SCALING, SURVIVAL AND EXTINCTION IN MANY-SPECIES SYSTEMS

by

Margarita Ifti

Fizikan, University of Tirana, 1990
M.S., Iowa State University, 1997

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Department of Physics and Astronomy)

We accept this dissertation as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

December 2003

Abstract

In this thesis we study the behaviour of three- and many-species systems, when the stochastic nature of the interactions within them is taken into consideration. We pay special attention to the appearance of cooperative states in them. We also study a model of development of cooperation, called Continuous Prisoner's Dilemma.

We start with the cyclic ABC model ($A + B \to 2B$, $B + C \to 2C$, $C + A \to 2A$), and its counterpart: the three-component neutral drift model ($A + B \to 2B$ or $2A$, $B + C \to 2C$ or $2B$, $C + A \to 2A$ or $2C$). In the former case, the mean-field approximation exhibits cyclic behaviour with an amplitude determined by the initial condition. When stochastic phenomena are taken into account, the amplitude of the oscillations drifts and eventually one and then two of the species go extinct. The second model remains stationary for all initial conditions in the mean-field approximation, and drifts when stochastic phenomena are considered. We analyzed the distribution of first extinction times of both models by simulations of the master equation, and from the point of view of the Fokker-Planck equation. Survival probability vs. time plots suggest an exponential decay. For the neutral model the extinction rate is inversely proportional to the system size, while the cyclic system exhibits anomalous behaviour for small system sizes. In the large sys-

tem size limit the extinction times for both models will be the same. This result is compatible with the smallest eigenvalue obtained from the numerical solution of the Fokker-Planck equation. We also studied the behaviour of the probability distribution. For evolutionary times, it becomes uniform, which is in agreement with the exponential decay. The exponential behaviour is found to be robust against certain changes, such as the three reactions having different rates. Finally, we study the model with biased drift, and show that for evolutionary times it behaves like the cyclic (deterministic) and neutral drift models.

When mutation or migrations are taken into consideration, there are three distinct regimes in the model. (*i*) In the "fixation" regime, the first extinction time scales with system size N and has exponential distribution with an exponent that depends on the mutation/migration probability μ. (*ii*) In the "diversity" regime, the system accepts the linear noise approximation, and exhibits Ornstein-Uhlenbeck process behaviour. All three species are present in the system at almost all times. The analytical results are checked against computer simulations of the master equation. (*iii*) In the critical regime, the first extinction time has a power-law distribution with exponent -1. The transition corresponds to a crossover from diffusive behaviour to Gaussian fluctuations around a stable solution. The critical μ is written as μ_0/N and μ_0 is determined from the simulations, as well as numerical solution of a nonlinear Fokker-Planck equation to be about 0.33. The above-described behaviour is observed for systems with four- or more species, and a possible mechanism for the appearance of symbiotic pairs is discussed.

The model is used for emulating a computer network with e-mail viruses. It is observed that clustering is necessary for pandemics to happen. We discuss differences between computer users, e-mail, and social network topologies and

their role in determining the nature of epidemics.

The evolution of cooperation is studied, using the Continuous Prisoner's Dilemma. This is a Prisoner's Dilemma in which the cooperation is not all or nothing, but rather goes through baby steps, from small to bigger degrees of assistance. It has been shown that in the presence of spatial structure the individual's investment reaches substantial levels, and fluctuates around those. We examine the effects of increasing the neighbourhood size, and study the limits at which the mean-field behaviour of pure defection becomes the prevailing one. It is observed that when the neighbourhood sizes for "playing against" and "comparing with" differ by more than 0.5, the cooperative behaviour is unsustainable. Further, neighbourhood size mutations are introduced, and the parameters are treated as phenotypes. The final state is a cooperative one, and the neighbourhood and dispersal parameters values are observed to converge toward one-another. When placed in a network, it is observed that the average degree of the network for which cooperation is still sustained practically does not depend on network size, and is much larger in the clustered social networks than in the distributed random networks. This further strengthens the argument that clustered spatial structures help the development and maintenance of cooperation.

Contents

List of Figures

Acknowledgements

I would first like to thank my supervisor, Dr. Birger Bergersen, for his help and financial support over the last many years. His patience, encouragement, and remarkable ability to make me look carefully at the "trees", while not losing sight of the "forest", have been an inspiration to me. I shall always try to do the same.

I would also like to thank my mentors over the years, Dr. Rexhep Meidani in Tirana and Dr. Constantine Stassis in Ames, for their encouragement and confidence in me.

Part of this thesis work (the chapter on the Continuous Prisoner's Dilemma) was done in collaboration with Dr. Michael Doebeli and Dr. Timothy Killingback.

I am also particularly appreciative of the many helpful discussions with Dr. Hendrik Blok.

I gratefully acknowledge the financial support, through the University Graduate Fellowship, of the University of British Columbia.

Finally, I am greatly indebted to my husband and friend, Paul, for continuously encouraging me in my studies and never hesitating to put his own work aside to help me with code writing and debugging, as well as being my "walking dictionary" when I needed help with reading the part of the

literature that was originally in Russian.

MARGARITA IFTI

The University of British Columbia

December 2003

Dedication

To my parents, Klara and Sarandi.

Se pas dimrit vjen një verë
që do kthehemi njëherë,
Pranë vatrës, pranë punës,
Anës Vjosës, anës Bunës.

Chapter 1

Introduction

There exist many examples of ensembles which influence one-another through competition, such as: populations of various biological species, political parties, businesses, countries, coupled chemical reactions, epidemiological systems, components of nervous systems, elementary excitations in matter, etc. Models for these ensembles have been constructed, either from first principles or intuitively. They yield rate equations, which in general are nonlinear, and contain a number of constants, generally unknown ones.

1.1 Infectious Diseases

In the 1960s, thanks to the effectiveness of improved sanitation, antibiotics, and vaccination programmes, humanity started being confident that infectious diseases would soon cease to exist. But they have continued to be the major causes of death in developing countries. Moreover, infectious disease agents adapt and evolve, so that new infectious diseases have emerged and some existing diseases have reemerged. Here we could mention newly identified diseases such as Lyme disease, Legionnaire's disease, toxic-shock syndrome, hepatitis C and E, hantavirus, and AIDS (HIV virus). Recently, SARS caused huge problems to public health systems, especially in China. Antibiotic-resistant strains of tuberculosis, pneumonia, and gonorrhea have evolved. Malaria, dengue, and yellow fever have reemerged and are spreading as climate changes occur. Cholera, plague, and hemorrhagic fevers continue to erupt occasionally. Infectious agents called prions have joined the already rich zoo of previously known ones: viruses, bacteria, protozoa, and worms. Prions are believed to be responsible for spongiform encephalopathies, e.g. BSE (mad cow), CJD, kuru, and scrapie in sheep.

The emerging and reemerging diseases have led to a revived interest in infectious diseases. Models, including stochastic ones, have become important tools in analyzing their spread and control. However, the mathematical epidemiologists mostly use the rate equation approach, which describes the evolution of the average values of the quantities of interest, and hence does not take into account the fluctuating nature of those quantities. Also, cyclic phenomena are often ignored when studying epidemiological processes, but nevertheless they can have important consequences. We contract most infective diseases only once in our lifetime, because our immune system has "memory". Vaccines are designed based on this knowledge, and they work quite efficiently. However, some diseases do reinfect us. Examples of these diseases would include: Pertussis (whooping cough), influenza, varicella (chicken pox) etc. Immunity to whooping cough is temporary, and decreases as the time after the most recent pertussis infection increases. Influenza comes in several types. Type A influenza has three subtypes in humans that are associated with epidemics and pandemics. Types B and C tend to be associated with local or regional epidemics. An infection or vaccination for one variant may give only partial immunity to another variant of the same subtype. The frequent changes at antigenic sites of the A subtype implies that vaccination programs cannot eradicate them, because the target is constantly moving. The varicella zoster virus (VZV) is the agent for varicella, also known as chickenpox. After the illness, the VZV becomes latent, but can reactivate in the form of shingles. There exists a varicella vaccine, but the vaccine–immunity wanes with time. Our model is motivated by this type of situation.

In epidemiological models populations are often categorized in three states: susceptible (S), infected (I) and recovered (R). There exists a vast literature about the so-called SIR models, where the "loss of immunity" step is not considered (e.g. the classic texts by Bailey [1], Anderson and May [2], the review by Hethcote [3]). We consider the case when the "loss of immunity" step is present, otherwise referred to as $SIRS$ models [4, 5].

1.2 Ecological Systems

It has been shown that many species of bacteria are able to produce toxic substances that are effective against bacteria of the same species that do not produce a resistance factor against the toxin [6, 7]. These bacteria exist in colonies of three possible types: sensitive (S), killer (K) and resistant (R). The killer type produces the toxin, and the resistance factor that protects it from its own toxin, at a metabolic cost. The resistant strain only produces the resistance factor (at a smaller metabolic cost). The sensitive strain produces neither. Clearly, the S colony can be invaded by a K colony, the K can be invaded by R, and the R can be invaded by S.

In our first model the population is categorized into three "species": A, B, C, and the rules of evolution are such that when an A meets a B, it becomes B, when B meets C, it becomes C, when C meets A, it becomes A. The total population size is conserved. In another context, this could be seen as a three-party voter model, when the follower of a certain party "converts" a follower of another certain party when they meet.

The reaction is similar to the cyclical "Rock-paper-scissors game", of which a biological example has been found recently [8, 9]. It is "played" by the males of a lizard species that exist in three versions: the blue throat male defends a territory that contains one female, the orange throat male defends a territory with many females, and the male with a throat with yellow stripes does not defend his own territory, but can sneak into the territory of the orange males and mate with their females. Hence, a "blue" population can be invaded by "orange" males, while an "orange" population is vulnerable to "yellow" males, who on their turn are at an disadvantage against the "blue" males, who defend their territory very well. Several spatial models that relate to such systems have been built [10–12, 107].

1.3 Evolutionary Theories

Darwin in his *Origin of Species* spoke of evolution, as well as proposing a mechanism for it, as "survival of fittest". It implied the prevalence of deterministic dynamics in the evolutionary process. The work of Wright [13] and Fisher [16] marked the beginning of a quantitative theory of the role of genetic drift in the evolutionary process outlined by Darwin. Wright argued that random processes were essential to every aspect of evolution, while Fisher took the opposite view that evolution worked by mass selection alone. In the 1960s Kimura [17] first combined population genetics theory with molecular evolution data [18] to arrive at a theory using genetic drift as the main force changing allele frequencies, known as the "Neutral Theory of Molecular Evolution" [19, 20]. Realizing that "one nucleotide pair has been substituted in the population every two years", he proposed that many of the substituted alleles must be neutral. In 1969, King and Jukes published "Non-Darwinian Evolution" [21], a paper that independently proposed that most aminoacid substitutions must be neutral. Advantageous mutations were thought to be so rare as to make only a negligible contribution to the totality of substitutions. A mutation whose selection coefficient is much less than the inverse of the population size behaves as if it were strictly neutral. The nearly neutral mutation theory deals with mutations whose selection coefficients are close to the inverse of the population size [22].

The existing models consider two alleles of the same locus, and treat the problem stochastically. The cyclic phenomena are usually ignored when studying the evolutionary processes. In order to compare the evolutionary path due to purely neutral drift to that due to cyclic competition, we would need a minimum model with three alleles present. Our second model is a three-component version of the famous Kimura model of neutral genetic drift [19,20]. In that case, when individuals from two species meet, their offspring may be either of the first or the second species, with equal probability.

In the Wright-Fisher model, each individual of the population is equally likely to interact with any other. This raises the question of the influence of spatial structures in the evolutionary process. Wright argued for the necessity of such spatial structures in order to facilitate "hopping over the potential barrier". Fisher downplayed the role of small population size and genetic drift in favour of mass selection. Debate over these issues continues to rage and the question is far from resolved [23–25]. Some of the questions concerning spatial structure were answered by Kimura and Weiss with their analysis of the

Stepping Stone model [26, 27]. Duty in his doctoral work [28] obtains results that highlight the importance of local interactions and spatial structure on large scale gene flow in distributed populations. In recent years the science of networks has made enormous progress [29].

1.4 This Work

This thesis will focus on idealized representation or modelling of the cyclic and neutral version of the three-, four-, or n-species model. We use the Fokker-Planck equation, as well as computer simulations to study the non-spatial model with and without mutations [30, 31], and "experiment" via computer simulation to study different versions of our model.

Computer simulation is necessary because many-agent models are impractical or even impossible to analyze by hand. Even the non-spatial models are too complex for an analytic treatment, and we need to use numerical methods for solving the nonlinear Fokker-Planck equations obtained. So we turn to computers, which are capable of performing millions of calculations rapidly, with no (significant) errors.

Simulation may seem inappropriate because, to develop a stochastic model, random events must be incorporated, but computers are incapable of generating truly random numbers. As an alternative, a number of algorithms have been constructed to produce pseudo-random numbers that pass all known statistical tests for randomness [32–34]. However, these generators still require a seed from the user—a random, initial number to begin the sequence. This flaw can often be a blessing because it offers replicability in our experiments—by seeding the simulation with the exact same number as a previous iteration, that particular realization of our system can be reproduced. (To generate independent realizations different seeds are used.)

The theory of stochastic processes is reviewed briefly in Chapter 2. In Chapter 3 we formally introduce the models, and use the Kramers-Moyal expansion to derive nonlinear Fokker-Planck equations for both. Also, there we report the results of our computer simulations for the non-spatial versions of both models. Chapter 3 also uses the "experimental" data from the simulations to derive the probability distribution function, and deals with the numerical solution of the Fokker-Planck equation. Chapter 4 adds mutations and/or migrations to the picture, considering that in the real world populations are organized in "patches" with a certain number of individuals, who then communicate and exchange individuals with one-another. One of these patches is studied by using the linear noise approximation of the master equation, finding the limits at which the approximation breaks down, and studying the phase transitions that occur. The critical values of the parameters are then calculated directly, using a nonlinear Fokker-Planck equation, obtained from the Kramers-Moyal expansion of the master equation. In Chapter 5 the model is generalized by considering four or more species, and showing that the picture obtained for three species still holds qualitatively. One possible mechanism for the appearance of symbiotic pairs is also discussed. Also, it is shown that the same phase transition will occur in generalized Lotka-Volterra systems. In Chapter 6 we put the cyclic model in a network, emulating the e-mail virus spread and survival in computer users networks, as well as the internet, providing also a comparison to the spreading of infectious diseases in human (social) networks. Chapter 7 deals with a slightly different issue: the evolution of cooperation, using a Continuous Prisoner's Dilemma model. The behaviour of the model is studied in different topologies, such as a 2D square lattice, as well as in networks. It is shown that clustering is an important factor that facilitates the development and maintenance of cooperation. The thesis closes with a discussion of some conclusions that can be drawn from the research and some ideas for future research.

Chapter 2

Stochastic Processes

This chapter is an overview of stochastic processes, and serves as an introduction to the terminology and notation used in this thesis. In it we discuss the classic approaches to working with stochastic processes: the master equation, the Fokker-Planck equation, and methods for its solution. Readers with experience in stochastic processes may only need to skim through this chapter, to familiarize themselves with the notation. Those, who are new in the field of stochastic processes should read the entire chapter, and may also need to check some of the references, particularly the books by van Kampen [35], Risken [36] and the recent book by Redner [37], even though the chapter is meant to be self-contained.

2.1 Stochastic Processes in Physics

The role of stochastic processes in physics has been subject of numerous studies. Stochastic processes enter in the physical picture of the world, because many physical phenomena evolve in time in a very complicated way. We are not able to calculate and/or observe them, but we can observe some average features, and they follow simple laws.

The use of probability considerations in physics is partly prompted for by the impossibility of knowing the microscopic state of the system. However, we know that some variables, such as pressure, can be averaged out, while some others, such as the composition of a gas, cannot. This is a property of the system, not a result of our inability to know the microscopic state of the system. Our experience has taught us that some (rapidly varying) microscopic variables are irrelevant and can be averaged out, while the remaining macroscopic variables are relevant, but we do not have a precise criterion of differentiation between them.

In statistical mechanics we follow the ergodic path: instead of averaging over time (which would be quite a difficult task,) we replace our system with a set of suitably chosen ones, and describe the structure of this set by a density function $\rho(x)$ which characterizes the initial distribution of the states of our ensemble. This way we effectively consider x to be a stochastic variable with distribution $\rho(x)$ (to the approximation of a normalization constant). The rest is trivial: choose the appropriate probability distribution, construct the physical quantity of interest, and start the engine, to calculate the expectation value and the moments. At the end, each physical quantity has become a stochastic process, whose average and moments can be related to the observations. Figuring out when and why the ergodic hypothesis works is the main task of equilibrium statistical mechanics.

