 \tag{2.4}$$
$$k_0^{(2)}(x,y) = \rho_0^2 + \delta_0(x-y)\rho_0$$

where $s_2 = 2[\delta_0(x-y)(\mu+\lambda) - a(y-x)\lambda]\rho_0$.

Below we derive differential equation (2.3) from the forward equations. Because the forward equations for $k_t^{(2)}(x,y)$ differ for the cases $x=y$ and $x \neq y$ we treat them separately. Subsequently, we use Fourier analysis to analyze equation (2.4).

Case 1, $x = y$

$$\begin{aligned} k_{t+dt}^{(2)}(x,x) \quad &= E[E[(n(t,x)+\xi_{dt}(t,x))(n(t,x)+\xi_{dt}(t,x))|\mathcal{F}_{\le t}]] \\ &= k_t^{(2)}(x,x) + E[2(\beta-\mu)n(t,x)^2 dt \\ &+2\lambda n(t,x)\sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt + 2\kappa n(t,x)\sum_z q(z)n(t,x+z)dt \\ &+(\beta+\mu)n(t,x)dt + \lambda\sum_{z\neq 0} a(z)(n(t,x+z)+n(t,x))dt \\ &+\kappa\sum_{z\neq 0} q(z)n(t,x+z)dt + O(dt^2)] \end{aligned}$$

Thus,

$$\begin{aligned} \frac{\partial k_t^{(2)}(x,x)}{\partial t} &= (2\lambda\mathcal{L}_{ax}+2\kappa\mathcal{L}_{qx})k_t^{(2)}(x,x)) + 2(\beta-\mu+\kappa)k_t^{(2)}(x,x) \\ &\quad + (\lambda\mathcal{L}_a+\kappa\mathcal{L}_q)k_t^{(1)}(x) + (\beta+\mu+\kappa+2\lambda)k_t^{(1)}(x) \end{aligned}$$

$$k_0^{(2)}(x,x) = \rho_0^2 + \rho_0$$

Case 2, $x \neq y$

$$\begin{aligned}
k^{(2)}_{t+dt}(x,y) \quad &= E[E[(n(t,x)+\xi_{dt}(t,x))(n(t,y)+\xi_{dt}(t,y))|\mathcal{F}_{\leq t}]] \\
&= k^{(2)}_t(x,y) + E[2(\beta-\mu)n(t,x)n(t,y)dt \\
&+\lambda n(t,x)\sum_{z\neq 0} a(z)(n(t,y+z)-n(t,y))dt + \kappa n(t,x)\sum_z q(z)n(t,y+z)dt \\
&+\lambda n(t,y)\sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt + \kappa n(t,y)\sum_z q(z)n(t,x+z)dt \\
&-\lambda a(x-y)(n(t,x)+n(t,y))dt + O(dt^2)]
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial k^{(2)}_t(x,y)}{\partial t} &= [\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay})+\kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy})]k^{(2)}_t(x,y)) + 2(\beta-\mu+\kappa)k^{(2)}_t(x,y) \\
&- \lambda a(x-y)(k^{(1)}_t(x)+k^{(1)}_t(y))
\end{aligned}$$

$$k^{(2)}(0,x,y) = \rho_0^2$$

2.3.3 Third moment

The differential equation and initial condition for the third moment $k^{(3)}_t(x,y,w)$ are:

$$\begin{aligned}
\frac{\partial k^{(3)}_t(x,y,w)}{\partial t} &= [\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay}+\mathcal{L}_{aw})+\kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy}+\mathcal{L}_{qw})]k^{(3)}_t(x,y,w)) \\
&+ 3(\beta-\mu+\kappa)k^{(3)}_t(x,y,w) + s_3
\end{aligned} \tag{2.5}$$

$$k_0^{(3)}(x,y,w) = \rho_0^3 + [\delta_0(x-y)+\delta_0(x-w)+\delta_0(y-w)]\rho_0^2 + [\delta_0(x-y)\delta_0(x-w)]\rho_0$$

where s_3, a source, is a function of the first and second moments:

$$\begin{aligned}
s_3 = (\beta+\mu+\kappa+2\lambda)\big(&\delta_0(x-y)[k_t^{(2)}(x,w) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,w)] \\
&+ \delta_0(x-w)[k_t^{(2)}(x,y) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,y)] \\
&+ \delta_0(y-w)[k_t^{(2)}(x,y) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,y)]\big) \\
&- \lambda\big(a(x-w)+a(y-w)\big)k_t^{(2)}(x,y) - \lambda\big(a(x-y)+a(y-w)\big)k_t^{(2)}(x,w) \\
&- \lambda\big(a(x-w)+a(y-x)\big)k_t^{(2)}(w,y) + \delta_0(x-y)\delta_0(x-w)(\beta-\mu+\kappa)k_t^{(1)}(x))
\end{aligned}$$

defining $a(0)=0$.

In the critical case ($\mu = \beta+\kappa$):

$$\frac{\partial k_t^{(3)}(x,y,w)}{\partial t} = [\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay}+\mathcal{L}_{aw}) + \kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy}+\mathcal{L}_{qw})]k_t^{(3)}(x,y,w)) + s_3 \tag{2.6}$$

$$k_0^{(3)}(x,y,w) = \rho_0^3 + [\delta_0(x-y)+\delta_0(x-w)+\delta_0(y-w)]\rho_0^2 + [\delta_0(x-y)\delta_0(x-w)]\rho_0$$

where s_3, a source, is a function of the first and second moments:

$$\begin{aligned}
s_3 = 2(\mu+\lambda)\big(&\delta_0(x-y)[k_t^{(2)}(x,w) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,w)] \\
&+ \delta_0(x-w)[k_t^{(2)}(x,y) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,y)] \\
&+ \delta_0(y-w)[k_t^{(2)}(x,y) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k_t^{(2)}(x,y)]\big) \\
&- \lambda\big(a(x-w)+a(y-w)\big)k_t^{(2)}(x,y) - \lambda\big(a(x-y)+a(y-w)\big) \\
&\cdot k_t^{(2)}(x,w) - \lambda\big(a(x-w)+a(y-x)\big)k_t^{(2)}(w,y)
\end{aligned}$$

Below we derive differential equation (2.5) from the forward equations. There are three cases to consider, $x=y=w$, $x=y\neq w$, and $x\neq y\neq w\neq x$.