But it still remains necessary to solve the microscopic equations of motion, in order to find the distribution for the physical quantity of interest, given the distribution of the states in phase space. At this point, assumptions are made, consisting of repeatedly averaging over the rapid irrelevant variables at successive times. This way we are left with the macroscopic evolution equations (for the slowly varying variables) such as the rate equations, the diffusion equations, the Malthus' law. There are deviations from these equations, which we call fluctuations. They are *the stochastic processes* we are concerned with.

In order to calculate the fluctuations exactly, we would have to solve the microscopic equations of motion. So, we have to treat them as stochastic processes. Fortunately for us, they obey simple laws, which can be established with the help of the same assumption of repeated randomness we used to establish the macroscopic laws. This approach is often referred to as *mesoscopic*. Then the basic hypothesis is that of repeated randomness, which manifests itself in the irreversibility of the macroscopic and mesoscopic equations, at a time when the microscopic equations of motion are themselves reversible.

2.1.1 Distribution Functions

The definition of a stochastic variable X requires specification of (i) the set of possible values, and (ii) the probability distribution over this set. If we have a stochastic process Y defined as in the previous paragraph, i.e. a function of another stochastic process X, then the probability density for the Y to take the value y at time t is:

$$P_1(y,t) = \int \delta(y - Y_X(t)) P_X(x) dx \tag{2.1}$$

If we want the probability that Y have the value y_1 at time t_1, and value y_2 at t_2, and so on til y_n, t_n is:

$$\begin{aligned} P_n(y_1,t_1;y_2,t_2;\dots;y_n,t_n) = \\ \int \delta(y_1 - Y_X(t_1))\delta(y_2 - Y_X(t_2)\dots\delta(y_n - Y_X(t_n))P_X(x)dx \end{aligned} \tag{2.2}$$

This defines a hierarchy of probability densities, which are "consistent" when:

(i) $P_n \geq 0$;

(ii) P_n does not change when we swap two pairs x_k, t_k and x_l, t_l;

(iii) $\int P_n(y_1, t_1; \ldots; y_n, t_n) dy_n = P_{n-1}(y_1, t_1; \ldots; y_{n-1}, t_{n-1})$;

(iv) $\int P_1(y_1, t_1) dy_1 = 1$.

Kolmogorov has proved that any set of functions obeying the "consistency conditions" determines a stochastic process [38].

The *conditional probability* $P_{2|1}(y_2, t_2|y_1, t_1)$ is the probability density for Y to take the value y_2 at t_2 if it had the value y_1 at t_1. It is non-negative and normalized:

$$\int P_{2|1}(y_2, t_2|y_1, t_1) dy_2 = 1 \tag{2.3}$$

By definition the $P_{l|k}$ is symmetric in the set of k pairs of variables.

A process is called *stationary* when all P_n depend on the time differences only, and it is called *Gaussian* when all its P_n are Gaussian distributions. A Gaussian process is then fully specified by its average $< Y(t) >$ and its second moment $< Y(t_1)Y(t_2) >$. Gaussian processes are very easy to handle, and have been studied very well. They often are used as an approximate description for physical processes (equivalent to the assumption that the higher cummulants are negligible). This assumption can be justified in many cases, but not always.

2.1.2 Markov Processes

A Markov process is by definition a stochastic process with the property that for any set of n successive times the following condition is fulfilled:

$$P_{n|n-1,n-2,\ldots,1}(y_n, t_n|y_{n-1}, t_{n-1}; \ldots; y_1, t_1) = P_{n|n-1}(y_n, t_n|y_{n-1}, t_{n-1}) \tag{2.4}$$

This condition means that the conditional probability density at time t_n given the value y_{n-1} at time t_{n-1} does not depend on the values of y at earlier times, i.e. the history of the system. $P_{n|n-1}$ is called the *transition probability*.

The Markov property guarantees then that the Markov process be fully defined by the two functions $P_1(y_1, t_1)$ and the conditional probability $P_{n|n-1}$. We can then reconstruct the whole hierarchy from the above two functions. This property singles out Markov processes by making them manageable, and that is another reason they are liked so much and used in many applications.

The best known example of a Markov process is *Brownian motion*, which is the observed motion of a "heavy" particle immersed in a fluid of "light" molecules, which collide with it randomly, e.g. a dust particle in air. The velocity of the Brownian particle varies by a large number of very small and uncorrelated jumps. The velocity of the Brownian particle is then a Markov process. Einstein and Smoluchowski realised that we actually observe the net displacement after many jumps in velocity. Hence, not only the velocity is a Markov process, but also the position of the particle in the timescale imposed by the experimental conditions is a Markov process.

Assume a closed physical system has a quantity $Y(t)$ which may be treated as a Markov process. When the system is in equilibrium, $Y(t)$ is a *stationary Markov process*. In that case P_1 is independent of time, and identical to the already familiar equilibrium distribution of the quantity Y, as given by equilibrium statistical mechanics. Stochastic processes that are both Markovian and stationary are particularly interesting, especially for describing equilibrium fluctuations. For those processes the transition probabilities do not depend on the two times, but only on the time interval, and we introduce a special notation:

$$P_{2|1}(y_2, t_2|y_1, t_1) = T_\tau(y_2|y_1) \tag{2.5}$$

with $\tau = t_2 - t_1$.

2.2 The Master Equation

The master equation offers the most flexible description for stochastic processes. Let us consider a dynamic system where the state (configuration) space is the set $\{C\}$. In this formalism C can represent almost any kind of variable: a configuration of spins on a lattice, a set of population numbers for a system $(n_1, n_2, \ldots, n_m)$, a set of binary numbers, and so on. Changes in this system happen by stochastic transitions from one state to another; those are Markov processes, and the transition rates, i.e. the probability of the transition per unit time from C to C' are denoted by $W(C|C')$. These rates fully specify the physical model, and the system is described by the probability of its being in state C at time t. The master equation is then a linear equation for the time derivative of the probability $P(C, t)$:

$$\frac{\partial P(C,t)}{\partial t} = \sum_{C'} [W(C|C')P(C',t) - W(C'|C)P(C,t)] \tag{2.6}$$

It expresses the conservation of probability: the first term represents the net flow of probability from any state C' into C, and the second term the flow out of C into any other state C'. It describes the *evolution of the probability distribution*, and not that of the state of the system. The underlying assumption is that the system is intrinsically stochastic. We will be concerned only with stochastic processes that have the Markov property, and furthermore are not explicitly time-dependent.

The following fundamental property of the master equation holds: *In the long-time limit all the solutions tend to the stationary solution(s).* The statement is strictly true only for a finite number of discrete states. When the state space is continuous (e.g. random walk), there are exceptions.

A special case is that of a closed (i.e. no exchange of matter with the external world), isolated (i.e. no external time-dependent forces are present), physical (i.e. there exists a Hamiltonian microscopical description) systems. The energy is then a constant of the motion, and the trajectory in phase space is confined to a single shell—the equilibrium statistical mechanics tells us there is an equilibrium distribution, which is called microcanonical ensemble. If there are other constants of the motion (angular momentum in a cylindrical vessel comes to mind), then the energy shell is split into subshells with a fixed value of the conserved quantity. According to the ergodic hypothesis, the system's motion then covers the whole subshell. When such a system can be described on a mesoscopic level by a master equation, the subshell can be split into cells, in such a way that the evolution of the system can be described by the transition probabilities between two cells n and n', $W_{n,n'}$. These transition probabilities satisfy the following relations:

$$\sum_{n'} W_{n,n'} p_{n'}^{(e)} = p_n^{(e)} \sum_{n'} W_{n',n} \tag{2.7}$$

where $p_n^{(e)}$ is the equilibrium probability for the state n. The transitions balance for each pair n, n' separately, if both the Hamiltonian, and the state variable n are even functions of the momenta. This is known as *detailed balance.* Since the probabilities $p_n^{(e)}$ are known from equilibrium statistical mechanics, the above is a property of the transition probabilities. The cyclic systems of this thesis violate detailed balance (processes have a preferred direction). Even when we do not allow the agents to migrate in and out, the system cannot be considered isolated–there must be some form of transfer of energy, which keeps the system out of equilibrium.

An example of a master equation would be that describing radioactive decay. The state space is the positive integers, and the master equation is written:

$$\frac{\partial P_n(t)}{\partial t} = \lambda[(n+1)P_{n+1}(t) - nP_n(t))] \tag{2.8}$$

It has one single parameter λ, the decay rate. Radioactive decay is a special case of *one-step* or *birth and death* processes. These stochastic processes can be modelled as taking place one event at a time, so that the transition rates have the form:

$$W(n|n') = \lambda(n')\delta_{n,n'-1} + \gamma(n')\delta_{n,n'+1} \tag{2.9}$$

where λ is the death rate and γ the birth rate. When n is restricted to positive integers, we usually have what is called a "natural boundary" at $n = 0$,

$$\lambda(0) = 0 \tag{2.10}$$

In this case the steady state is easy to find, because, if the time derivative of the probability is zero for all n, we must have:

$$\lambda(n)P_n = \gamma(n-1)P_{n-1} \tag{2.11}$$

We can therefore express P_n in terms of P_0 in the following way:

$$P_n = \frac{\gamma(n-1)\gamma(n-2)\ldots\gamma(0)}{\lambda(n)\lambda(n-1)\ldots\lambda(1)}P_0 \tag{2.12}$$

and then determine P_0 from the normalization condition:

$$\sum_n P_n = 1 \tag{2.13}$$

2.3 One-Step Processes

The "one-step processes" are defined as a continuous time Markov process whose range consists of integers n and where only transitions between adjacent sites are allowed. The master equation is of the form:

$$\dot{p_n} = r_{n+1}p_{n+1} + g_{n-1}p_{n-1} - (r_n + g_n)p_n \tag{2.14}$$

where r_n is the probability per unit time of a "jump to the left"' i.e. n decreases by 1, and g_n is the probability per unit time of a "jump to the right", i.e. n increases by 1. Examples of one-step processes are, except for the already introduced radioactive decay, the absorption or emission of particles or photons, generation and recombination processes in semiconductors, birth and death of individuals in a population (hence the name "birth and death processes"), etc. The range for n can be all the integers, the positive integers, or a finite interval of integers.

Based on the coefficients r_n and g_n, one-step processes can be classified as "random walks" when the coefficients are constant and independent of n, "linear one-step processes" when the coefficients are linear functions of n, and "nonlinear" in all the other cases. A special case is the so-called "Poisson process", which is a random walk with jumps to the right alone, with a constant jump rate, and jumps that happen at random times.

Birth and death processes are usually solved by the use of the *generating function*:

$$F(z,t) = \sum_{n=0}^{\infty} P_n(t) z^n \tag{2.15}$$

From the generating function, we can find the moments by taking derivatives with respect to z and then setting $z = 1$. For example, the expectation value for n is

$$< n >= \frac{\partial F(z,t)}{\partial z}|_{z=1} \tag{2.16}$$

Often the m^{th} moment

$$\phi_m \equiv < n(n-1)(n-2)\ldots(n-m+1) >= \frac{\partial^m F(z,t)}{\partial z^m}|_{z=1} \tag{2.17}$$

is referred to as the *factorial moment.* From the factorial moments we can always get back the probabilities:

$$P_n = \frac{1}{n!} \sum_{m=n}^{\infty} \frac{(-1)^m}{m!} \phi_m \tag{2.18}$$

To illustrate the use of the generating function, we take our example of radioactive decay above and multiply both sides of the master equation by z^n. Then we sum over n to produce the equation:

$$\frac{\partial}{\partial t} F(z,t) = -\lambda(z-1) \frac{\partial}{\partial z} F(z,t) \tag{2.19}$$

Suppose we start with a Poisson distribution of particles:

$$P_n(0) = \frac{n_0^n e^{-n_0}}{n!} \tag{2.20}$$

and hence the generating function at the start is:

$$F(z,0) = e^{n_0(z-1)} \tag{2.21}$$

Now the solution to Eq. (2.19) will be:

$$F(z,t) = e^{n_0 e^{-\lambda t}(z-1)} \tag{2.22}$$

which brings us to the expectation of n:

$$< n(t) >= n_0 e^{-\lambda t} \tag{2.23}$$

and the standard deviation:

$$\sigma \equiv \sqrt{< n^2(t) > - < n(t) >^2} = n_0 e^{-\lambda t} \tag{2.24}$$

Often use is made of the "step operator" ϵ, defined as follows:

$$\epsilon f(n) = f(n+1), \qquad \epsilon^{-1} f(n) = f(n-1) \tag{2.25}$$

The master equation for the birth and death processes can be written:

$$\dot{p}_n = (\epsilon - 1) r_n p_n + (\epsilon^{-1} - 1) g_n p_n \tag{2.26}$$

In many cases the physical situation imposes boundaries on the range of values of n; in that case we have to treat the boundaries separately, and write separate equations for the boundaries (assumed to be $n = 0$ and $n = N$)

$$\dot{p}_0 = r_1 p_1 - g_0 p_0 \tag{2.27}$$

$$\dot{p}_N = g_{N-1} p_{N-1} - r_N p_N \tag{2.28}$$

If $r_0 = 0$ and the general equation is valid down to $n = 1$, the boundary $n = 0$ is called "natural". An example would be radioactive decay, where the number of radioactive nuclei cannot be negative. However, something peculiar might happen at low n and the analytic character of r_n is broken, as happens in diffusion-controlled reactions. A boundary that is not natural is called "artificial". Artificial boundaries are defined as boundaries at which the occupation probability of some sites obeys special equations, and r_n and g_n that apply to the other n do not apply to these special sites. One category is the "reflecting" boundaries, where the total probability is conserved, and the "absorbing" boundaries, where the probability vanishes. A "pure" boundary is one in which only the end site requires a special equation. An absorbing pure boundary is referred to as "totally absorbing".

2.4 First-Passage Processes

Suppose a random walker starts at the origin. The question is: how long will it take this walker to reach the site N for the first time? This first passage time is a random variable, since it is different for different realizations of the walk; as such, it has a probability distribution, there is also a mean first passage time. It also makes sense to talk about a survival probability. Clearly, the first passage problem is identical with the absorbing boundary problem, and it can be formulated for an arbitrary one-step problem. The topic has been discussed in books by Redner [37], Spitzer [39], and also in books on stochastic processes, e.g. van Kampen [35], Risken [36], Gardiner [40].

2.5 The Fokker-Planck Equation

The Fokker-Planck equation is a generalized "diffusion" partial differential equation that comes in two slightly different forms. The most common form describes the time evolution of the probability distribution of a continuous random variable, which we will denote as $W(x,t)$. This equation was first used by Fokker [41] and Planck [42] to describe the Brownian motion of particles. A very thorough reference on applications and methods for solving the Fokker-Planck equation is the book by Risken [36]. Risken also shows how the Fokker-Planck can be regarded as a special form of the master equation for continuous random variables.

2.5.1 The Forward Equation

The more general form of the Fokker-Planck equation is also known as the *forward Kolmogorov equation* and is written not for $W(x,t)$, but in terms of the *conditional probability density* $P(x,t|x_0,t_0)$, which is the probability density for x at time $t \geq t_0$, given the initial condition $W(x,t_0) = \delta(x - x_0)$. At this point it is important to emphasize the distinction between these two forms, because later we will introduce the *backward equation*—another equation for the conditional probability $P(x,t|x_0,t_0)$. Following Risken, we write the forward equation as:

$$\frac{\partial}{\partial t}P(x,t|x_0,t_0) = -\frac{\partial}{\partial x}A(x)P(x,t|x_0,t_0) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x)P(x,t|x_0,t_0) \quad (2.29)$$

or, defining the linear differential operator:

$$\mathcal{L}_{FP}(x) \equiv -\frac{\partial}{\partial x}A(x) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x), \quad (2.30)$$

we can write:

$$\frac{\partial}{\partial x}P(x,t|x_0,t_0) = \mathcal{L}P(x,t|x_0,t_0) \quad (2.31)$$

Often the operator $\mathcal{L}_{FP}$ is referred to as the *Liouville operator* of the Fokker-Planck equation. Multiplying Eq. (2.31) by $W(x_0, t_0)$, integrating over x_0, and recognizing that

$$W(x,t) = \int_{\{x_0\}} P(x,t|x_0,t_0)W(x_0,t_0)dx_0 \tag{2.32}$$

produces the more commonly known Fokker-Planck equation for $W(x,t)$, given an arbitrary initial distribution of x at $t = 0$:

$$\frac{\partial}{\partial t}W(x,t) = \mathcal{L}_{FP}(x)W(x,t) \tag{2.33}$$

The first term of the right hand side of Eq. (2.30), containing $A(x)$, is referred to as the "drift" term. The second term, containing $B(x)$, is called the "diffusion" term. In the simplest diffusion process $B(x)$ is just a constant.

The two terms have a relatively simple interpretation. Suppose we start with the initial state x_0, i.e. $W(x,t_0) = \delta(x - x_0)$. Then $A(x_0)$ describes the instantaneous rate of change of the expectation of $x - x_0$ (i.e. $< x > -x_0$), while $B(x_0)$ gives the instantaneous rate of change of the variance of x (i.e. $< x^2 > -x_0^2$).