Case 1, $x = y = w$

$$\begin{aligned}
k^{(3)}_{t+dt}(x,x,x) \quad &= E[E[(n(t,x)+\xi_{dt}(t,x))(n(t,x)+\xi_{dt}(t,x))(n(t,x)+\xi_{dt}(t,x))|\mathcal{F}_{\leq t}]] \\
&= k^{(3)}_t(x,x,x) + E[3(\beta+\kappa-\mu)n(t,x)^3 dt \\
&+3\lambda n(t,x)^2 \sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt \\
&+3\kappa n(t,x)^2 \sum_{z} q(z)(n(t,x+z)-n(t,x))dt \\
&+3(\beta+\mu)n(t,x)^2 dt + 3\lambda n(t,x)\sum_{z\neq 0} a(z)(n(t,x+z)+n(t,x))dt \\
&+3\kappa n(t,x)\sum_{z} q(z)n(t,x+z)dt + (\beta-\mu)n(t,x)dt \\
&+\lambda \sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt + \kappa \sum_{z} q(z)n(t,x+z)dt + O(dt^2)]
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial k^{(3)}_t(x,x,x)}{\partial t} &= (3\lambda\mathcal{L}_{ax} + 3\kappa\mathcal{L}_{qx})k^{(3)}_t(x,x,x)) + 3(\beta-\mu+\kappa)k^{(3)}_t(x,x,x) \\
&+ 3(\beta+\mu+\kappa+2\lambda)k^{(2)}_t(x,x) + (3\lambda\mathcal{L}_{ax} + 3\kappa\mathcal{L}_{qx})k^{(2)}_t(x,x) \\
&+ (\beta-\mu+\kappa)k^{(1)}_t(x) + (\lambda\mathcal{L}_a + \kappa\mathcal{L}_q)k^{(1)}_t(x)
\end{aligned}$$

$$k^{(3)}_0(x,x,x) = \rho_0^3 + 3\rho_0^2 + \rho_0$$

Case 2, $x = y \neq w$

$$
\begin{aligned}
k^{(3)}_{t+dt}(x,x,w) \quad &= E[E[(n(t,x)+\xi_{dt}(t,x))(n(t,x)+\xi_{dt}(t,x))(n(t,w)+\xi_{dt}(t,w))|\mathcal{F}_{\leq t}]] \\
&= k^{(3)}_t(x,x,w) + E[3(\beta-\mu)n(t,x)^2 n(t,w)dt \\
&+2\lambda n(t,x)n(t,w)\sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt \\
&+\lambda n(t,x)^2 \sum_{z\neq 0} a(z)(n(t,w+z)-n(t,w))dt \\
&+2\kappa n(t,x)n(t,w)\sum_z q(z)n(t,x+z)dt + \kappa n(t,x)^2 \sum_z q(z)n(t,w+z)dt \\
&-2\lambda n(t,x)a(x-w)\big(n(t,x)+n(t,w)\big) + (\beta+\mu)n(t,x)n(t,w) \\
&+\lambda n(t,w)\sum_{z\neq 0} a(z)(n(t,x+z)+n(t,x))dt \\
&+\kappa n(t,w)\sum_z q(z)n(t,x+z)dt + \lambda a(x-w)(n(t,x)-n(t,w))dt + O(dt^2)]
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{\partial k^{(3)}_t(x,x,w)}{\partial t} &= \big(\lambda(2\mathcal{L}_{ax}+\mathcal{L}_{aw}) + \kappa(2\mathcal{L}_{qx}+\mathcal{L}_{qw})\big)k^{(3)}_t(x,x,w)) \\
&+ 3(\beta-\mu+\kappa)k^{(3)}_t(x,x,w) - 2\lambda a(w-x)(k^{(2)}_t(x,x) + k^{(2)}_t(x,w) \\
&+ (\beta+\mu+\kappa+2\lambda)k^{(2)}_t(x,w) + (\lambda\mathcal{L}_{ax}+\kappa\mathcal{L}_{qx})k^{(2)}_t(x,w) \\
&+ \lambda a(w-x)(k^{(1)}_t(x) - k^{(1)}_t(w))
\end{aligned}
$$

$$
k^{(3)}_0(x,x,w) = \rho_0^3 + \rho_0^2
$$

Case 3, $x \neq y \neq w \neq x$

$$
\begin{aligned}
k^{(3)}_{t+dt}(x,y,w) \quad &= E[E[(n(t,x)+\xi_{dt}(t,x))(n(t,y)+\xi_{dt}(t,y))(n(t,w)+\xi_{dt}(t,w))|\mathcal{F}_{\le t}]] \\
&= k^{(3)}_t(x,y,w) + E[3(\beta-\mu)n(t,x)n(t,y)n(t,w)dt \\
&+\lambda n(t,x)n(t,y)\sum_{z\neq 0} a(z)(n(t,w+z)-n(t,w))dt \\
&+\lambda n(t,x)n(t,w)\sum_{z\neq 0} a(z)(n(t,y+z)-n(t,y))dt \\
&+\lambda n(t,y)n(t,w)\sum_{z\neq 0} a(z)(n(t,x+z)-n(t,x))dt \\
&+\kappa n(t,x)n(t,y)\sum_{z} q(z)n(t,w+z)dt + \kappa n(t,x)n(t,w)\sum_{z} q(z)n(t,y+z)dt \\
&+\kappa n(t,y)n(t,w)\sum_{z} q(z)n(t,x+z)dt \\
&-\lambda a(x-w)n(t,y)\big(n(t,x)+n(t,w)\big)dt - \lambda a(y-w)n(t,x) \\
&\cdot\big(n(t,y)+n(t,w)\big)dt - \lambda a(x-y)n(t,w)\big(n(t,x)+n(t,y)\big)dt + O(dt^2)]
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\frac{\partial k^{(3)}_t(x,y,w)}{\partial t} &= \big(\lambda(\mathcal{L}_{ax}+\mathcal{L}_{ay}+\mathcal{L}_{aw}) + \kappa(\mathcal{L}_{qx}+\mathcal{L}_{qy}+\mathcal{L}_{qw})\big)k^{(3)}_t(x,y,w)) \\
&+ 3(\beta-\mu+\kappa)k^{(3)}_t(x,y,w) - \lambda\big(a(w-x)+a(y-w)\big)k^{(2)}_t(x,y) \\
&- \lambda\big(a(x-y)+a(y-w)\big)k^{(2)}_t(x,w) - \lambda\big(a(x-w)+a(y-x)\big)k^{(2)}_t(w,y)
\end{aligned}
$$