The Fokker-Planck equation can be generalized to M variables $x_1, \ldots, x_M = \{x\}$ and has the form:

$$\frac{\partial}{\partial t}W(\{x\},t) = \left(-\sum_{i=1}^{M} \frac{\partial}{\partial x_i}A_i(\{x\}) + \frac{1}{2}\sum_{i,j=1}^{M} \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}B_{i,j}(\{x\}) \right) W(\{x\},t) \tag{2.34}$$

Generally, the drift vector $A_i(\{x\})$ and the diffusion tensor $B_{ij}(\{x\})$ depend on all the variables $\{x\}$.

2.5.2 Solving the One-variable Fokker-Planck Equation

The Fokker-Planck equation is linear in $W(\{x\}, t)$; however, two classes of Fokker-Planck equations are discussed in literature. They are called "linear" and "nonlinear". This choice of terminology is somewhat confusing, since the two categories refer to different aspects of the Fokker-Planck equation. This distinction is important to make, so it is included here.

A *linear* Fokker-Planck equation of the form of Eq. (2.34) is one where the drift vector $A_i(\{x\})$ is linear in the variables $x_1, \ldots, x_M = \{x\}$ and the diffusion tensor $B_{ij}(\{x\})$ is constant (i.e. independent of $\{x\}$, but still a tensor). Linear Fokker-Planck equations have Gaussian distributions for the stationary, as well as non-stationary solutions, and this is true even if $A_i(\{x\})$ is explicitly time-dependent. A *nonlinear* Fokker-Planck equation has a diffusion tensor $B_{ij}(\{x\})$ that is explicitly dependent on at least one of the variables of the ensemble $\{x\}$. A more appropriate term for this type of stochastic process is *heterogeneous diffusion.* It can be shown [43] that even for nonlinear Fokker-Planck equations, the probability distribution is Gaussian for short enough time intervals, "short" being determined from the parameter of nonlinearity, as well as the characteristic times of the deterministic system, and the magnitude of the stochastic excitations.

Fokker-Planck equations describing heterogeneous diffusion are difficult to solve, except for the one-variable case, where a change of variables produces a constant diffusion term and a modified drift term. We can think of the heterogeneous diffusion as giving rise to "fluctuation generated forces". For the one-variable Fokker-Planck equation of Eq. (2.33), the steady state distribution $W_s(x)$ can always be found:

$$W_s(x) = \frac{C}{B(x)} e^{2\int^x \frac{A(y)}{B(y)} dy} \tag{2.35}$$

where C is a normalization constant.

2.5.3 The Backward Equation

The backward Kolmogorov equation, or *adjoint equation*, as it is sometimes called in the spirit of quantum mechanics, is also an equation for the conditional probability $P(x, t|x_0, t_0)$. However, the partial derivatives are with respect to the initial time t_0 and initial state x_0. We write it as:

$$\frac{\partial}{\partial t_0} P(x, t|x_0, t_0) = \mathcal{L}^{\dagger}_{FP}(x_0) P(x, t|x_0, t_0) \tag{2.36}$$

and the adjoint operator $\mathcal{L}^{\dagger}_{FP}(x_0)$ corresponding to $\mathcal{L}_{FP}(x)$, defined in Eq. (2.30) is given by:

$$\mathcal{L}^{\dagger}_{FP}(x_0, t_0) = A(x_0)\frac{\partial}{\partial x_0} + \frac{1}{2} B(x_0) \frac{\partial^2}{\partial x_0^2} \tag{2.37}$$

It is obtained by: (i) changing the sign of the drift term, (ii) reversing the order of differentiation and multiplication by $A(x_0)$ and $B(x_0)$, and (iii) changing the partial derivatives with respect to x to those with respect to x_0.

2.6 System Size Expansions: From Master Equation to Fokker-Planck Equation

We need to describe how we can arrive at a Fokker-Planck description for a given stochastic process. We would have to start with a master equation, written in terms of microscopic states and transition rates, and then systematically derive a Fokker-Planck equation for some sort of coarse-grained variables. Along these lines, van Kampen [35, 44, 45, 54] has formulated a general technique for treating the birth-death and one-step processes, which involves an expansion of the master equation in inverse powers of $\Omega^{1/2}$, which is a measure of the system size ($< n > \propto \Omega$ is usually the volume).

The Ω-expansion of van Kampen begins with the Ansatz that the fluctuations in a microscopic variable n are $n - < n > \propto \sqrt{\Omega}$. This Ansatz is quite reminiscent of the Central Limit Theorem for random variables. For a single variable we describe this situation by writing

$$n = \Omega\phi(t) + \sqrt{\Omega}x \tag{2.38}$$

where $\phi(t)$ is the non-fluctuating "macroscopic" variable and x, a fluctuating continuous variable with probability distribution $W(x,t)$. When n fluctuates by an amount Δn, x will fluctuate by $\Delta x = \Delta n/\sqrt{\Omega}$. Then:

$$\frac{\partial W}{\partial x} = \Omega \frac{\partial P}{\partial n} \tag{2.39}$$

and

$$\frac{\partial P}{\partial t} = \frac{1}{\sqrt{\Omega}} \frac{\partial W}{\partial t} - \frac{d\phi}{dt} \frac{\partial W}{\partial x} \tag{2.40}$$

Introduce the raising and lowering operators as follows:

$$\epsilon[f(n)] = f(n+1)$$
$$\epsilon^{-1}[f(n)] = f(n-1)$$

For the special case of one step birth and death process, the master equation can be written:

$$\frac{\partial P(n,t)}{\partial t} = \{(\epsilon - 1)\lambda(n) + (\epsilon^{-1} - 1)\gamma(n)\}P(n,t) \tag{2.41}$$

A formal Taylor expansion of the raising and lowering operators gives:

$$\begin{aligned} \epsilon - 1 &= \frac{1}{\sqrt{\Omega}}\frac{\partial}{\partial x} + \frac{1}{2\Omega}\frac{\partial^2}{\partial x^2} + \ldots \\ \epsilon^{-1} - 1 &= -\frac{1}{\sqrt{\Omega}}\frac{\partial}{\partial x} + \frac{1}{2\Omega}\frac{\partial^2}{\partial x^2} + \ldots \end{aligned}$$

After substituting into the master equation, and sorting terms according to their order in Ω, we obtain a deterministic ordinary differential equation for $\phi(t)$ that is precisely the one obeyed by the expectation $< n(t) > /\Omega$ under the stochastic process, and a Fokker-Planck equation for the fluctuating part $W(x,t)$. Furthermore, the expansion is systematic in that its validity is quite apparent. The appropriateness of the expansion can be briefly summarized by saying that the stability of the differential equation for $\phi(t)$ in the standard dynamical systems sense implies the validity of the expansion for large enough system size. The reader should consult the references [35, 44, 45, 54] to determine what "large enough" means. The difficulty arises when the system does not have a stable steady state, but we have to rather consider fluctuations about a limit cycle or a centre. There are systems that do not have a single steady state solution as such. In these systems, the fluctuations grow unchecked not only along the solution curves, but also perpendicular to them. An important subclass is formed by systems for which the Ω expansion yields a Fokker-Planck equation in which the coefficient(s) in front of the first order derivative (i.e. the drift coefficient) is zero. These are known as master equations of *diffusion type*. In such cases it can be shown [35] that the macroscopic solution is unstable, and the fluctuations are expected to grow. The Ω expansion breaks down after a transient period. The case of phase transitions is another one, when the Ω expansion breaks down, since again the fluctuations are large. The important point to make here is that van Kampen's method fails near the instabilities, or critical points, as they are called in statistical physics. This is unfortunate, since these points in the phase diagram are precisely the most "interesting" ones.

In the case of critical fluctuations, the expansion in powers of $\Omega^{1/2}$ is not the correct starting point any more [35]. The fluctuations now will be proportional to a higher power of Ω. This is the *enhancement of fluctuations near the critical point*. In this case we must take the fluctuations to be proportional to another power of Ω, which according to Kubo et al. [46], is $3/4$. Furthermore, their distribution is not Gaussian any more. The expansion in powers of $\Omega^{-3/4}$ is valid only at *exactly the critical point*. At all the other points, no matter how close to the critical point, the expansion in powers of $\Omega^{-1/2}$ is the appropriate one [35]. However, in the proximity of the critical points that expansion becomes less and less useful. The fact that at the critical point the expansion changes abruptly is called Stokes' phenomenon [47].

So, the question remains: does there exist a Fokker-Planck representation of the master equation which is asymptotic in system size, but retains an accurate description of the fluctuations, even near the critical points? By accurate we mean both qualitatively, i.e. the phase diagram is preserved, and quantitatively, i.e. the critical exponents characterizing the instabilities are unchanged. The answer is yes, as was pointed out by Horsthemke and Brenig [48, 49]. Their work was based on the mathematically rigorous results of Kurtz [50, 51], who proved similar results for the asymptotic representation of stochastic processes by stochastic differential equations. Horsthemke and Brenig's results are simple to state: instead of the change of variables of Eq. (2.38), we make the change of variables:

$$n = \Omega x \tag{2.42}$$

In other words, we transform to the intensive variable $x = n/\Omega$. This method is also recommended by van Kampen [35] when dealing with diffusion type equations. Expansion of the master equation now produces a non-linear Fokker-Planck equation that preserves the character of the fluctuations near unstable critical points. This is what is otherwise known as the Kramers-Moyal expansion, and is essentially a Taylor expansion of the master equation in powers of Ω. It is very important to know how many terms of expansion must be taken into account. The Pawula theorem [52] states that for a positive transition probability, the Kramers-Moyal expansion may stop either after the first, or the second term, but if it does not stop after the second term, it must contain an infinite number of terms. If the expansion is stopped after the second term, it then yields a Fokker-Planck equation [36]. We will illustrate this method in Chapter 3, and further in Chapter 4, when we derive a non-linear diffusion approximation for the ABC model, and study the phase transition in the presence of mutations/migrations.

2.7 Solution of the Fokker-Planck Equation

In general, it is very difficult to obtain analytical solutions for the many-variable Fokker-Planck equation. The difficulty in solving it usually increases with increasing number of variables. However, its similarity to the Schrödinger equation suggests that we can use the methods that are used to solve that equation. At first, therefore, one may try to eliminate some variables. If the drift and diffusion coefficients do not explicitly depend on some variables, those variables can be separated. If some variables decay very rapidly (fast variables) they may be eliminated adiabatically. In special cases the problem can be transformed into a Hermitian one, or the variational method can be used. A very effective way is always the numerical solution, the same way a differential equation is solved.

Often the method of expansion into complete sets is used. This method can be used when we have complete orthonormalized sets of functions, which satisfy the boundary conditions for the variables of the problem. Because the sets are complete, the probability density may be expanded as a series involving these eigenfunctions. By substituting the W expansion back into the Fokker-Planck equation, we obtain an infinite system of coupled differential equations for the expansion coefficients. If the infinite system is truncated at some order, we may solve it numerically, leading to approximate solutions of the Fokker-Planck equation. This is essentially the path we will follow in Chapters 3 and 4 to obtain the smallest eigenvalue of the Fokker-Planck equation for our system.

2.8 Langevin Equations

Many stochastic processes are formulated as stochastic differential equations. Such processes are most easily visualized in terms of ordinary differential equations, to which we add a random perturbation or "noise" term. In the physics literature first order stochastic differential equations are often called *Langevin equations*, due to the work of Langevin [53], who also investigated Brownian motion.

A single-variable Langevin equation is usually written in the physics literature as

$$\frac{dx(t)}{dt} = A(x) + \sqrt{B(x)}\eta(t) \tag{2.43}$$

where $A(x)$ is the deterministic force and $\eta(t)$, a Gaussian distributed random impulse with zero mean and autocorrelation

$$< \eta(t)\eta(t') >= \delta(t - t') \tag{2.44}$$

The main advantage of this formulation is its conceptual simplicity. It is easy to simulate a Langevin equation by first integrating the deterministic part for a time interval Δt, followed by adding the random "kick" of $\sqrt{B(x)}\eta(t)\Delta t$. The disadvantages of using Langevin equations are twofold. First of all, Eq. (2.43) as it is written, is ambiguous. This point is discussed in detail by van Kampen [35, 54]. The problem is that $\eta(t)$ is a stochastic process with discrete jumps. When $B(x)$ is not a constant, one must choose what value of x is used to compute the noise. Should it be the value before, or after the jump? Or perhaps an average of the two? The Ito interpretation [55] is to use the value before the jump

$$x(t + \delta t) = x(t) + A(x)\delta t + \sqrt{B(x)} \int_t^{t+\delta t} \eta(t')dt' \tag{2.45}$$

and is typically appropriate for internal fluctuations, like those of birth-death processes. It can be shown that this interpretation of Eq. (2.43) is equivalent to the Fokker-Planck equation:

$$\frac{\partial}{\partial t}W(x,t) = -\frac{\partial}{\partial x}A(x)W(x,t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}B(x)W(x,t) \tag{2.46}$$

Alternatively, we can use the Stratonovich interpretation [56] and take the average:

$$x(t+\delta t) = x(t) + A(x)\delta t + \sqrt{B(\frac{x(t+\delta t)+x(t)}{2})} \int_t^{t+\delta t} \eta(t')dt' \qquad (2.47)$$

The latter interpretation is appropriate for external noise. For example, a system coupled to an environment where the Gaussian noise is never really "white" (delta-function correlated), but has a non-zero relaxation time. It can be shown that this interpretation leads to the Fokker-Planck equation:

$$\frac{\partial}{\partial x}W(x,t) = -\frac{\partial}{\partial x}A(x)W(x,t) + \frac{1}{2}\frac{\partial}{\partial x}\sqrt{B(x)} \cdot \frac{\partial}{\partial x}\sqrt{B(x)}W(x,t) \qquad (2.48)$$

The second limitation of the Langevin equation formulation is that the form of the noise term $B(x)$ must come from a more detailed knowledge of the stochastic process. As pointed out by van Kampen [35, 54], both of these problems relating to specification of the noise are resolved, if one starts with a microscopic description in terms of the master equation, and proceeds using the system size expansion discussed in the previous section to arrive at a Fokker-Planck equation. One can then use the equivalence described above to get the appropriate Langevin equation and interpretation. This procedure is convenient for simulating a Fokker-Planck equation using the Langevin formulation.

The one-variable Langevin equation described above can be easily generalized to several degrees of freedom by using a system of ordinary differential equations supplemented with appropriate and cross correlators for the noise terms. It can be further extended to continuous systems with infinitely many degrees of freedom.

2.9 Conclusions

In this thesis we consider the master equation to be the governing equation of stochastic processes, and van Kampen's Ω-expansion is the appropriate way of solving the master equation when the system approaches a steady state, and the numbers involved are large, which means the fluctuations are relatively small. The terms of order $\Omega^{1/2}$ yield the macroscopic rate equations, while those of order Ω^0 a linear Fokker-Planck equation with time-dependent coefficients. This equation describes small fluctuations around the macroscopic solution (linear noise approximation). The process described by this linear Fokker-Planck equation is an Ornstein-Uhlenbeck process. A Langevin approach can also be used, but it fails for fluctuations around time-dependent solutions.

In higher orders, additional powers are added to the coefficients of the Fokker-Planck equation, and higher derivatives appear. The stationary distribution is not necessarily Gaussian; however, the moments can be computed successively. The same remains true for a multivariate master equation. Exceptions are the solutions that are unstable (as in the case of phase transitions) or diffusion type master equations. In these cases the fluctuations grow, as to make the expansion valid only for short times. In many cases the Kramers-Moyal expansion can be used, but it leads to a nonlinear Fokker-Planck equation, usually difficult to solve. Therefore, there are no recipes as to how to handle and study the stochastic processes that are valid in all cases.

Chapter 3

Survival and Extinction in Non-Spatial Cyclic and Neutral Three-Species Systems

The equilibrium systems are quite important to study, but most of the systems in the world are not in equilibrium. One of the pioneers in the study of non-equilibrium systems in human society was Vilfredo Pareto. In his work [14,15] he locates six residues as major motivations of human actions, and discusses the reasons and the process in which elites with (mainly) these residues cyclically replace one-another at the helm of power. According to him, "the history of man is the history of the continuous replacement of elites", and "there is a rhythm of sentiment which we can observe in ethics, in religion, and in politics as waves resembling the business cycle" [15].

Biological systems also represent themselves as non-equilibrium systems. R. A. Fisher [16] was the first to mathematically analyse the problem of genetic drift using a partial differential equation known as the Fokker-Planck (or forward Kolmogorov equation) discussed in the previous chapter. In this chapter we expand on our study of cyclic and neutral three-species systems [30]. We use a non-linear Fokker-Planck equation to describe the time-evolution of the probability distribution of three variables, which take values in the [0,1] interval. These variables represent the frequency of some given genes or alleles in a fixed-size population. The Fokker-Planck equation is a continuous-time "diffusion" equation that describes the large system size, long-time limit of our model. The Fokker-Planck equation is most familiar to physicists, since it resembles the Schrödinger equation, and is more closely related to developments in modern statistical mechanics.

3.1 Description of the Model

Consider a system in which three species A, B, C are competing in a way described by the reaction: $A + B \to 2B$, $B + C \to 2C$, $C + A \to 2A$. The model is discrete, both in time and space. It consists of a finite population of N individuals, which are of one of the three types: A, B, or C. Using the genetics terminology, this would be called a "one locus, three alleles" model. The population evolves under well-defined rules as follows: two individuals are picked at random. One dies without reproducing, while the other replicates. This way, not all the individuals contribute to the next time step. The two versions of our model differ in the way it is decided which individual will die and which one will reproduce.

In the first version, which we call the cyclic competition one, the A individuals "are eaten" by the B individuals, who go on and reproduce, when an AB pair is picked, the B individuals "are eaten" by the C individuals, who then reproduce, when a BC pair is picked, and the C individuals "are eaten" by the A individuals, who then reproduce, when an AC pair is picked. This model, which could be thought of as a real-life game of rock-paper-scissors, is important in ecology, since it addresses the issue of maintaining biodiversity (in e.g. food webs). In the context of ecological systems, this is how one would model the predator-prey interaction: the predator eats the prey, and then reproduces (the reproduction process happens at metabolic cost). The model is similar to the epidemiological $SIRS$ model, in which immunity to a certain pathogen is temporary, and a recovered individual eventually goes back to being susceptible. However, the analogy is not complete, particularly since the loss of immunity process does not happen through contact, but rather at a certain rate. As far as evolutionary processes are concerned, the cyclic scenario corresponds to the offspring of two individuals being always of the same type as one of them, and never as the other.

The rate equations for this system will be:

$$\begin{aligned} N\frac{dA}{dt} &= AC - AB \\ N\frac{dB}{dt} &= BA - BC \\ N\frac{dC}{dt} &= CB - CA \end{aligned} \tag{3.1}$$

with $A + B + C = N =$ const. We assume the rates are the same, in which case a time-rescaling will remove them from the equations. These equations can be rewritten as:

$$\begin{aligned} N\frac{d}{dt}\ln A &= C - B \\ N\frac{d}{dt}\ln B &= A - C \end{aligned} \tag{3.2}$$

$$N\frac{d}{dt}\ln C = B - A$$

which leads to the second conservation rule: $ABC = H =$ const.

The above model contrasts with the neutral drift model ($A + B \to 2A$ or $2B$, $B + C \to 2B$ or $2C$, $C + A \to 2C$ or $2A$). In this version of the model, an encounter of an A and a B will result in the death of *either* the A or the B with equal probability, and the reproduction of the remaining one. A similar scenario will follow the encounter of a B and a C, and the encounter of a C and an A. This scenario corresponds to the neutral genetic drift of Kimura [17–20]. The rate equations for this version of the model will be written:

$$\frac{dA}{dt} = \frac{dB}{dt} = \frac{dC}{dt} = 0 \tag{3.3}$$

3.2 The Fokker-Planck Equation for Our Model

In Chapter 2, we presented the Kramers-Moyal expansion, which is an asymptotic expansion of the master equation that produces a non-linear Fokker-Planck equation. We will use that expansion here to derive a non-linear diffusion equation for our model.

The dynamics of the model consists of three types of random events that occur independently in time. This means the stochastic process is Markovian and the time intervals between similar events are exponentially distributed.

In the expansion below, the system size parameter will be the total number of individuals N. It is the same parameter we called Ω in Chapter 2. In general, Ω is not the total number of particles, but the volume. For this model the total number is a conserved quantity, and therefore we can take $\Omega = N$.

Assuming that the system is subject to stochastic noise due to Poisson birth and death processes (intrinsic noise), we get the master equation for the cyclic competition model:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \frac{1}{N}[(A-1)(C+1)P(A-1,B,C+1,t) - ACP(A,B,C,t)+$$
$$+(A+1)(B-1)P(A+1,B-1,C,t) - ABP(A,B,C,t)+$$
$$+(B+1)(C-1)P(A,B+1,C-1,t) - BCP(A,B,C,t)] \tag{3.4}$$

For the neutral drift case the master equation is:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \frac{1}{N}[\frac{1}{2}(A-1)(C+1)P(A-1,B,C+1,t)+$$
$$+\frac{1}{2}(A+1)(C-1)P(A+1,B,C-1,t) - ACP(A,B,C,t)+$$
$$+\frac{1}{2}(A+1)(B-1)P(A+1,B-1,C,t)+\frac{1}{2}(A-1)(B+1)P(A-1,B+1,C,t)-$$
$$-ABP(A,B,C,t) + \frac{1}{2}(B+1)(C-1)P(A,B+1,C-1,t)+$$
$$+\frac{1}{2}(B-1)(C+1)P(A,B-1,C+1,t) - BCP(A,B,C,t)] \tag{3.5}$$

3.2.1 Expansion of the Master Equation

At this point we can carry out the expansion of the master equation, using the classical technique. We will use the system size expansion of Horsthemke and Brenig [48, 49]. However, the notation and style is closer to van Kampen [35, 44].

Now we introduce the "shift" operators, defined by

$$\epsilon_A f(A,B,C) = f(A+1,B,C) \tag{3.6}$$

and

$$\epsilon_A^{-1} f(A,B,C) = f(A-1,B,C) \tag{3.7}$$

and likewise for B and C operators. The master equation for the cyclic competition model now reads:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \frac{1}{N}[(\epsilon_C\epsilon_A^{-1}-1)AC+(\epsilon_A\epsilon_B^{-1}-1)AB+(\epsilon_B\epsilon_C^{-1}-1)BC]P(A,B,C,t) \tag{3.8}$$

and for the neutral drift one:

$$\begin{aligned}\frac{\partial P(A,B,C,t)}{\partial t} = \frac{1}{N}[(\frac{1}{2}(\epsilon_C\epsilon_A^{-1}+\epsilon_A\epsilon_C^{-1})-1)AC+(\frac{1}{2}(\epsilon_A\epsilon_B^{-1}+\epsilon_B\epsilon_A^{-1})-1)AB+ \\ +(\frac{1}{2}(\epsilon_B\epsilon_C^{-1}+\epsilon_C\epsilon_B^{-1})-1)BC]P(A,B,C,t)\end{aligned} \tag{3.9}$$

Next we transform to the intensive quantities

$$x_A = \frac{A}{N}, x_B = \frac{B}{N}, x_C = \frac{C}{N} \tag{3.10}$$

The shift operators become

$$\begin{aligned}\epsilon_{A(B,C)} &= \sum_j \frac{1}{j!}N^{-j}\frac{\partial^j}{\partial x_A^j(x_B^j,x_C^j)} \\ \epsilon_{A(B,C)}^{-1} &= \sum_j \frac{-1^j}{j!}N^{-j}\frac{\partial^j}{\partial x_A^j(x_B^j,x_C^j)}\end{aligned} \tag{3.11}$$

Further we use the rules of transformation of random variables to define:

$$W(x_A,x_B,x_C,t) = P(A,B,C,t)\cdot N^3 \tag{3.12}$$

Before we go ahead with the expansion, a few comments are necessary. From the rate equations, it is clear that our system does not have one single steady state or limit cycle. Instead, the mean-field solution is completely dependent on the initial conditions, and we do not have a macroscopic solution to expand about, but rather an infinity of neutrally stable cycles (for the competition case) or points (for the neutral model). The above expansion, otherwise known as Kramers-Moyal expansion, is discussed by van Kampen [35, 44]. It is a risky expansion, mainly because derivatives of orders higher than two are not small themselves. Also, with time, large fluctuations may occur, and since our system presents itself as a fluctuation-driven one, it becomes even more important to watch out for these sorts of complications.

In the case of neutral drift, the second order term is the leading one, so we would expect this Ansatz to work for large system sizes. In the cyclic competition model, the first order term does not drive the system out of the (neutrally) stable trajectories, and again it would be the second order term which would essentially determine the fate of the system. So, we also set out to investigate the applicability of the Kramers-Moyal expansion to the cyclic competition and neutral genetic drift three-species systems.

Now we transform the master equation into an equation for $W(x, y, z, t)$ grouping terms of the same order in N. The terms of order N^1 cancel, and the first terms in the expansion are of order N^0; keeping only those gives us the first order equation for the cyclic competition case:

$$\frac{\partial W}{\partial t} = [(\frac{\partial}{\partial x_A} - \frac{\partial}{\partial x_C})x_A x_C + (\frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_A})x_B x_A + (\frac{\partial}{\partial x_C} - \frac{\partial}{\partial x_B})x_C x_B]W \quad (3.13)$$

For the neutral drift case the first-order term is identically zero (in agreement with the rate equations).

The second order term is obtained by considering the terms of order N^{-1}:

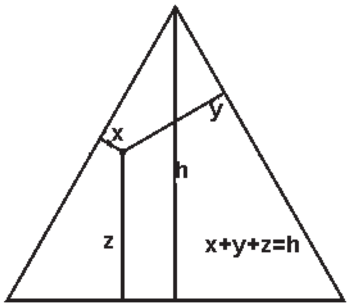

Figure 3.1: In the equilateral triangle the sum of distances of any point from the sides of the triangle is equal to the height of the triangle (unity.)

$$\frac{1}{2N}[(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_A x_C + (\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_B} + \frac{\partial^2}{\partial x_B^2})x_A x_B +$$
$$+(\frac{\partial^2}{\partial x_B^2} - 2\frac{\partial^2}{\partial x_B \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_B x_C]W \tag{3.14}$$

and is identical for both cyclic competition and neutral drift versions.

3.2.2 The First Order Term

The concentrations x_A, x_B, x_C of a three–component system are commonly represented by the distances of a point inside an equilateral triangle of unit height from its sides (Fig. 3.1):

$$x_A + x_B + x_C = 1$$
$$\frac{\partial}{\partial x_C} = -\frac{\partial}{\partial x_A} - \frac{\partial}{\partial x_B}$$

For the neutral drift case the first order equation (with an identically zero right hand side) tells us that in the mean–field approximation the system will remain in the initial state forever.

After some algebra, the first order equation for the cyclic competition system becomes:

$$\frac{\partial W}{\partial t} = \frac{\partial}{\partial x_A}(x_A x_C - x_A x_B)W + \frac{\partial}{\partial x_B}(x_A x_B - x_B x_C)W + \frac{\partial}{\partial x_C}(x_B x_C - x_A x_C)W \tag{3.15}$$

At the centre of the triangle ($x_A = x_B = x_C = 1/3$) all three expressions on the right hand side of the first order equation (3.15) above are zero, so in the first order (mean field) approximation the cyclic competition system will stay at the state with equal concentrations of each component forever. It can be verified that the product $H = x_A x_B x_C$ is a solution of the first-order equation, as is any function of that product. The lines $x_A x_B x_C$ = const. will then represent possible trajectories of the system. These trajectories are closed, since there is no term to drive the system out of them. In this (mean-field) approximation, the cyclic model exhibits global stability with constant system size N [57, 58]. The populations of each species oscillate out of phase with one-another, and the amplitude remains constant [59, 60]. These are the same trajectories that are obtained when one solves the rate equations (3.1). Fig. 3.2 shows some of those closed trajectories.

3.2.3 The Second Order Term

The second order term (3.14), which represents the "diffusion term" must be added to the first order equation to give the Fokker-Planck equation for our system. This term is identical for both the cyclic competition and the neutral genetic drift cases. As we will see, this makes both systems behave essentially the same, at least for reasonably large system sizes.

Van Kampen in his book [35] gives many examples of systems that

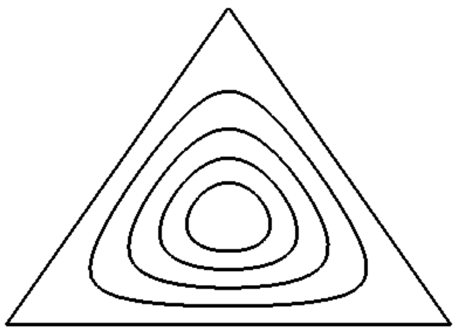

Figure 3.2: Some of the closed trajectories of the system when only the first order term is considered.

approach a steady state; in the competition case the mean field solution does not result in a steady state, but rather cycles with different values of the constant of motion $H = x_A x_B x_C$. For this reason, for the cyclic competition model we are unable to make a system–size expansion of the type envisioned by van Kampen [35, 44], and Horsthemke and Brenig [48, 49].

The trajectories of a given realisation of the system can easily be obtained by simulation: they initially spiral out of the centre, with the populations of each species oscillating out of phase with one–another, so that the total size of the population is conserved. The amplitude drifts until one of the populations becomes zero (species goes extinct). In other words, the slow second order term makes the system cross from a closed (mean-field) trajectory to a neighbouring one. This would suggest an adiabatic approximation [36]. In the neutral drift model, the fluctuations drive the system from one point–solution of the mean-field equations to a neighbouring one.

3.3 Computer Simulations

Before reporting the results of our computer simulations, we will point out the important differences between our simulations for continuous time reaction-diffusion models with those commonly used in statistical physics. One of the major differences involves time.

In the standard Monte-Carlo method, time is discrete and measured in steps. At each step, one tries to change the configuration according to some allowed "rules". The new configuration is then either accepted or discarded, according to probabilities that depend on the difference of the new configuration with the old configuration. For example, when simulating equilibrium systems like the Ising model [61], at each step a site is selected and a spin is flipped. If the new configuration has lower energy ($\Delta E < 0$), it is accepted. If it has higher energy ($\Delta E > 0$)), it is accepted with probability $e^{-\Delta E/k_B T}$. In the present problem there is no analogue to the energy E. Instead we are led to simulate the dynamics of the master equation, which describes a continuous time Markov process, where the "event" times are exponentially distributed with a characteristic time for each event that depends on the configuration.

The computer simulations of both systems start with an equal number of A, B, C, (start at the centre) and generate times for the next possible reaction event with exponential distribution as $-N \cdot \ln(rn)/(AB)$, (for $A + B \to 2B$ reaction, and similarly for the other two reactions,) where rn is a random variable with uniform distribution in [0,1] (this way the event times have exponential distribution, and the events are independent) [62]. The reaction which occurs first is then picked and the system is updated. The process is repeated until one of the species, and then another one, go extinct.

Fig. 3.3 shows the variation with time in the number of A's for a re-

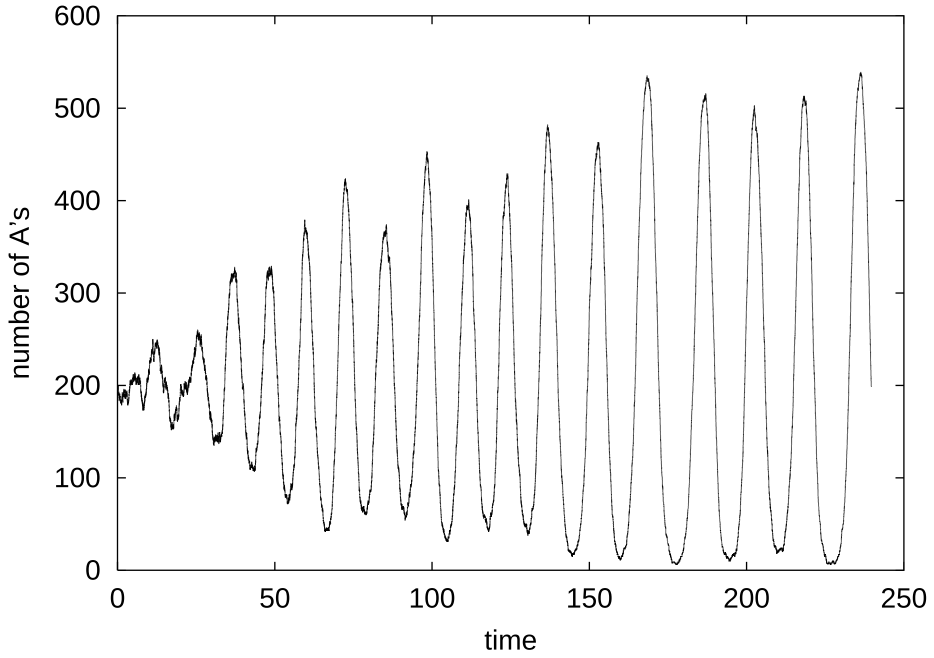

Figure 3.3: Variation over time in the total number of the A species for a realisation of the cyclic competition system ($N = 600$).

alization of the system in the cyclic competition scenario, and Fig. 3.4 shows the evolution of the number of A population in the neutral drift case (in both realizations the total population size is $N = 600$.) From the time series for the population numbers it looks as if the neutral drift model is some sort of "adiabatic" approximation of the cyclic competition model. If in the neutral drift model the population number is random walking, in the cyclic competition one the amplitude of oscillations is random walking! In other words, if we average the cyclic competition model over the cyclic orbits, we get essentially the same behaviour as the neutral genetic drift model. Then the fluctuations drive both models from a neutrally stable cycle (point) to the neighbouring ones, and these fluctuations remain quite small by virtue of the vicinity of the neighbouring cycle (point.) It is quite funny how we all tend to go around in circles in life, even though the point we end up to is still the same!