$$k^{(3)}_0(x,y,w) = \rho_0^3$$

2.3.4 Recursive moment equations

It is clear, by inspecting the differential equations and their derivation for the first three moments, that the differential equations for all moments can be stated. The differential equations take a recursive form; specifically, the differential equation for the nth moment involves only the nth and $(n-1)$th moments. The differential equations become more complicated as n increases. For $x_1, x_2, ..., x_n$ distinct, however, the recursive form for the differential equations for $k^{(n)}_t(x_1, \cdots, x_n)$ may be written simply in the critical case ($\mu = \beta + \kappa$):

$$\frac{\partial k_t^{(1)}(x)}{\partial t} = (\beta + \kappa - \mu) k_t^{(1)}(x) = 0 \Rightarrow k_t^1(x) \equiv \rho_0$$

$$\begin{aligned}\frac{\partial k_t^{(n)}(x_1, ..., x_n)}{\partial t} &= \left(\lambda \sum_{j=1}^{n} \mathcal{L}_{ax_j} + \kappa \sum_{j=1}^{n} \mathcal{L}_{qx_j} \right) k_t^{(n)}(x_1, ..., x_n) \\ &\quad - \lambda \sum_{i=1}^{n} \left(k_t^{(n-1)}(x_1, ..., x_{i-1}, x_{i+1}, ..., x_n) \sum_{j \neq i} a(x_i - x_j) \right)\end{aligned}$$

with initial conditions

$$k_0^{(n)}(x_1, ..., x_n) = \rho_0^n, \; n \geq 1.$$

2.4 Fourier analysis of the second moment

We begin by determining the Fourier transforms of $\mathcal{L}_a$ and $\mathcal{L}_q$. Define $\Phi_a(\varphi) := \sum_{z \neq 0} a(z)(1 - \cos(\varphi \cdot z))$ and $\Phi_q(\varphi) := \sum_{z \neq 0} q(z)(1 - \cos(\varphi \cdot z))$. Then

$$-\widehat{\mathcal{L}_a f(x)} = \hat{f}(\varphi) \Phi_a(\varphi)$$

$$-\widehat{\mathcal{L}_q f(x)} = \hat{f}(\varphi) \Phi_q(\varphi)$$

This follows from:

$$\begin{aligned}-\widehat{\mathcal{L}_a f(x)} &= -\sum_x e^{i(\varphi \cdot x)} \sum_{z \neq 0} a(z)(f(x+z) - f(x)) \\ &= -\sum_{z \neq 0} a(z) \left(e^{-i(\varphi \cdot z)} \sum_x e^{i(\varphi \cdot (x+z))} (f(x+z) - \sum_x e^{i(\varphi \cdot x)} f(x)) \right) \\ &= -\sum_{z \neq 0} a(z)(e^{-i(\varphi \cdot z)} - 1) \hat{f}(\varphi) = \hat{f}(\varphi) \sum_{z \neq 0} a(z)(1 - cos(\varphi \cdot z)).\end{aligned}$$

Note that the Fourier transform of $\delta_0(x - y)$ is:

$$\sum_{x,y} \delta_0(x-y) e^{i(\varphi_1 \cdot x) + i(\varphi_2 \cdot y)} = \sum_x 1 \cdot e^{i((\varphi_1 + \varphi_2) \cdot x)} = (2\pi)^2 \delta_0(\varphi_1 + \varphi_2).$$

We will need, also, letting $< x, y >$ denote a point in $\mathbb{Z}^2$ x $\mathbb{Z}^2$:

$$\begin{aligned}\hat{a}(x-y) &= \sum_{<x,y>} e^{i((\varphi_1\cdot x)+(\varphi_2\cdot y))}a(y-x) = \sum_{<x,z>} e^{i((\varphi_1\cdot x)+(\varphi_2\cdot(x+z)))}a(z) \\ &= \sum_{<x,z>} e^{i(((\varphi_1+\varphi_2)\cdot x)+(\varphi_2\cdot z))}a(z) = \sum_x e^{i((\varphi_1+\varphi_2)\cdot x)}\sum_z e^{i(\varphi_2\cdot z)}a(z) \\ &= \sum_x e^{i((\varphi_1+\varphi_2)\cdot x)}\sum_z a(z)cos(\varphi_2\cdot z) = 4\pi^2\delta_0(\varphi_1+\varphi_2)(1-\Phi_a(\varphi_2)).\end{aligned}$$

Define $\psi_t(x, y) := k_t^{(2)}(x, y) - \rho_0^2$. Then, using equation (2.4), we get the Fourier transform:

$$\begin{aligned}\frac{\partial\hat{\psi}(t,\varphi_1,\varphi_2)}{\partial t} &= -\left(\lambda(\Phi_a(\varphi_1)+\Phi_a(\varphi_2))+\kappa(\Phi_q(\varphi_1)+\Phi_q(\varphi_2))\right)\hat{\psi}(t,\varphi_1,\varphi_2)+\hat{s}_2 \\ \hat{\psi}(0,\varphi_1,\varphi_2) &= 4\pi^2\rho_0\delta_0(\varphi_1+\varphi_2). \qquad (2.7)\end{aligned}$$

This has the solution:

$$\hat{\psi}(t,\varphi_1,\varphi_2) = \left(4\pi^2\rho_0\delta_0(\varphi_1+\varphi_2) - \frac{\hat{s}_2}{A}\right)e^{-At} + \frac{\hat{s}_2}{A},$$

where $A := \lambda(\Phi_a(\varphi_1) + \Phi_a(\varphi_2)) + \kappa(\Phi_q(\varphi_1) + \Phi_q(\varphi_2))$.