We investigated the probability distribution of first extinction times. For that we ran fifty thousands copies of the system for each (different) size of the system, and the number of surviving systems was plotted vs. time. In the semilog axes we get a straight line, except for the few "rare events". The slope of this straight line is -3.01 for system size 3000, -2.98 for system size 6000, -2.99 for system size 9000. Fig. 3.5 shows the plots for these three different system sizes: 3 000, 6 000, and 9 000.

The time dependence of survival probability scales approximately with $1/N$, but there is still a very weak dependence left for small N. This can be justified by looking at the Fokker-Planck equation for this system, which contains both a term of order N^0 and N^{-1}.

To further investigate the behaviour of the system, we looked at the cumulative, conditional on being alive (which is equivalent to a normalization condition,) probability distribution for $H = x_A x_B x_C$. If $P(H < h|\text{alive})$ in-

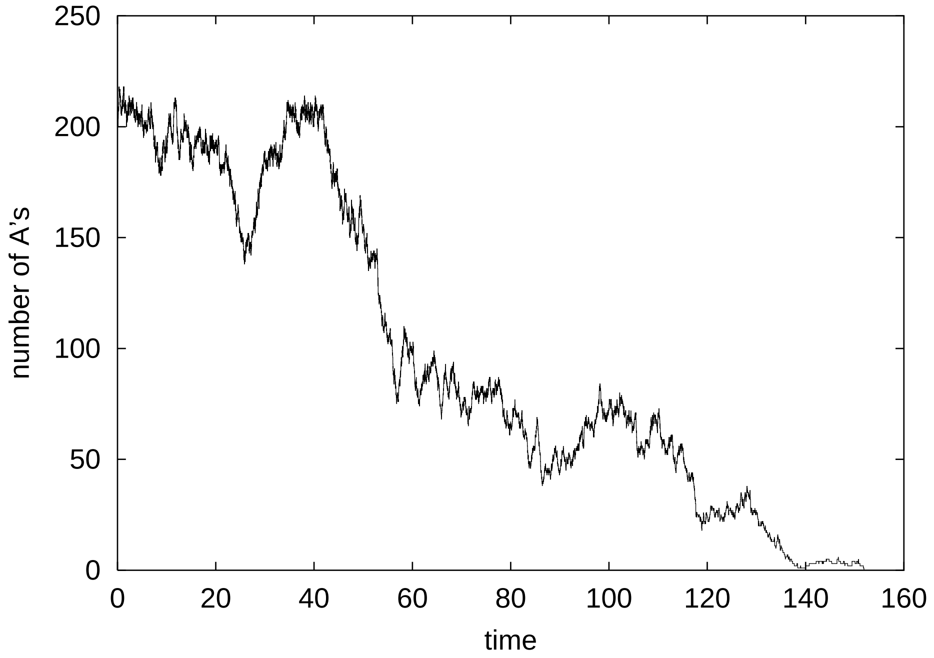

Figure 3.4: Variation over time in the total number of the A species for a realisation of the neutral drift system ($N = 600$).

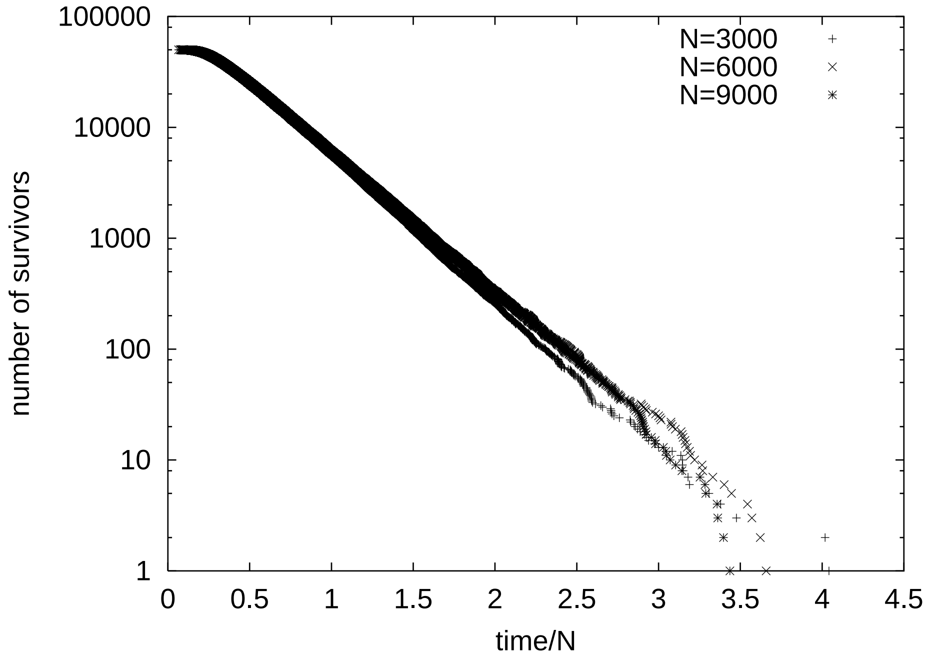

Figure 3.5: Number of surviving systems vs. time plots for three different system sizes, cyclic competition model.

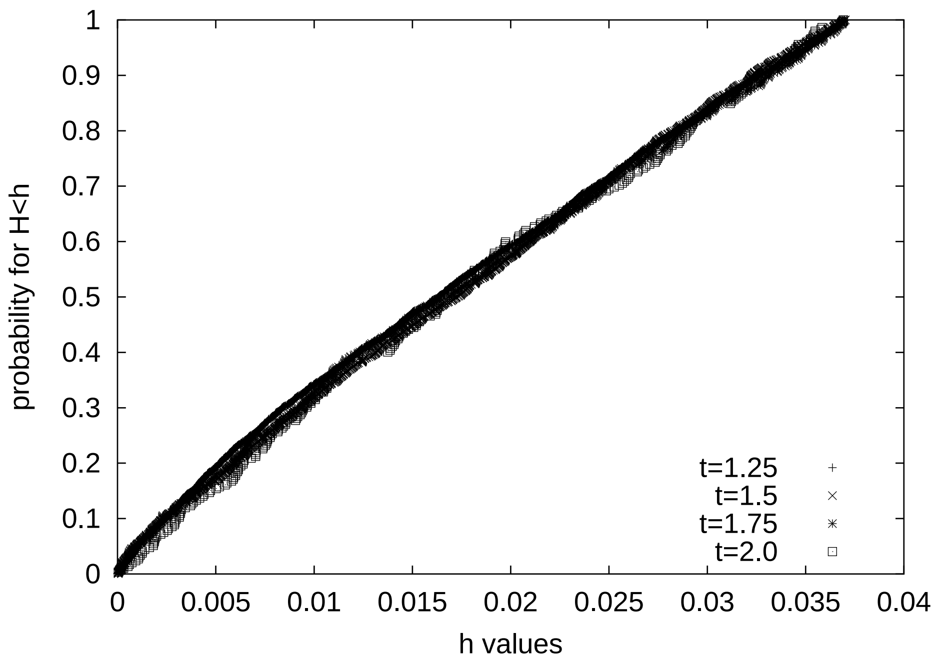

Figure 3.6: Probability distribution for the $H = x_A x_B x_C$ quantity for our system at $t = 1.25$, $t = 1.5$, $t = 1.75$, $t = 2.0$.

creases linearly with H for small H, the extinction rate will be non-zero, and we will get exponential decay. This corresponds to a constant distribution for $P(H|\text{alive})$. Fig. 3.6 shows these plots for $t = 1.25$, $t = 1.5$, $t = 1.75$, and $t = 2.0$ in linear axes. It can be seen clearly that the plot is a straight line, and the slope becomes a bit steeper for small H.

For the neutral drift case the data from the simulations for the survival probability vs. time in semilog axes give us a slope of -2.992 for system size 6000, -2.991 for system size 3000.

These plots are shown in Fig. 3.7. From the plot it can be seen clearly that the data collapse when time is rescaled with $1/N$, and that is in very good agreement with the fact that there is only a term of the order $1/N$ present in the Fokker-Planck equation for the neutral drift system. In Fig. 3.8 we show

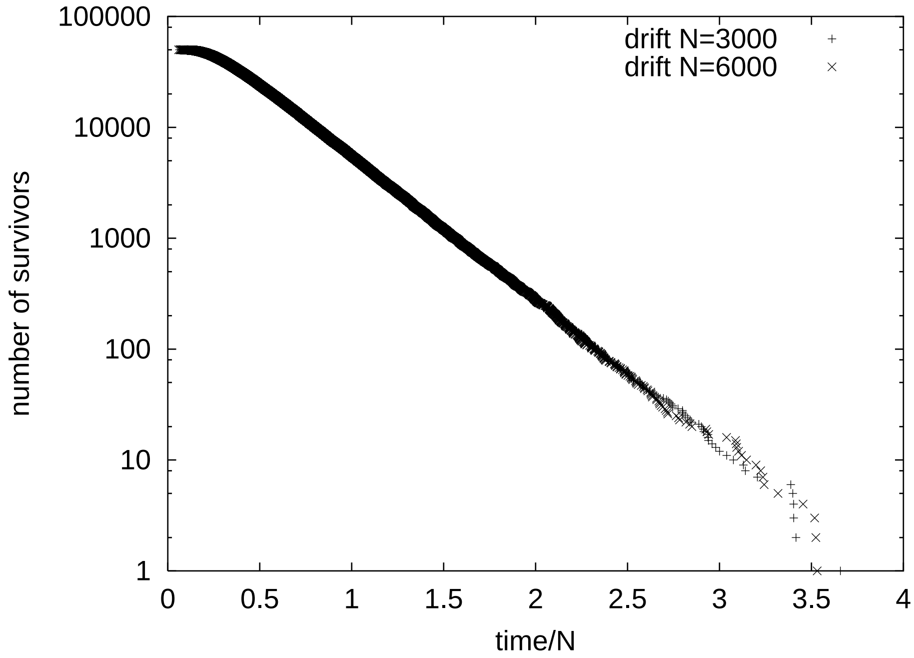

Figure 3.7: Number of survivors vs. time plots for the neutral drift case for system sizes $N = 3000$ and $N = 6000$.

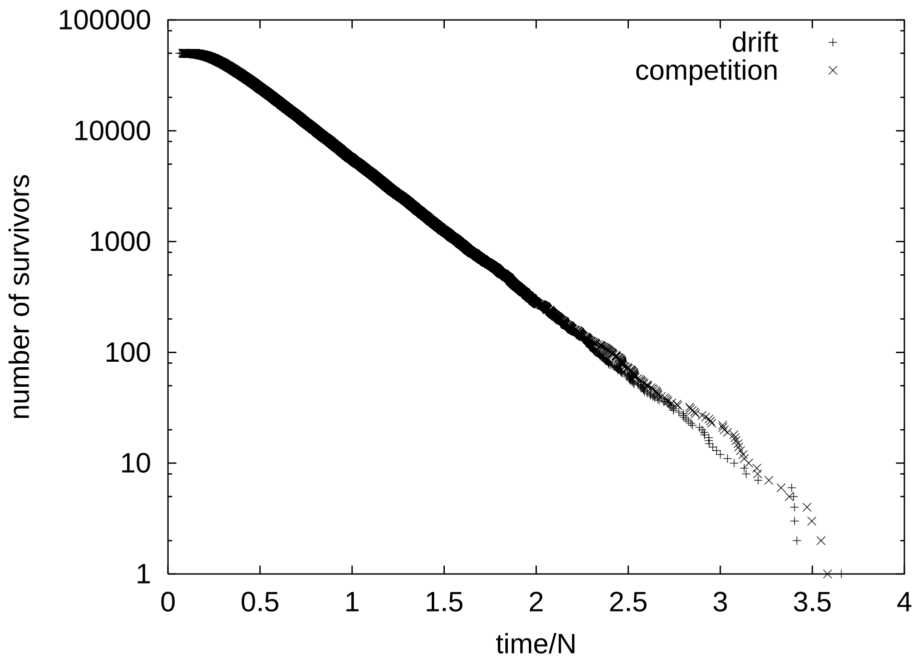

Figure 3.8: Number of survivors vs. time plots for the cyclic and drift case for system size $N = 6\ 000$.

the number of survivors vs. time plots for cyclic (competition) and drift cases together, for system size $N = 6\ 000$. It is clear that they agree.

3.4 Case when Rates Are not the Same

Now consider the cyclic (competition) case when the rates of the three reactions are not the same, i. e. the rate equations read like:

$$
\begin{aligned}
N\frac{dA}{dt} &= c_{13}AC - c_{12}AB \\
N\frac{dB}{dt} &= c_{21}AB - c_{23}BC \\
N\frac{dC}{dt} &= c_{32}BC - c_{31}AC
\end{aligned}
\tag{3.16}
$$

When the matrix c_{ij} is symmetric (which guarantees that N is conserved), we can always rescale the A, B, C, and the equations read:

$$\begin{aligned} N\frac{dA}{dt} &= A(\beta C - \gamma B) \\ N\frac{dB}{dt} &= B(\gamma A - \alpha C) \\ N\frac{dC}{dt} &= C(\alpha B - \beta A) \end{aligned} \tag{3.17}$$

Again, by adding these three equations, we see that $A + B + C = N$ =const. We can further manipulate the rate equations,

$$\begin{aligned} \frac{d}{dt}\ln A &= \beta C - \gamma B \\ \frac{d}{dt}\ln B &= \gamma A - \alpha C \\ \frac{d}{dt}\ln C &= \alpha B - \beta A \end{aligned} \tag{3.18}$$

where the system size N is absorbed into t, rescaling it. By adding these equations, we can find the second integral of motion to be $A^\alpha B^\beta C^\gamma = H$ =const. The new centre will then be not at the point (1/3, 1/3, 1/3), but at $A = \alpha, B = \beta, C = \gamma$. The lines along which the quantity $A^\alpha B^\beta C^\gamma$ is constant are shown in Fig. 3.9.

We need to check whether the exponential decay behaviour is universal when the symmetry is broken this way. For this we ran fifty thousand simulations for different combinations of α, β, and γ, for system size $N = 3000$, starting from the new centre ($A = N \cdot \alpha$, $B = N \cdot \beta$, $C = N \cdot \gamma$.) The number of survivors vs. time plots were again obtained, and those plots show that the exponential decay behaviour is robust. The time scale is now dependent on the initial distance from the boundary, with the equal rates case having

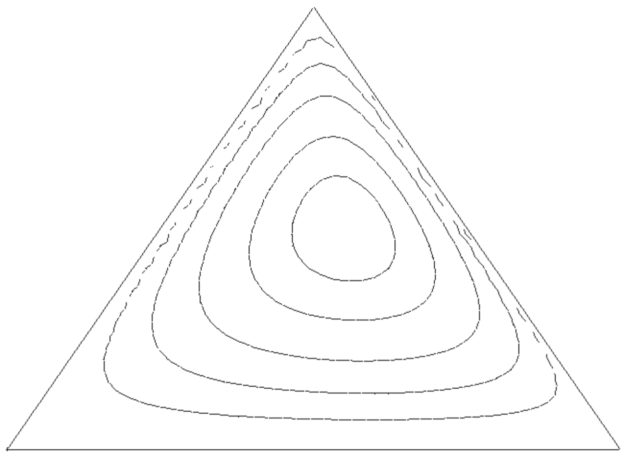

Figure 3.9: The H =const. lines for the case when the three reactions happen at different rates. Here α =0.2, β =0.3, γ =1$-\alpha-\beta$ =0.5

the largest distance, and hence the largest time scale (and the smallest slope.) Some of those plots are shown in Fig. 3.10. In that figure, the times for the equal rates case (for which the sum of rates is 3) are multiplied by three, to make them comparable to the times for the unequal rates case, for which the sum of rates is 1.

3.5 Probability Distribution for Long Times

To fully check the results of our computer simulations, we need to look at the shape of the probability distribution function and its evolution with time. In order to get an expression for W, we looked at our "experimental" data. We need to find a complete set of eigenfunctions of the Laplace equation, which satisfy the symmetry of the equilateral triangle, and also the Dirichlet problem, i.e. vanish at the boundary. Having this complete set of eigenfunctions for the

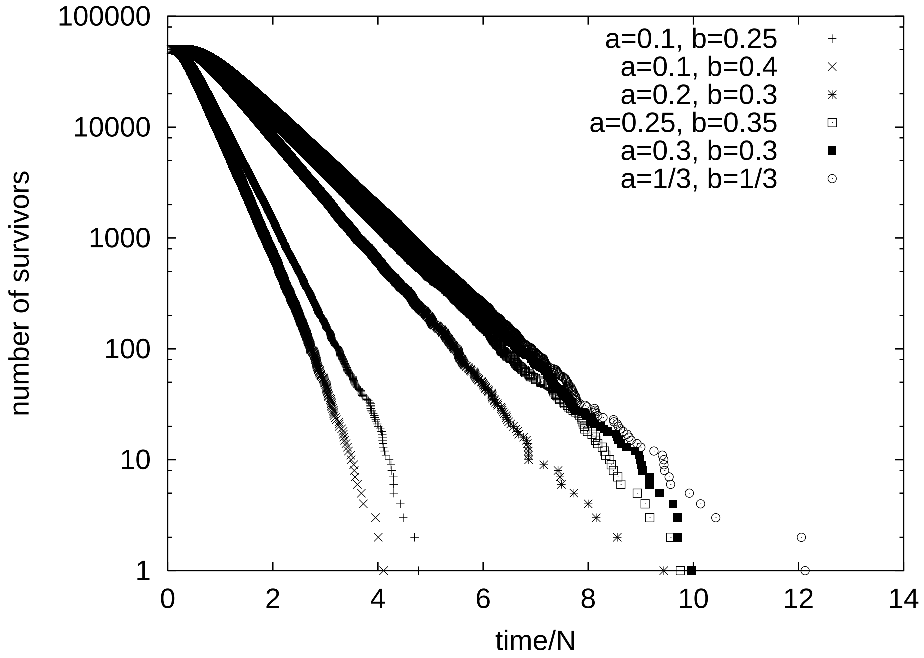

Figure 3.10: Number of survivors vs. time plots for the unequal rates case for system size 3000 and different combinations of α and β.

equilateral triangle, the probability density can be expanded as:

$$W = \sum_{m,n} c_{m,n}\phi_{m,n} \tag{3.19}$$

where $\phi_{m,n}$ are the above mentioned symmetric eigenfunctions. The problem of finding solutions of the equation $\Delta\phi + \lambda\phi = 0$ in an equilateral triangle, with zero boundary conditions (i.e. $\phi = 0$ at the boundary of the triangle) has been studied by Lamé [63] already in the nineteenth century.

3.5.1 The Eigenvalues of an Equilateral Triangle

Lamé [63] obtained an integral quadratic form for the eigenvalues. However, in his book, the complete list of eigenvalues is not clearly specified, and he gives no methods for computing the multiplicities of the eigenvalues. The problem has been revisited by Pinsky [64], who gives an up-to-date treatment of this problem.

The non-normalized eigenfunctions are then obtained from the general expression:

$$\phi_{(m,n)}(x_A, x_B) = \sum_{(m,n)} \pm \exp[\frac{2\pi i}{3}(nx_A + mx_B)] \tag{3.20}$$

where we are using the relation $x_A + x_B + x_C = 1$ and summation over the index pair (m, n) means summation for $(-n, m-n)$, $(-n, -m)$, $(n-m, -m)$, $(n-m, n)$, (m, n), $(m, m-n)$ [64]. The following conditions must also be fulfilled: $m+n$ is a multiple of 3, but $m \neq 2n$, and $n \neq 2m$. An eigenfunction is said to be symmetric when $R{\cdot}\phi = \phi$, and complex when $R{\cdot}\phi = \exp(\pm 2\pi i/3)\phi$, with R the operator of rotation by 120^0. An eigenvalue belongs to a symmetric eigenfunction iff m is also a multiple of 3. The eigenvalue belongs to a complex eigenfunction iff m is congruent to $+1$ or -1, modulo 3. In particular, an

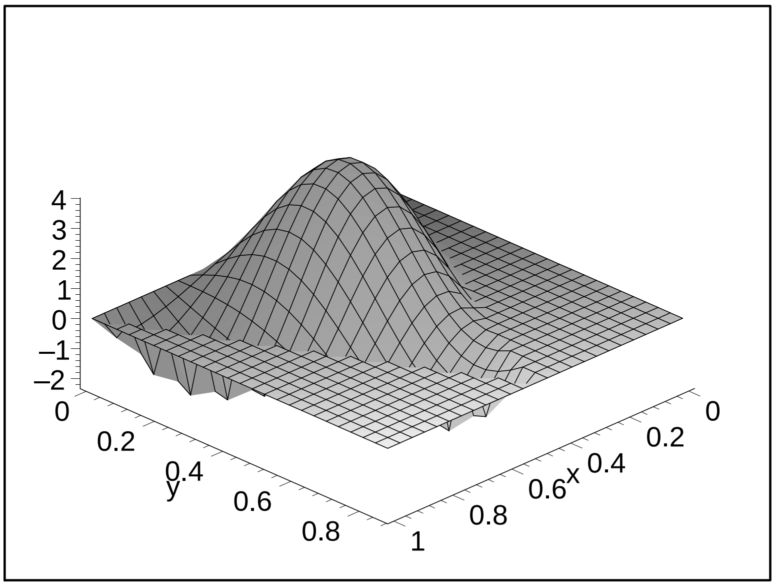

Figure 3.11: The Lamé function for $(m, n) = (1, 4)$ does not satisfy the Dirichlet boundary condition.

eigenvalue cannot belong to both a complex eigenfunction and a symmetric one. The following are then symmetric eigenfunctions and give a complete list of the simple (non-degenerate) eigenvalues:

$$\phi_p(x_A, x_B, x_C) = \sin(2\pi p x_A) + \sin(2\pi p x_B) + \sin(2\pi p x_C) \tag{3.21}$$

where $p = 1, 2, \ldots,$. It was necessary to check these functions, since we need to consider all the symmetric eigenfunctions which are non-zero at the centre of the triangle. Indeed, for the $(m, n) = (3, 0)$ we do obtain the first of the above series of symmetric eigenfunctions $p = 1$. If $m = -n$, $m = 2n$, or $n = 2m$, the eigenfunction is identically zero. And if the $n + m$ is not an integer multiple of 3, we get a function that does not satisfy the Dirichlet boundary condition, as shown in Fig. 3.11 for the pair $(m, n) = (1, 4)$.

The next eigenfunction with the desired symmetry is obtained for the

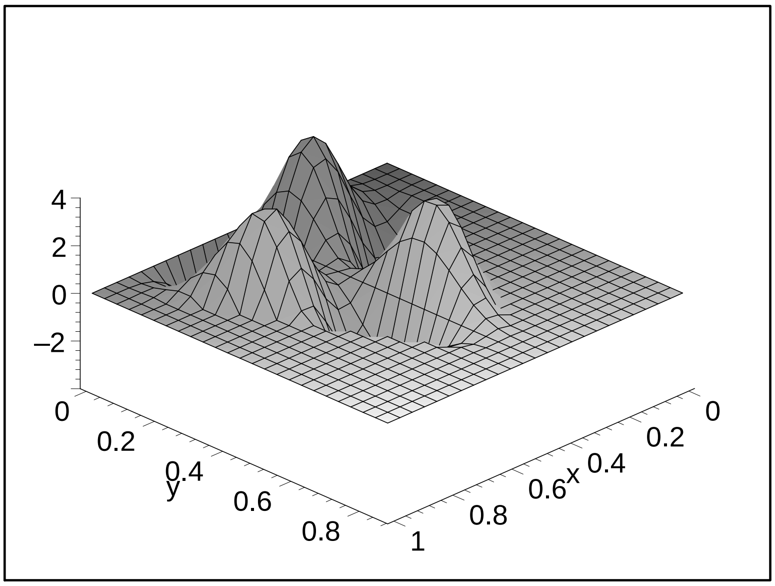

Figure 3.12: The real part of the Lamé function for $(m, n) = (6, -3)$ is antisymmetric with respect to a reflection about the three bisectors.

pair $(m, n) = (6, 0)$, and is the second non-degenerate eigenfunction from the list above $(p = 2)$. For the pair $(m, n) = (6, -3)$, the real part is antisymmetric with respect to a reflection about the three bisectors, as in Fig. 3.12.

The imaginary part of this function is symmetric. This is the first in the series of two-fold degenerate symmetric eigenfunctions, not mentioned by Pinsky in [64]. A plot of this function is given in Fig. 3.13.

We also verified these doubly degenerate functions to be orthogonal to the original eigenfunctions. Here is a list of the first few symmetric eigenfunctions:

$$\phi_{(3,0)} = \sin 2\pi x_B - \sin 2\pi(x_A + x_B) + \sin 2\pi x_A)$$

$$\phi_{(6,0)} = \sin 4\pi x_B - \sin 4\pi(x_A + x_B) + \sin 4\pi x_A)$$

$$\phi_{(6,-3)} = \sqrt{2}(\cos \pi(2x_A + x_B) \sin 5\pi x_B -$$

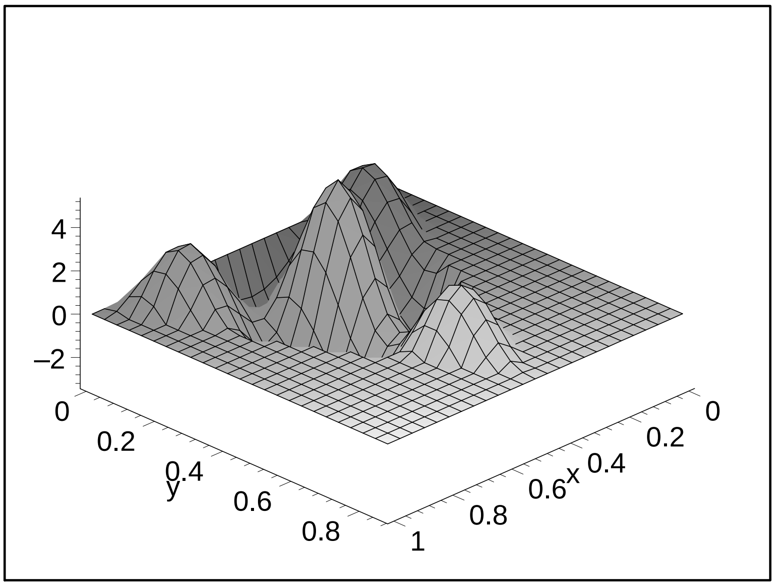

Figure 3.13: The Lamé function for $(m,n) = (6,-3)$ has a symmetric imaginary part which is non-zero at the centre of the triangle.

$$
\begin{aligned}
-\cos 2\pi(2x_A + x_B)\sin 4\pi x_B - \cos 3\pi(2x_A + x_B)\sin \pi x_B) \\
\phi_{(9,0)} = \sin 6\pi x_B - \sin 6\pi(x_A + x_B) + \sin 6\pi x_A) \\
\phi_{(3,12)} = \sqrt{2}(\cos \pi(2x_A + x_B)\sin 7\pi x_B - \\
-\cos 3\pi(2x_A + x_B)\sin 5\pi x_B - \cos 4\pi(2x_A + x_B)\sin 2\pi x_B) \\
\phi_{(12,0)} = \sin 8\pi x_B - \sin 8\pi(x_A + x_B) + \sin 8\pi x_A) \\
\phi_{(6,-9)} = \sqrt{2}(\cos 3\pi(2x_A + x_B)\sin 7\pi x_B + \\
+\cos 5\pi(2x_A + x_B)\sin \pi x_B - \cos 2\pi(2x_A + x_B)\sin 8\pi x_B) \\
\phi_{(15,0)} = \sin 10\pi x_B - \sin 10\pi(x_A + x_B) + \sin 10\pi x_A)
\end{aligned}
\tag{3.22}
$$

and now the pattern is established. These functions all normalize to $\frac{3\sqrt{3}}{2}$, and not to one.

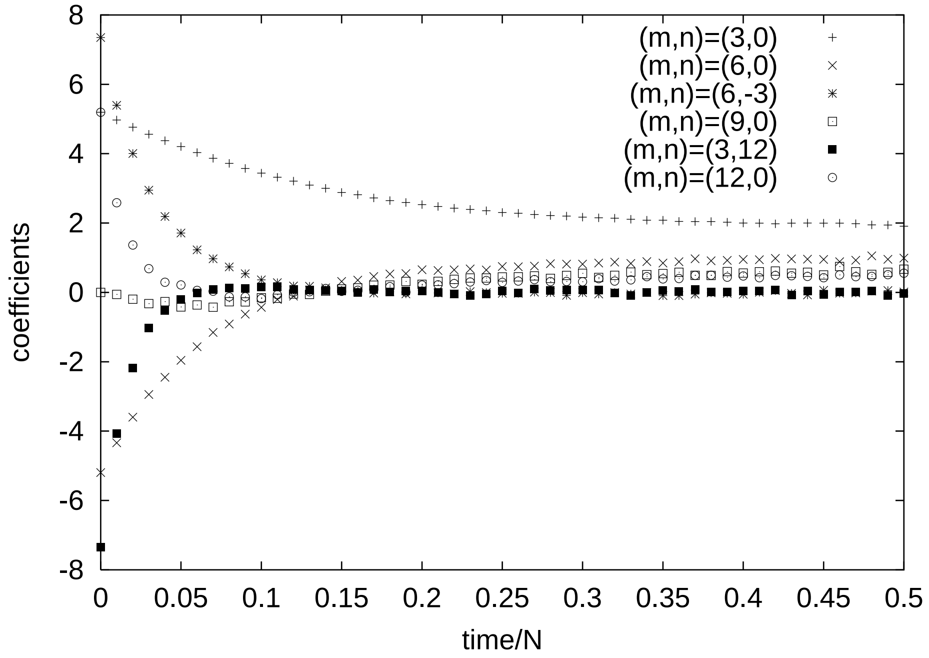

Figure 3.14: The variation with time of the coefficients for the first six functions in the expansion of $W(t)$: $(m,n) = (3,0)$, $(m,n) = (6,0)$, $(m,n) = (6,-3)$, $(m,n) = (9,0)$, $(m,n) = (3,12)$, $(m,n) = (12,0)$.

3.5.2 The Probability Distribution Function

In order to calculate the expansion coefficients, we took snapshots of the system at given times, and then used the relation:

$$c_{(m,n)}(t) \propto \frac{1}{p} \sum_{l=1}^{p} \phi_{(m,n)}(x_{Al}(t), x_{Bl}(t)) \tag{3.23}$$

where p is the number of experimental points. The time-evolution of the expansion coefficients is given in Fig. 3.14. The doubly–degenerate functions die out quite soon, while the Lamé symmetric functions persist.

With the coefficients obtained this way we can then construct the probability density at different times. Some snapshots at the $W(t)$ are shown in Fig. 3.15.

At $t = 0$ the $W(0)$ is a δ-function. As time passes, for $t = 0.1$ we see the W starts to spread, and it spreads even more for $t = 0.2$, becoming a "cake" for very long times, for both the cyclic system and the neutral drift one. For the neutral drift model, this "cake" has previously been obtained by Kimura [65]. The "cake" obtained this way shows some Gibbs oscillations, which are due to the inclusion of only a few terms in the expansion, and the presence of the absorbing boundary. For a uniform probability distribution, the calculated expansion coefficients are proportional to $1/m$, where m is the first index in the pair (m, n). It can be seen that our "experimental" expansion coefficients approach those values.