Also:

$$\hat{s}_2 = 2\rho_0\left((\mu+\lambda)\hat{\delta}_0(x-y) - \lambda\hat{a}(y-x)\right) = 8\pi^2\rho_0\delta_0(\varphi_1+\varphi_2)(\mu+\lambda\Phi_a(\varphi_2)).$$

Thus, for $\varphi_1 + \varphi_2 = 0$:

$$\hat{\psi}(t,\varphi_1,\varphi_2) = \left(4\pi^2\rho_0 - \frac{4\pi^2\rho_0(\mu+\lambda\Phi_a(\varphi_1))}{\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2)}\right)e^{-2(\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2))t} + \frac{4\pi^2\rho_0(\mu+\lambda\Phi_a(\varphi_1))}{\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2)},$$

and for $\varphi_1 + \varphi_2 \neq 0$: $\hat{\psi}(t, \varphi_1, \varphi_2) = 0$.

This shows that the process is stationary in space, because $\psi(t, x, y)$ and, therefore, $k^{(2)}(t, x, y)$ are functions of $y - x$ only.

Define $\hat{\psi}(\varphi_1, \varphi_2) := \lim_{t\to\infty} \hat{\psi}(t, \varphi_1, \varphi_2)$. Then if $\int_{T^2}\int_{T^2} \frac{d\varphi_1 d\varphi_2}{\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2)} < \infty$,

$$\hat{\psi}(\varphi_1,\varphi_2) = \delta_0(\varphi_1+\varphi_2)\frac{4\pi^2\rho_0(\mu+\lambda\Phi_a(\varphi_1))}{\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2)}$$

Taking the inverse Fourier transform:

$$\psi(x,y) = \frac{1}{(2\pi)^4}\int\limits_{T^2}\int\limits_{T^2}\delta_0(\varphi_1+\varphi_2)\frac{4\pi^2\rho_0(\mu+\lambda\Phi_a(\varphi_1))}{\lambda\Phi_a(\varphi_1)+\kappa\Phi_q(\varphi_2)}e^{-i(\varphi_1\cdot x)-i(\varphi_2\cdot y)}d\varphi_1 d\varphi_2$$
$$= \frac{\rho_0}{(2\pi)^2}\int\limits_{T^2}\frac{\mu+\lambda\Phi_a(\varphi)}{\lambda\Phi_a(\varphi)+\kappa\Phi_q(\varphi)}e^{-i(\varphi\cdot(y-x))}d\varphi.$$

When this integral exists—the issue is the singularity at $\varphi = 0$—$k^{(2)}(x,y)$ will be finite and, therefore, the process will be transient and patches will not occur.

The integral exists if $\int\limits_{D^2}\frac{d\varphi}{\lambda\Phi_a(\varphi)+\kappa\Phi_q(\varphi)} < \infty$ for some region $D^2 \subset T^2$ around $(0,0)$. We show this is true if $\lambda\Phi_a(\varphi)+\kappa\Phi_q(\varphi) = O(|\varphi|^\alpha)$ for $0<\alpha<2$. First, however, we prove the following lemma.

Lemma 2.4. *Suppose the migration process or immigration process has the spatial distribution*

$$f(z) = \frac{h(\theta)}{|z|^{2+\alpha}}\left(1+O\left(\frac{1}{|z|^2}\right)\right),\ z\neq 0$$

with $0<\alpha<2$, $\theta = \arg\frac{z}{|z|} \in [-\pi,\pi) = T^1$, $h\in C^2(T^1)$, $h>0$ *and so satisfies the heavy tails assumption. Then, as* $|\varphi|\to 0$, $\Phi_f(\varphi) = O(|\varphi|^\alpha)$.

Proof. We have $\Phi_f(\varphi) := \sum_{z\neq 0} f(z)(1-\cos(\varphi\cdot z))$. Let us consider the following integral $I(\varphi)$, which will give a good approximation of $\Phi_f(\varphi)$, $\varphi\in[-\pi,\pi)^2 = T^2$:

$$I(\varphi) = \int\limits_{\mathbb{R}^2-A(0)}\frac{d\vec{x}}{|\vec{x}|^{2+\alpha}}h(\frac{\vec{x}}{|\vec{x}|})(1-\cos(\varphi\cdot\vec{x}))$$

Here, $A(0) = \{\vec{x} : |x_1|\le\frac{1}{2}, |x_2|\le\frac{1}{2}\}$ and, in general, $A(\vec{n}) = \{\vec{x} : |x_1-n_1|\le\frac{1}{2}, |x_2-n_2|\le\frac{1}{2}\}$, $\vec{n} = (n_1,n_2)\in\mathbb{Z}^2$. Note that in $I(\varphi)$, $\arg\varphi\in T^1$ can be arbitrary, but because of the singularity at $\varphi = 0$ we are concerned primarily with the situation $|\varphi| << 1$.

Put $I_{\vec{n}}(\varphi) = \int\limits_{A(n)}\frac{h(\frac{\vec{x}}{|\vec{x}|})}{|\vec{x}|^{2+\alpha}}(1-\cos(\varphi\cdot\vec{x}))dx$, $I(\varphi) = \sum_{\vec{n}\neq 0} I_{\vec{n}}(\varphi)$, and make the substitution

$\vec{x} = \vec{n} + \vec{w}$, $\vec{w} \in A(0)$. Then, $|\vec{x}| = (|\vec{n}|^2 + 2(\vec{n}\cdot\vec{w}) + |\vec{w}|^2)^{1/2}$. It gives

$$\frac{\vec{x}}{|\vec{x}|} = \frac{\vec{n}}{|\vec{n}|} + \frac{\vec{w}}{|\vec{n}|} - \frac{\vec{n}(\vec{n}\cdot\vec{w})}{|\vec{n}|^3} + O\left(\frac{1}{|\vec{n}|^2}\right)$$

and

$$\frac{1}{|\vec{x}|^{2+\alpha}} = \frac{1}{|\vec{n}|^{2+\alpha}}\left(1 - \frac{(2+\alpha)(\vec{n}\cdot\vec{w})}{|\vec{n}|^2} + O\left(\frac{1}{|\vec{n}|^2}\right)\right)$$