3.5.3 Numerical Solution of the Fokker-Planck Equation

The second–order Fokker–Planck equation for the drift case accepts solutions of the form:

$$W(x_A, x_B, x_C, t) = e^{-\lambda t} W(x_A, x_B, x_C) \tag{3.24}$$

with λ the smallest eigenvalue of the spatial equation. In other words, we need to solve the differential equation:

$$\frac{1}{N} L_{FP} W + \lambda W = 0 \tag{3.25}$$

where the L_{FP} operator has the form

$$\begin{aligned} L_{FP} = \frac{1}{2}[(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_A x_C + [(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_B} + \\ + \frac{\partial^2}{\partial y^2})x_A x_B + [(\frac{\partial^2}{\partial x_B^2} - 2\frac{\partial^2}{\partial x_B \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_B x_C] \end{aligned} \tag{3.26}$$

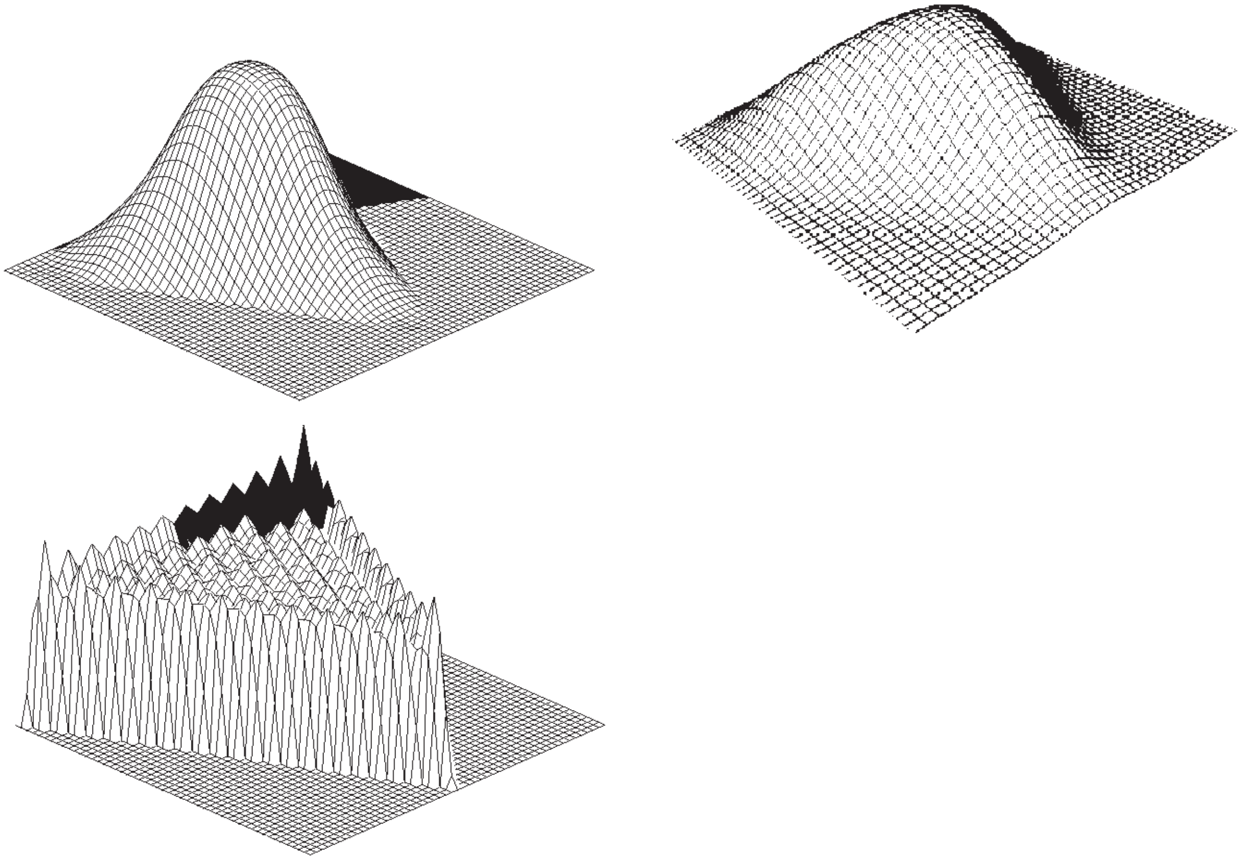

Figure 3.15: Snapshots at the probability density function at $t = 0.1$, $t = 0.2$, and after a very long time.

The N from the left hand side denominator can be absorbed into the exponential, resulting in time-rescaling (instead of time t, the time variable is now t/N). The form of the second-order term suggests using some numerical method for its solution. Since we already have a complete set of eigenfunctions for the symmetry of the problem at hand, and those eigenfunctions satisfy the same boundary conditions as the solution W, the Ritz-Galerkin method seems like a good choice. There are many numerical methods books which describe the method, so we will only outline it in the following section.

3.5.4 Ritz and Galerkin Methods for Solving Differential Equations

Consider the general differential equation

$$f(x, \dot{x}, \ddot{x}, \ldots, t) = 0 \tag{3.27}$$

with f a nonlinear function of its arguments. Assume a solution of the form:

$$\Phi(t) = \sum_{i=0}^{N} c_i \phi_i(t) \tag{3.28}$$

where $\phi_i(t)$ are a set of linearly independent functions, chosen appropriately for the given problem, and c_i's are constants we need to adjust to make the function $\Phi(t)$ as close to the exact solution as possible. The set of the functions $\phi_i(t)$ is chosen in such a way as to make the task of solving the problem as easy as possible. Given that, we must make sure that the solution satisfies the required initial and boundary conditions.

Next, we need to establish a criterion to estimate the acuracy of the assumed solution. Keep in mind that if $\Phi(t)$ were the exact solution, it would satisfy the differential equation, and would make the right hand side identically zero. Being just an approximate solution, the right hand side is not exactly zero, but some "residual" $e(t)$, which measures the error in the approximate solution. Since the "residual" must measure the error over a wide time interval, it will be necessary to account for all the time. In analogy with the method of least squares for fitting the best line through a set of experimental points, we take not the "residual" but its square as a measure of the goodness of the approximate solution. This means that the integral

$$E = \int_a^b e^2(t)dt \tag{3.29}$$

must be minimized. Here the interval of interest for the variable is $[a, b]$. (Otherwise, the integral must be taken over the area or volume of interest.) Then the integral E must be minimized by adjusting the constants c_i. Otherwise the differential of E when the variables are c_i (and they are independent) ought to be zero. This means that

$$\frac{\partial E}{\partial c_i} = \frac{\partial}{\partial c_i} \int e^2(t)dt = 0 \tag{3.30}$$

or

$$\int e(t) \frac{\partial e(t)}{\partial c_i} dt = 0 \tag{3.31}$$

where $i = 0, 1, 2, \ldots, N$. These equations can be solved for the coefficients c_i. This procedure is called the Ritz method and Eq. (3.31) is called the Ritz condition.

The Galerkin method is identical to the Ritz method, except that the Ritz condition (3.31) is replaced by the Galerkin condition:

$$\int e(t)\phi_i(t)dt = 0 \tag{3.32}$$

Here $\phi_i(t)$ are the same set of linearly independent functions used in the Ritz method. In Eq. (3.32) the residual is said to be "orthogonal" to all the ϕ_i $i = 0, 1, 2, \ldots, N$.

3.5.5 The Galerkin Method Solution of Our Eigenvalue Problem

For our problem, the Galerkin method of solution will produce equations of the form:

$$\oint \phi_i(x_A, x_B, x_C) L_{FP} \sum c_j \phi_j(x_A, x_B, x_C) dA = 0 \tag{3.33}$$

where the area of integration is that of an equilateral triangle with height equal to unity. The integration is done by transforming into the cartesian coordinates x and y using the transformation formulae

$$\begin{aligned} x_A &= 1 - x - \frac{y}{\sqrt{3}} \\ x_B &= x - \frac{y}{\sqrt{3}} \\ x_C &= \frac{2y}{\sqrt{3}} \end{aligned} \tag{3.34}$$

A surface integral over the triangle can then be written

$$\int_0^{\sqrt{3}/2} \int_{\frac{y}{\sqrt{3}}}^{1-\frac{y}{\sqrt{3}}} f(y, x) dx dy \tag{3.35}$$

Keeping only the first six non-degenerate (Lamé) eigenfunctions in the expansion of W, and applying the Galerkin formulae (3.33), we obtain six independent algebraic equations, linear in c_i, which contain seven unknowns: the c_i's, as well as the eigenvalue λ. Since we are only interested in λ, we can set one of the c_i's, say c_1, equal to one, and express the others in terms of c_1. All this information was fed into the Maple machine, which gave a value of $\lambda = 3.01$ for the smallest eigenvalue, and coefficients c_i (starting from $i = 2$) that go down as $1/i$, as is the case with the "experimentally" obtained coefficients. Hence a very good agreement between our "experimental" data and the numerical solution results.

3.6 Case when Drift Is Biased

So far we have studied a three-component version of the famous Kimura model of neutral genetic drift [19,20] in which the offspring of two individuals can be of either type with equal probability. We compared it to the cyclic case, in which deterministic dynamics is prevalent in the evolutionary process. In biological literature we often encounter the notion of "selection coefficient", which relates to the case when the dynamics is not completely neutral, but rather exhibits a (slight) bias toward one of the species. It is known that a mutation whose selection coefficient is much less than the inverse of the population size behaves as if it were strictly neutral. The nearly neutral mutation theory deals with mutations whose selection coefficients are close to the inverse of the population size [22].

The cyclic model would correspond to a completely biased drift, while the neutral model to a completely unbiased one. It is of interest to our biologist colleagues to check that the behaviour of our model is monotonic, when the model is biased at different degrees. For that we will start by assuming that the reaction $A + B \to 2A$ happens with probability p, and the alternative one,

$A + B \rightarrow 2B$, with probability $q = 1 - p$. The master equation for such a system will read:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \frac{1}{N}[(p\epsilon_C\epsilon_A^{-1} + q\epsilon_A\epsilon_C^{-1} - 1)AC + (p\epsilon_A\epsilon_B^{-1} + q\epsilon_B\epsilon_A^{-1} - 1)AB +$$
$$+(p\epsilon_B\epsilon_C^{-1} + q\epsilon_C\epsilon_B^{-1} - 1)BC]P(A,B,C,t) \tag{3.36}$$

The Kramers-Moyal expansion of this master equation can be performed as it was done for the purely cyclic or neutral models. The resulting Fokker-Planck equation is:

$$\frac{\partial W}{\partial t} = (p-q)[(\frac{\partial}{\partial x_A} - \frac{\partial}{\partial x_C})x_A x_C + (\frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_A})x_B x_A +$$
$$+(\frac{\partial}{\partial x_C} - \frac{\partial}{\partial x_B})x_C x_B]W + \frac{1}{2N}[(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_A x_C +$$
$$+(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_B} + \frac{\partial^2}{\partial x_B^2})x_A x_B + (\frac{\partial^2}{\partial x_B^2} - 2\frac{\partial^2}{\partial x_B \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_B x_C]W \tag{3.37}$$

This equation combines the first order term from the equation for the cyclic competition model (3.15), multiplied by $s = p - q$, and the second order term from both the cyclic and neutral model (3.14). The work we described in this chapter shows that the first order term only drives the system around the closed mean-field trajectories shown in Fig. 3.2. The term that actually drives the system toward the boundary is the second order term, or the diffusion one. This term is identical for the cyclic, neutral and partially neutral versions of the model, and it does not depend on s. We would then expect the partially neutral model to exhibit the same behaviour the two "extremes" exhibit, as far as extinction times and exponents are concerned. This hypothesis was verified by running simulations of our system for different values of p, i.e. also different values of the selection coefficient s. Some of the survival probability vs. time plots are shown in Fig. 3.16. In that figure we show the survival probability

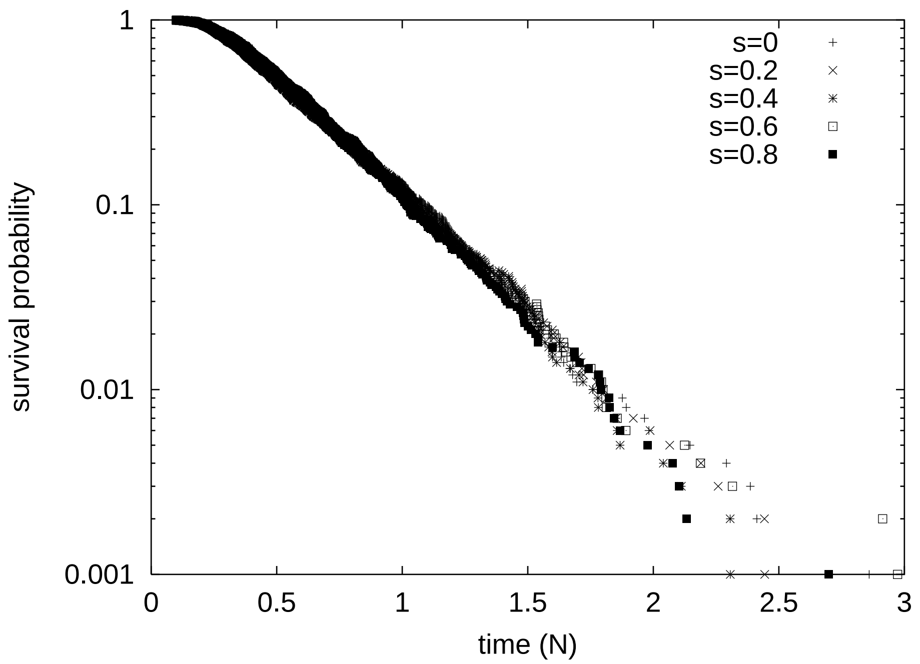

Figure 3.16: Number of survivors vs. time plots for the case when $s = p - q$ is between zero and one for system size $N = 6\ 000$. The extinction times plot for the drift case is shown for comparison.

vs. time plot for the pure neutral drift case for comparison (the cyclic model plots agree with the neutral model ones). The agreement is apparent.

3.7 Conclusions

In this chapter we have considered an ABC model in both the cyclic competition and neutral genetic drift versions, and studied the long-term behaviour of both these models. The number of the A, B, C species in the cyclic competition case oscillates with a drifting with time amplitude, until one of the species (and then the next one in the cycle) goes extinct. In the neutral drift case the number of the A, B, C species drifts, until one of them becomes zero. Afterwards the three allele neutral model becomes a two allele one, and one of the remaining species becomes extinct, leaving only one of them in the picture. In the cyclic case the amplitude of the oscillations drifts, while in the neutral one the population numbers drift, making the neutral model seem as an adiabatic approximation of the cyclic one.

In both scenarios the number of survivors vs. time plots show an exponential decay, with the same exponent: -3. The exponent does not depend on the system size. The extinction times scale exactly as the size N, at least when the population size is reasonably large. This conclusion is verified by checking the distribution of the product $H = x_A x_B x_C$ on the condition that the system be alive. The plots for long enough times are straight lines, indicating a linear probability distribution for the H quantity. This is compatible with the exponential decay. It is very interesting and important to note that there is no difference in the time scale for the ensemble of species in cyclic competition and the case of neutral genetic drift.

The probability distribution function for the whole system has been reconstructed, based on "experimental" data, i.e. data from simulations. The eigenfunctions of an equilateral triangle that satisfy the Dirichlet boundary conditions were used as expansion basis, and the expansion coefficients were calculated, using a sample of 3000 copies of the system as a source for the data. The expansion coefficients decay as $1/n$ (with n the index of the eigenfunction), indicating an essentially flat probability distribution function, which was verified by plotting the function. It looks like a "cake" in which Gibbs oscillations are present, reflecting the finite number of terms taken into account, as well as the presence of the absorbing boundary.

The result of the simulations was verified by writing and solving the Fokker–Planck equation for the second model. It is interesting to note that this is one of the rare cases in which the Kramers-Moyal expansion of the master equation yields the correct results. Finally, the robustness of the exponential decay for the survival probability is checked against variations in the rate of the three different reactions in the system. The slope in this case depends only on the distance of the centre (starting point) from the boundary, i.e. the relationship between the rates of different reactions.

In the case when the selection coefficient is not zero or one, but takes on intermediate values, the Fokker-Planck equation suggests that the extinction behaviour of the system should not change. This was verified by computer simulations of the model for different values of the selection coefficient. They show a good agreement between the survival probability vs. time curves for different selection coefficients.

Chapter 4

Crossover Behaviour of 3-Species Systems with Mutations or Migrations

4.1 Introduction

So far we have considered a non-spatial version of the ABC model in both the cyclic competition and neutral genetic drift versions, and studied the evolutionary time behaviour of such a model. The number of the A, B, C species oscillates with an amplitude that drifts with time, until one of the species (and then the second one) goes extinct, i.e. biodiversity is lost. The number of survivors vs. time plots show an exponential decay. It is very interesting to note that there is no difference in the extinction time scale for the ensemble of competing species and the case of neutral drift. This allows us to think of the neutral model as an "adiabatic approximation" of the cyclic system. Considering that the cyclic competition system would be a minimal model (three alleles) of Darwinian evolution picture, it seems that, by just looking at the results, we cannot tell which one must have been the mechanism for the evolution!

There is growing concern about the effects of habitat fragmentation in the survival of a species [66]. Small, isolated forest fragments lose species–we have many examples in the case of fast deforestation worldwide [70]. Some species were lost within a few years of their isolation from the once–continuous forest [71]. Oscillations in the number of one "species" are observed in the case of epidemics, as well as ecology. Throughout this century we have observed changes in patterns of epidemics [72]. For some diseases, the major transitions have been between regular cycles and irregular epidemics, and from regionally synchronized oscillations to complex, spatially incoherent epidemics. Sinervo et al. [8, 73] observed spatio-temporal oscillations in the number of lizards (male and female) that employ different mating strategies.

Many population ecologists then conclude that a very important issue is the synchrony of population dynamics in different habitat patches [67–69, 74,75]. If the population of a certain species goes extinct in one patch while it still survives in other patches, then the hope is that what is known as "rescue effect" can prevent global extinction [67]. Otherwise, the population ecologists say, if there is synchrony of population dynamics between patches, the species is doomed to go extinct altogether. In this framework, there have been many proposed strategies for preservation, one of the most debated of which is the construction of "conservation corridors" that would make it possible for individuals to move along habitat patches [76, 77].

Considering that each habitat patch can be thought of as one copy of our non-spatial ensemble, the broad range of extinction times might help explain the persistence (for some time) of species that exist in different "versions", such as certain lizards [8, 73], or even some kinds of bacteria. Applied to its epidemiological scenario, the result would not explain the endemicity of certain diseases, (namely those with mutating pathogen,) since there should be outbreaks and pandemics in patches (areas, cities). Since neither the animals, nor the humans are immobile, it seemed reasonable to study the behaviour of the three-species cyclic and/or neutral system in the presence of mutations or migration. In the following section we introduce our model, and in the next ones, we present the results of our calculations and simulations.