Using the Taylor series expansion of the integrant we can present $I_{\vec{n}}(\varphi)$ in the form

$$\begin{aligned} I_{\vec{n}}(\varphi) &= \frac{1}{|\vec{n}|^{2+\alpha}}\int\limits_{A(0)}[h\left(\frac{\vec{n}}{|\vec{n}|}\right) + \frac{1}{|\vec{n}|}\nabla h\left(\frac{\vec{n}}{|\vec{n}|}\right)\cdot\vec{w} - \frac{(\vec{n}\cdot\nabla h)(\vec{n}\cdot\vec{w})}{|\vec{n}|^3} + O(\frac{1}{|\vec{n}|^2})] \\ &\cdot[1 - \frac{(2+\alpha)(\vec{n}\cdot\vec{w})}{|\vec{n}|^2} + O\left(\frac{1}{|\vec{n}|^2}\right)][1 - \cos(\vec{n}\cdot\varphi)\cos(\vec{w}\cdot\varphi) + \sin(\vec{n}\cdot\varphi)\sin(\vec{w}\cdot\varphi)]d\vec{w} \\ &= \frac{1}{|\vec{n}|^{2+\alpha}}\int\limits_{A(0)}[h\left(\frac{\vec{n}}{|\vec{n}|}\right) + \frac{\nabla h\cdot\vec{w}}{|\vec{n}|} - \frac{(\vec{n}\cdot\nabla h)(\vec{n}\cdot\vec{w})}{|\vec{n}|^3} - h\left(\frac{\vec{n}}{|\vec{n}|}\right)\frac{(2+\alpha)(\vec{n}\cdot\vec{w})}{|\vec{n}|^2} + O\left(\frac{1}{|\vec{n}|^2}\right)] \\ &\cdot[(1-\cos(\vec{n}\cdot\varphi)) + \cos(\vec{n}\cdot\varphi)(1-\cos(\vec{w}\cdot\varphi)) - \sin(\vec{n}\cdot\varphi)\sin(\vec{w}\cdot\varphi)]d\vec{w} \end{aligned}$$

For large $|\vec{n}|$, the leading term in the expansion of $I_{\vec{n}}(\varphi)$ is equal to

$$\psi_{\vec{n}}(\varphi) := \frac{1}{|\vec{n}|^{2+\alpha}}h\left(\frac{\vec{n}}{|\vec{n}|}\right)(1 - \cos(\vec{n}\cdot\varphi))$$

and $\sum\limits_{\vec{n}\neq 0}\psi_{\vec{n}}(\varphi) = \Phi_f(\varphi)$.

Let us note also that the integrals containing linear functions of $\vec{w}$ are vanishing:

$$\int\limits_{A(0)}(\vec{c}\cdot\vec{w})d\vec{w} = 0,\ \forall\vec{c}\in\mathbb{R}^2$$

$$\int\limits_{A(0)}(1 - \cos(\vec{w}\cdot\vec{\varphi}))(\vec{c}\cdot\vec{w})d\vec{w} = 0$$

The integrals $\int\limits_{A(0)}\sin(\vec{w}\cdot\vec{\varphi})(\vec{c}\cdot\vec{w})d\vec{w}$ are non-vanishing but they have ahead the odd over $\vec{n}$ factor $\sin(\vec{n}\cdot\vec{\varphi})$ and after summation over $\vec{n}\neq 0$ the corresponding contribution is equal to zero.

The remainders $O(\frac{1}{|\vec{n}|^2})$ will give a contribution of order $O(|\varphi|^2)$, $|\varphi| \to 0$ (since $\sum_{\vec{n}\neq 0} \frac{1}{|\vec{n}|^{4+\alpha}}(1-\cos(\vec{n}\cdot\varphi)) \in C^2(T^2)$ $\forall(\alpha > 0)$.

Finally,

$$\Phi_f(\varphi) = I(\varphi) + O(|\varphi|^2),\ |\varphi| \to 0$$

with

$$\begin{aligned} I(\varphi) &= \int\limits_{\mathbb{R}^2 - A(0)} \frac{h(\frac{\vec{x}}{|\vec{x}|})}{|\vec{x}|^{2+\alpha}}(1-\cos(\varphi\cdot\vec{x}))d\vec{x} \\ &= \int\limits_{|\vec{x}|\geq 1} \frac{h(\frac{\vec{x}}{|\vec{x}|})}{|\vec{x}|^{2+\alpha}}(1-\cos(\varphi\cdot\vec{x}))d\vec{x} \end{aligned}$$

If $\vec{x} = (x_1, x_2) = r(\cos\theta, \sin\theta)$, $\varphi = |\varphi|(\cos\gamma, \sin\gamma)$, then

$$\begin{aligned} \Phi_f(\varphi) &= \int\limits_1^\infty \frac{dr\cdot r}{r^{2+\alpha}} \int\limits_{-\pi}^{\pi} h(\theta)(1-\cos(r|\varphi|\cdot|\cos(\theta-\gamma)|))d\theta + O(|\varphi|^2) \\ &= \int\limits_{-\pi}^{\pi} h(\theta) \int\limits_1^\infty \frac{dr}{r^{1+\alpha}}(1-\cos(\varepsilon r))d\theta + O(|\varphi|^2) \end{aligned}$$

where $\varepsilon = |\varphi||\cos(\theta-\gamma)|$. Using the substitution $t = \varepsilon r$ we obtain

$$\Phi_f(\varphi) = O(|\varphi|^2) + |\varphi|^\alpha \int\limits_{-\pi}^{\pi} d\theta h(\theta)|\cos(\theta-\gamma)|^\alpha \cdot \int\limits_{|\varphi||\cos(\varphi-\theta)|}^{\infty} \frac{1-\cos t}{t^{1+\alpha}}dt$$

But $\int\limits_{|\varphi||\cos(\varphi-\theta)|}^{\infty} \frac{1-\cos t}{t^{1+\alpha}}dt = c_\alpha - O(|\varphi|^{2-\alpha})$, with $c_\alpha := \int\limits_0^\infty \frac{1-\cos t}{t^{1+\alpha}}dt$. Set $\mathcal{H}(\gamma) := \int\limits_{-\pi}^{\pi} h(\theta)|\cos(\theta-\gamma)|^\alpha d\theta$, $\gamma = \arg\varphi$, $\mathcal{H}(\gamma) \in C(T^1)$, $\mathcal{H}(\gamma) > 0$. Then:

$$\Phi_f(\varphi) = c_\alpha|\varphi|^\alpha\mathcal{H}(\gamma) + O(|\varphi|^2),\ |\varphi| \to 0$$

□

Under the assumptions of the Lemma, therefore, in some region $D^2 \subset T^2$ around $(0,0)$, we have $\Phi_a(\varphi) \sim |\varphi|^\alpha$ or $\Phi_q(\varphi) \sim |\varphi|^\alpha$, $0 < \alpha < 2$. It follows that $\int\limits_{D^2} \frac{d\varphi}{\alpha\Phi_a(\varphi_1) + \kappa\Phi_q(\varphi_2)} < \infty$ for the region D^2, and, therefore that the inverse Fourier transform $\psi(x, y)$ exists.

This, in turn, means that the process for $d = 2$ is transient for $0 < \alpha < 2$ and so clusterization does not occur.

For $d = 1$, with appropriate adjustments, such as setting $a(z)$ or $q(z)$ to $\frac{c}{|z|^{2+\alpha}}$ for some constant c, $0 < c < \infty$, the process is transient if $0 < \alpha < 1$. The proof is nearly identical.

2.5 The variance of the population of a region Q_r

We consider the critical branching process with migration and immigration in one or two dimensions, as described previously. In this section we derive the asymptotic for the variance of the number of particles in a region as the size of the region increases. Q_r denotes a ball of radius r with the center at the origin, let $n(Q_r)$ denote the number of particles in Q_r as $t \to \infty$. We have the following theorem:

Theorem 2.5. *Assume the critical branching process with migration and immigration analyzed above. In particular, assume* $0 < \alpha < d$ *for dimension* $d \leq 2$*. Then, for* $d \leq 2$*, if* Q_r *denotes a ball of radius* r *with the center at the origin, then as* r *increases, the variance of the number of particles in* Q_r *grows as* $r^{d+\alpha}$*.*

Proof. Let $n(Q_r)$ denote the number of particles in Q_r as $t \to \infty$ and, as earlier, define $\psi_t(x, y) := k_t^{(2)}(x, y) - \rho_0^2$. Then, the variance of $n(Q_r)$ is:

$$\mathrm{Var}\,(n(Q_r)) = \sum_{x,y \in Q_r} k^{(2)}(x, y) - \sum_{x,y \in Q_r} \rho_0^2 = \sum_{x,y \in Q_r} \psi(x, y).$$

Because of the spatial invariance of $k^{(2)}$ and ψ, we set $B(x - y) := \psi(x, y)$.

As r increases, $\sum_{x,y \in Q_r} \psi(x, y) \to \int_{Q_r} \int_{Q_r} B(x - y) dx dy$. Calling this V, we use Fourier transforms to obtain its asymptotic as r grows large.

$$\begin{aligned}
V := &\int\limits_{Q_r}\int\limits_{Q_r} B(x-y)dxdy = \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} I_{Q_r}(x)I_{Q_r}(y)B(x-y)dxdy \\
&= \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} dxdy\frac{1}{(2\pi)^{3d}}\int\limits_{T^d}\int\limits_{T^d}\int\limits_{T^d} \hat{I}_{Q_r}(\varphi_x)e^{-i(\varphi_x\cdot x)}\hat{I}_{Q_r}(\varphi_y)e^{-i(\varphi_y\cdot y)}\hat{B}(\varphi_z)e^{-i(\varphi_z\cdot(x-y))}d\varphi_x d\varphi_y d\varphi_z \\
&= \int\limits_{\mathbb{R}^d}\int\limits_{\mathbb{R}^d} dxdy\frac{1}{(2\pi)^{3d}}\int\limits_{T^d}\int\limits_{T^d}\int\limits_{T^d} \hat{I}_{Q_r}(\varphi_x)\hat{I}_{Q_r}(\varphi_y)\hat{B}(\varphi_z)e^{-i((\varphi_x+\varphi_z)\cdot x)}e^{-i((\varphi_y-\varphi_z)\cdot y)}d\varphi_x d\varphi_y d\varphi_z \\
&= \frac{1}{(2\pi)^d}\int\limits_{T^d}\int\limits_{T^d}\int\limits_{T^d} \delta_0(\varphi_x+\varphi_z)\delta_0(\varphi_y-\varphi_z)\hat{I}_{Q_r}(\varphi_x)\hat{I}_{Q_r}(\varphi_y)\hat{B}(\varphi_z)d\varphi_x d\varphi_y d\varphi_z \\
&= \frac{1}{(2\pi)^d}\int\limits_{T^d} \hat{I}_{Q_r}(-\varphi_z)\hat{I}_{Q_r}(\varphi_z)\hat{B}(\varphi_z)d\varphi_z.
\end{aligned}$$

We can write the Fourier transform of the indicator function I as: :

$$\hat{I}_{Q_r}(\varphi) = \int\limits_{Q_r} e^{i(\varphi,x)}dx = \left(\frac{2\pi r}{|\varphi|}\right)^{d/2} J_{d/2}(r|\varphi|),$$

where $J_{d/2}$ is the Bessel function of the first kind of order $\frac{d}{2}$ [Gikhman and Skorokhod (1974)]. Also including the result for $\hat{B}(\varphi_z) = \rho_0\frac{\mu+\lambda\Phi_a(\varphi)}{\lambda\Phi_a(\varphi)+\kappa\Phi_q(\varphi)}$ from section 2.4, we obtain:

$$V = \rho_0 r^d \int\limits_{T^d} \frac{1}{|\varphi|^d}\left(J_{d/2}(r|\varphi|)\right)^2 \frac{\mu+\lambda\Phi_a(\varphi)}{\lambda\Phi_a(\varphi)+\kappa\Phi_q(\varphi)}d\varphi. \tag{2.8}$$

For $d = 2$, we have from Lemma 2.1 that there is some $\varepsilon > 0$ such that for $|\varphi| < \varepsilon$ either $\Phi_a(\varphi) \sim |\varphi|^\alpha$ or $\Phi_q(\varphi) \sim |\varphi|^\alpha$ or both. Together with the symmetry of Φ_a and Φ_q, this means we can write $\lambda\Phi_a(\varphi) + \kappa\Phi_q(\varphi) = |\varphi|^\alpha A(|\varphi|)$, where for $|\varphi| < \varepsilon$ we have $A(|\varphi|) \sim 1$ and for $|\varphi| \geq \varepsilon$ we have $0 < A(|\varphi|) \leq \frac{2(\lambda+\kappa)}{\varepsilon^\alpha}$.

We now can write (2.8) in polar coordinates. This gives:

$$V \sim \rho_0 r^d \int\limits_0^R \frac{c_d}{|\varphi|^{1+\alpha}}\left(J_{d/2}(r|\varphi|)\right)^2 \frac{\mu+\lambda\Phi_a(|\varphi|)}{A(|\varphi|)}d|\varphi|,$$

where $c_d = \frac{d\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$.

We set $x := r|\varphi|$ to get:

$$V \sim \rho_0 c_d r^{d+\alpha} \int_0^{rR} \frac{1}{x^{1+\alpha}} \left(J_{d/2}(x)\right)^2 \frac{\mu + \lambda\Phi_a(\frac{x}{r})}{A(\frac{x}{r})} dx.$$

Then, as $r \to \infty$ we have $A(\frac{x}{r}) \to A(0) \sim 1$ and $\Phi_a(\frac{x}{r}) \to 0$. Consequently, as $r \to \infty$,

$$V = \frac{\rho_0 c_d \mu}{A(0)} r^{d+\alpha} \int_0^{\infty} \frac{1}{x^{1+\alpha}} \left(J_{d/2}(x)\right)^2 dx.$$

Finally, for small x, $\left(J_{d/2}(x)\right)^2 \leq cx^d$ and as x becomes large, $\left(J_{d/2}(x)\right)^2 < \frac{1}{x}$, which means that $\int_0^{\infty} \frac{1}{x^{1+\alpha}} \left(J_{d/2}(x)\right)^2 dx$ converges, giving that:

$$V \xrightarrow[r\to\infty]{} c_0 r^{d+\alpha}$$

where $c_0 = \frac{\rho_0 c_d \mu}{A(0)} \int_0^{\infty} \frac{1}{x^{1+\alpha}} \left(J_{d/2}(x)\right)^2 dx$.

Except for a different c_0 the same argument holds for dimension $d = 1$.

□

Note that the result for our process in Theorem 2.2 contrasts with case of independent particles in a region. There, the variance of the number of particles grows simply as r^d. Our result can be used to establish a central limit theorem for our process, although we do not do so in this paper.

2.6 The subcritical case ($\mu > \beta + \kappa$)

Unsurprisingly, in the subcritical case the population degenerates everywhere. We establish this by showing that the first and second moments tend to 0 as t increases.

Theorem 2.6. *Assume the branching process with migration and immigration analyzed above, but in the subcritical condition. That is, assume that $\mu > \beta + \kappa$. Then, as $t \to \infty$, the first moment $k^{(1)}(t, x) \to 0$ and the second moment $k^{(2)}(t, x, y) \to 0$.*

Proof. For the first moment we have:

$$k^{(1)}(t,x) = \rho_0 e^{-(\mu-\beta-\kappa)t} \xrightarrow[t\to\infty]{} 0.$$

The Fourier transform of the differential equation for the second moment, using equation (2.3) and the subsequent equation for s_2, is:

$$\begin{aligned}\frac{\partial \hat{k}^{(2)}(t,\varphi_1,\varphi_2)}{\partial t} =& -\left(\lambda(\Phi_a(\varphi_1)+\Phi_a(\varphi_2))+\kappa(\Phi_q(\varphi_1)+\Phi_q(\varphi_2))\right)\hat{k}^{(2)}(t,\varphi_1,\varphi_2)\\ &+2(\beta+\kappa-\mu)\hat{k}^{(2)}(t,\varphi_1,\varphi_2)+\hat{s}_2\\ \hat{k}^{(2)}(0,\varphi_1,\varphi_2) =&\ 2\pi\rho_0^2+4\pi^2\rho_0\delta_0(\varphi_1+\varphi_2).\end{aligned}$$

Here,

$$\hat{s}_2 = 4\pi^2\delta_0(\varphi_1+\varphi_2)\rho_0(\beta+\mu+\kappa+2\lambda\Phi_a(\varphi_2))e^{-(\mu-\beta-\kappa)t}.$$

To simplify the calculation, set:

$$\begin{aligned}&A := -\left(\lambda(\Phi_a(\varphi_1)+\Phi_a(\varphi_2))+\kappa(\Phi_q(\varphi_1)+\Phi_q(\varphi_2))\right)\hat{k}^{(2)}(t,\varphi_1,\varphi_2)-2(\mu-\beta-\kappa)<0,\\ &B := 4\pi^2\delta_0(\varphi_1+\varphi_2)\rho_0(\beta+\mu+\kappa+2\lambda\Phi_a(\varphi_2)),\\ &C := \mu-\beta-\kappa>0.\end{aligned}$$

We have $\frac{\partial \hat{k}^{(2)}(t,\varphi_1,\varphi_2)}{\partial t} = A\hat{k}^{(2)}(t,\varphi_1,\varphi_2)+Be^{-Ct}$, which has the solution:

$$\hat{k}^{(2)}(t,\varphi_1,\varphi_2) = \frac{-B}{A+C}e^{-Ct}+\left(\hat{k}^{(2)}(0,\varphi_1,\varphi_2)+\frac{B}{A+C}\right)e^{At}.$$

Noting that $-\infty < A+C < 0$, that $C>0$ and $A<0$, and that B is finite, we have that:

$$\hat{k}^{(2)}(\infty,\varphi_1,\varphi_2) := \lim_{t\to\infty}\hat{k}^{(2)}(t,\varphi_1,\varphi_2) = 0.$$

□

2.7 Conclusion

For the critical case ($\mu = \beta + \kappa$), the existence of the inverse Fourier transform $\psi(\infty, x, y)$, shown in section 2.4, means that the correlation function,

$$k^{(2)}(\infty, x, y) = \rho_0^2 + \rho_0 \int\limits_{T^2} \frac{\mu + \lambda\Phi_a(\varphi)}{\lambda\Phi_a(\varphi) + \kappa\Phi_q(\varphi)} e^{-i(\varphi\cdot(y-x))} d\varphi,$$

is finite. This means that for the critical case in two dimensions the demographic process does not lead to clusterization; it will not produce patches.

In more detail, if $\Phi_a(\varphi) \sim |\varphi|^\alpha$, $0 < \alpha < 2$, then the key underlying process is migration with long jumps, which is transient and, therefore, does not produce patches. If $\Phi_q(\varphi) \sim |\varphi|^\alpha$, $0 < \alpha < 2$, then the key underlying process is immigration with long-distance effects. This process cannot be transient in the usual sense but it is analogous to a transient random walk in that the probability of infinite occurrences of immigration at a particular site is 0. It, too, does not produce patches.

Thus, under the conditions of heavy tails migration, immigration influenced by particles at distant locations, or both, the population will reach a stable distribution without clusterization. The distribution that results will depend on the initial density of the population, ρ_0. In particular, as ρ_0 increases we have:

$$k^{(2)}(\infty, x, y) \xrightarrow[t\to\infty]{} \rho_0^2.$$

Finally, on a more general level, we note that a common element of the models of both social processes is the effect of a small proportion of exceptionally active particles. The dynamics of the second model of the spread of rumors are governed by the presence of a minority of spreaders, particles that are the only ones that can spread a rumor. Clusterization in the demographic model is avoided if there are particles that make long jumps when they migrate to a new location, or if immigrants are positively affected by particles at a long distance, one interpretation of which is that some exceptional particles are capable of having long-distance effects.

REFERENCES

Ash, T. G. (1993): The Magic Lantern: The Revolutions of 1989 as Witnessed in Warsaw, Budapest, Berlin, and Prague. New York: Vintage Books

Birkner, M. (2003): Particle Systems with Locally Dependent Branching: Long-Time Behaviour, Genealogy and Critical Parameters. Ph. D thesis, Johann Wolfgang Goethe-Universitat

Daley, D. J. and Kendall, D. G. (1965): Stochastic rumours. Journal of the Institute of Mathematics and its Applications, 1, 42–55

Fayard, J., Klein, E. K. and Lefevre, F. (2007): Long distance dispersal and the fate of a gene from the colonization front. Paper presented at Dispersal Conference, Paris

Feller, W. (1971): An introduction to probability theory and its applications, Vol. II, Second edition. New York: Wiley

Foster, J. (1971): A Limit Theorem for a Branching Process with State Dependent Immigration. Annals of Mathematics and Statistics, 42, 1773–1776

Galton, F. (1 April 1873): Problem 4001. Educational Times

Galton, F. and Watson, H. (1874): On the probability of the extinction of families. Journal of the Royal Anthropological Institute, 4, 138–144

Gikhman, I. and Skorokhod, A. (1974): The Theory of stochastic processes. Berlin: Springer-Verlag

Goodwin, A. (1966): The French Revolution. London: Hutchinson

Greenwood, M. J. and Hunt, G. L. (2003): The early history of migration research. International Regional Science Review, 26, 3–37

Haldane, J. B. S. (1927): A mathematical theory of natural and artificial selection. V. Proc. Cambridge Phil. Soc. 23, 838–844

Harris, T. (1963): The Theory of branching processes. Berlin: Springer

Hawkins, D. and Ulam, S. (1944): Theory of multiplicative processes. Los Alamos National Laboratory (265). – Declassified document

Ivanov, B. G. (1980): The Branching Diffusion with Immigration. Journal of Applied Probability, 17, 1–15

Jensen, H. J. (1998): Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems. Cambridge, UK: Cambridge University Press

Kawazu, K. and Watanabe, S. (1971): Branching Processes with Immigration and Related Limit Theorems. Theory of Probability and Its Applications, 16, 36–54

Kendall, D. G. (1966.): Branching processes since 1873. J. London Math. Soc. 41, 385–406

Kolmogorov, A. N. and Dmitriev, N. A. (1947): Branching stochastic processes. Doklady Akad. Nauk U.S.S.R. 56, 5–8

Lee, E. S. (2006): A Theory of migration. Demography, 3, 47–57

Lefèvre, C. and Picard, P. (1994): Distribution of the final extent of a rumour process. Journal of Applied Probability, 31, 244–249

Li, Z. (2006): A Limit Theorem for Discrete Galton-Watson Branching Processes with Immigration. Journal of Applied Probability, 43, 289–295

Lotka, A. J. (1931): The extinction of families. I. J. Washington Acad. Sci. 21, 377–380

Lowry, I. (1966.): Migration and metropolitan growth: two analytical models. San Francisco: Chandler

Maki, D. P. and Thompson, M. (1973.): Mathematical Models and Applications. Englewood Cliffs, NJ: Prentice-Hall

Malthus, T. (1826.): An Essay on the Principle of Population. 6th edition. London: John Murray

Milos, P. (2009): Occupation times of subcritical branching immigration systems with Markov motions. Stochastic Processes and their Applications, 119, 3211–3237

Pakes, A. (1971a): A Branching Process with a State-Dependent Immigration Component. Advances in Applied Probability, 3, 301–314

Pakes, A. (1971b): On the critical Galton-Watson process with immigration. Journal of the Australian Mathematical Society, 12, 476–482

Pakes, A. (1986): Some Properties of a Branching Process with Group Immigration and Emigration. Advances in Applied Probability, 18, 628–645

Pearce, C. E. M. (2000): The Exact Solution of the General Stochastic Rumour. Mathematical and Computer Modelling, 31, 289–298

Pittel, B. (1990): On a Daley-Kendall model of random rumours. Advances in Applied Probability, 27, 14–27

Rapoport, A. and Rebhun, L. J. (1952.): On the mathematical theory of rumour spread. Bulletin of Mathematical Biophysics, 14, 375–383

Ravenstein, E. G. (1885): The Laws of Migration. Journal of the Statistical Society of London, 48, 167–235

Steffenson, J. F. (1930): On Sandsynligheden for at Afkommet uddr. Matem. Tiddskr. B,, 19–23

Sudbury, A. (1985): The proportion of the population never hearing a rumour. Journal of Applied Probability, 22, 443–446

Watson, R. (1988): On the size of a rumour. Stochastic Processes and Applications, 27, 141–149

Whitmeyer, J. M. (2009): The perils of assuming a single equilibrium. Sociological Perspectives, 52, 557–579

Whitmeyer, J. M. and Yeingst, C. (2006): Modeling Coleman's friendly association networks. Social Science Research, 35, 642–667

Yaglom, A. (1947): Certain limit theorems of the theory of branching random processes. Doklady Akad. Nauk USSR, 56, 795–798

Yamazato, M. (1975): Some Results on Continuous Time Branching Processes with State Dependent Immigration. J. Math. Soc. Jpn, 27, 479–496