4.2 The Model

Consider a non-spatial (well-stirred) system in which three species A, B, C are competing in a way described by the reactions: $A+B \to 2B$ at a rate AB/N, $B+C \to 2C$, $C+A \to 2A$ at corresponding rates. Add to it the reactions: $A \to B$ at rate μA, $A \to C$ at rate μA, $B \to A$ at rate μB, $B \to C$ at rate μB, $C \to A$ at rate μC, and $C \to B$ at rate μC, where μ is the probability of mutations per individual particle in unit time.

The rate equations for this system will be:

$$\begin{aligned}
\frac{dA}{dt} &= \frac{AC}{N} - \frac{AB}{N} + \mu B + \mu C - 2\mu A \\
\frac{dB}{dt} &= \frac{BA}{N} - \frac{BC}{N} + \mu A + \mu C - 2\mu B \\
\frac{dC}{dt} &= \frac{CB}{N} - \frac{CA}{N} + \mu A + \mu B - 2\mu C
\end{aligned} \tag{4.1}$$

with $A+B+C = N = \text{const}$. If the neutral drift system is considered instead, i.e. when $A+B \to 2A$ or $2B$ with equal probability (and similarly for the two other reactions), the rate equations will only contain the terms that depend on μ, and the first two terms will be absent.

In another version, mutations are replaced by migration into and out of the "island". In other words, the following migrations are happening: A out at rate $3\mu A$, B out at rate $3\mu B$, C out at rate $3\mu C$, A, B, or C in at constant rate of $\mu \cdot N$, where N is the system size (to make it comparable to the rate of leaving the island. This would correspond to the average number of individuals in other patches, which are the source of our "immigrants".) The rate equations will be the same as those for the system with mutations. The above equations have a fixed point at $A = B = C = N/3$, and it is a stable solution.

However, the rate equations are just a "mean field" approximation; they only describe the behaviour of the average values of the individual populations. In the real world, the system is subject to stochastic noise due to birth and death processes (intrinsic noise), which we take to be Poisson-distributed. The random nature of these processes need be taken into consideration, if we want to get the correct picture of the evolution of the system. For that we ought to write the master equation, and then solve it: first expand it, obtaining a Fokker-Planck equation, and solve the F-P equation. There are two most commonly used expansions of the master equation: the Kramers-Moyal expansion, which is essentially a Taylor expansion in powers of system size N (we used it in the previous chapter, to study the isolated three-species systems), and the Ω– expansion of van Kampen [35], which is an expansion in powers of $\sqrt{N}$. The first method produces a nonlinear Fokker-Planck equation, which, as we saw previously, is quite difficult to solve. The second method is quite systematic, and yields satisfactory results when the system has a single stable point, as the rate equations suggest is the case with the present model. In the absence of this stable solution, the task becomes very difficult [30].

Using the "shift" operators notation:

$$\epsilon_A f(A,B,C) = f(A+1,B,C)$$
$$\epsilon_A^{-1} f(A,B,C) = f(A-1,B,C)$$

the master equation for the cyclic competition system with mutations reads:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \{\frac{1}{N}[(\epsilon_C\epsilon_A^{-1}-1)AC + (\epsilon_A\epsilon_B^{-1}-1)AB + (\epsilon_B\epsilon_C^{-1}-1)BC] +$$
$$+\mu[(\epsilon_A\epsilon_B^{-1}+\epsilon_A\epsilon_C^{-1}-2)A + (\epsilon_B\epsilon_C^{-1}+\epsilon_B\epsilon_A^{-1}-2)B +$$
$$(\epsilon_C\epsilon_A^{-1}+\epsilon_C\epsilon_B^{-1}-2)C]\}P(A,B,C,t) \tag{4.2}$$

while that for the neutral drift system with mutations:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \{\frac{1}{2N}[(\epsilon_C\epsilon_A^{-1}+\epsilon_A\epsilon_C^{-1}-2)AC +$$
$$(\epsilon_A\epsilon_B^{-1}+\epsilon_B\epsilon_A^{-1}-2)AB + (\epsilon_B\epsilon_C^{-1}+\epsilon_C\epsilon_B^{-1}-2)BC] +$$
$$+\mu[(\epsilon_A\epsilon_B^{-1}+\epsilon_A\epsilon_C^{-1}-2)A + (\epsilon_B\epsilon_C^{-1}+\epsilon_B\epsilon_A^{-1}-2)B +$$
$$(\epsilon_C\epsilon_A^{-1}+\epsilon_C\epsilon_B^{-1}-2)C]\}P(A,B,C,t) \tag{4.3}$$

and that for the cyclic system with migrations:

$$\frac{\partial P(A,B,C,t)}{\partial t} = \{\frac{1}{N}[(\epsilon_C\epsilon_A^{-1}-1)AC + (\epsilon_A\epsilon_B^{-1}-1)AB + (\epsilon_B\epsilon_C^{-1}-1)BC] +$$
$$+3\mu[(\epsilon_A-1)A + (\epsilon_B-1)B + (\epsilon_C-1)C +$$
$$+\frac{N}{3}(\epsilon_A^{-1}+\epsilon_B^{-1}+\epsilon_C^{-1}-3)]\}P(A,B,C,t) \tag{4.4}$$

4.3 The "Fixation" Regime

In the previous chapter [30] we studied the cyclic competition and neutral genetic drift systems in absence of mutations. In the mean-field approximation (rate equations) the cyclic system has an infinity of neutrally stable solutions: any trajectory that conserves the integral of motion $H = ABC/N^3$ is a stable trajectory of the system. On the other hand, the neutral drift system has an infinity of neutrally stable points: any state of the system is stable in the mean-field picture. The story is different, when stochastic birth-and-death processes are taken into account: then the individual population distribution drifts, until one of the species, and then another one, go extinct. In other words, there is fixation of the population onto one of the varieties, and the order parameter, which can be defined as the product $H = ABC/N^3$ becomes zero. The survival probability versus time (in units of N) plot exhibits exponential decay with slope -3, which was verified by numerically solving the corresponding Fokker-Planck equation. This equation was obtained from a Kramers-Moyal expansion, since the infinite multiplicity of stable solutions (points or limit cycles) makes the van Kampen method impractical to apply. The first extinction time scales with the total population size N. This behaviour is the same for both the cyclic competition and the neutral drift models, the second one behaving as an "adiabatic approximation" to the first one. For both models, the probability distribution function becomes uniform within a short time, i.e. any point inside the triangle that represents the phase space of the system is equally probable.

When μ is small (the meaning of small will become clear in section 4.5) the system with mutations/migrations is in the "fixation" regime, in which the results of our previous study [30] are applicable qualitatively. In that regime, the stochastic processes drive the system towards the (absorbing) boundary, in which at least one of the species has met extinction, and the order parameter is zero. This was verified by computer simulations of the master equations (4.2, 4.3, 4.4). We observed a "fixation" regime, in which the main reaction is the cyclic one, and there are only occasional mutations/migrations, which are not frequent enough to prevent fixation. In other words, the system is experiencing "forces" due to fluctuations, and "forces" due to mutations (migrations). The fluctuations are pushing the system toward the boundary, while the mutations (migrations) push toward the centre. For small mutation/migration rate the "fluctuation-generated forces" are stronger. Fig. 4.1 shows the variation with time of the population of C's for a realisation of the system in the "fixation" regime. Occasionally, an individual of the competing species (the one next in the "food chain") is introduced in the system by mutations (or migrations), and then it either causes an occasional spike in the time series (like the one we see in Fig. 4.1 between $t = 25$ and $t = 30$), or it "eats up" the old species completely and replaces it in the system (which is what happens between $t = 35$ and $t = 40$ in our plot).

We investigated the behaviour of the probability distribution for the first crossing of the boundary, i.e. when one of the populations becomes zero for the first time (extinction of order parameter) as the mutation/migration rate increases. For that we ran ten thousand copies of the system for each value of the mutation rate. The survival probability (as defined above) was plotted vs. time. In the "fixation regime" the decay is still exponential, but the exponent varies with the mutation/migration rate, getting close to zero as the system approaches criticality. Fig. 4.2 shows a plot of the exponent as a function of mutation probability μ for system size $N = 150$ and $N = 300$.

We ran statistics for the order parameter at different values of the mu-

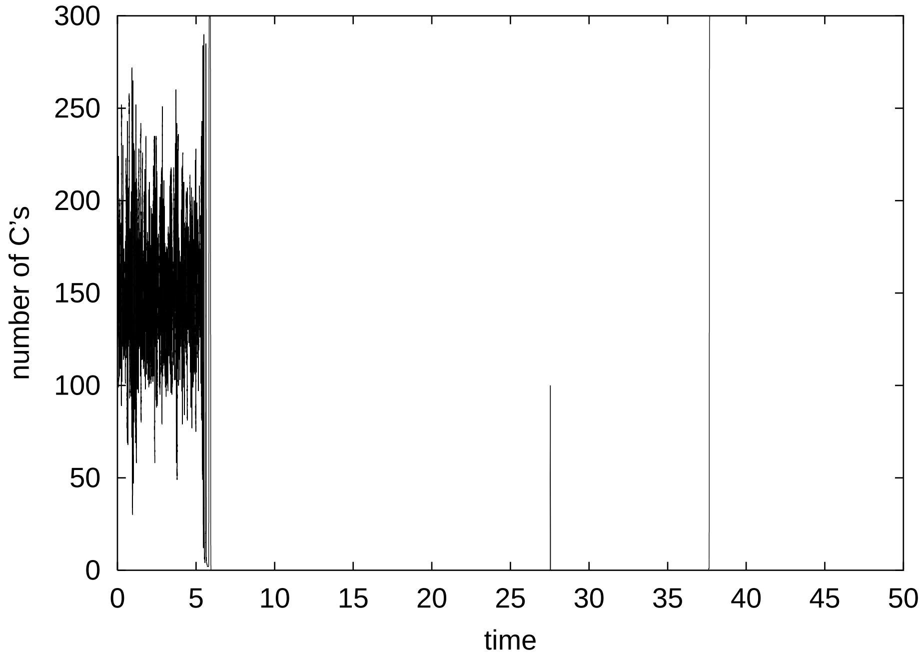

Figure 4.1: The time series for the number of C species in the "fixation" regime (here $\mu = 0.4 \cdot 10^{-3}$, and system size $N = 300$).

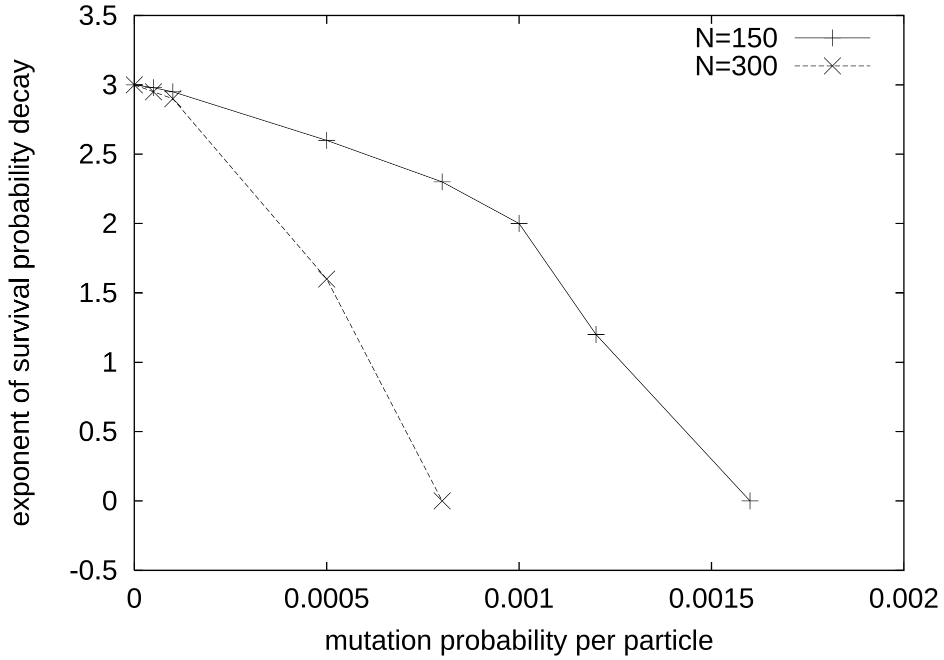

Figure 4.2: Dependence of the exponent of the decay of the survival probability on the mutation rate in the "extinction" regime for system size $N = 150$ and $N = 300$.

tation rate μ within the "fixation" regime. For each value, 3000 copies of the system were simulated, and the average order parameter was calculated as a function of time. Some of those plots are shown in Fig. 4.3 for system size $N = 300$. For all the copies of the system, we start at the centre, i.e. order parameter $H_0 = 1/27$. When the mutation rate is low (in our plot $\mu = 0.0004$,) the average order parameter goes down quickly, and then remains about constant (but very small) for long times, due to the existence of copies of the system which are still non-extinct. As the mutation rate increases, the plot of the average order parameter with time "flattens out"; the drop is slower, over a longer time interval. This can be seen in two intermediate plots in Fig. 4.3, which correspond to $\mu = 0.001$ and $\mu = 0.0014$. The same sort of evolution of the order parameter with time when the mutation rate is varied is obtained for the neutral drift case. When the mutation rate gets close to the critical value, the average order parameter initially drops, and then decays slowly over a very long time interval.

4.4 The "Diversity" Regime

If only mutations (migrations) are present, the system remains near the centre point ($A = B = C = N/3$), and the order parameter remains considerably above zero; in other words, all three species are present in the system. One can occasionally observe temporary extinctions, but this happens very seldom, and for very long times; when the system size is very large, it essentially never happens. Since the boundary is not absorbing, occasional mutations/migrations will return the system to the state with maximal symmetry (biodiversity) where all three varieties coexist. When both the cyclic/neutral drift mechanism and mutations (migrations) are present, and μ is above critical, the muta-

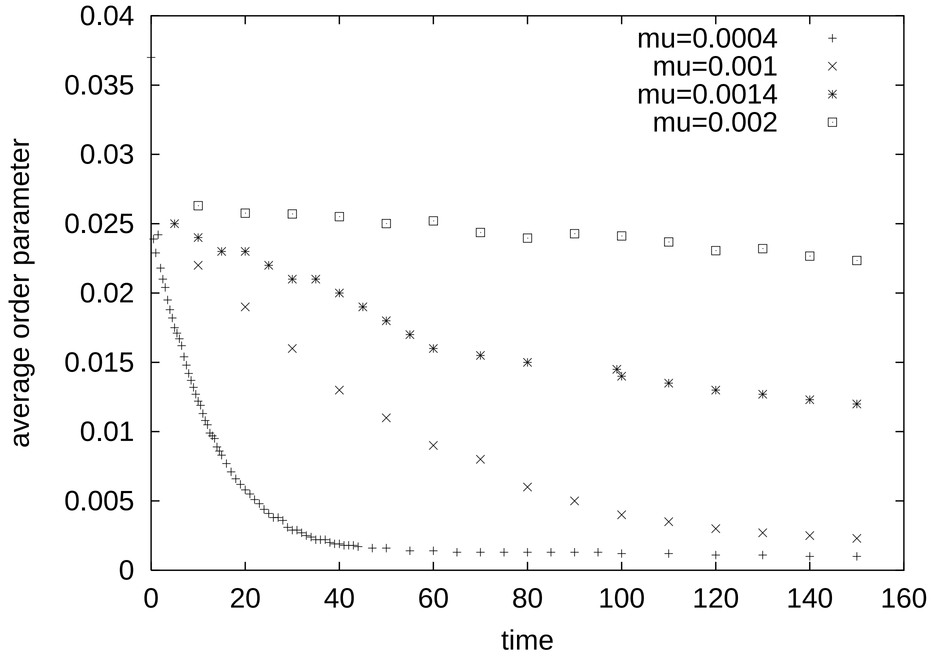

Figure 4.3: Time series of the order parameter for system size $N = 300$, and different values of μ. For $\mu = 0.002$, Eq. (4.23) gives $\eta \simeq 0.026$.

tions manage to keep the system maximally disordered, since they are stronger than the fluctuations (which, as we saw, try to drive the system toward the boundary, i.e. fixation, and keep it there). Compared to the situation with no/low mutations/migrations, where the boundary is (practically) absorbing, and the final state of the system is a "pure" one (with only one of the species present), the mutation/migration rate acts then as a "temperature".

Work has been done on the two allele almost neutral drift model with mutations [28]. The almost neutral model with mutations, preserving the total number of individuals, has only one degree of freedom, and allows one to derive an "effective potential" from the Fokker-Planck equation, obtained by a Kramers-Moyal expansion of the master equation. For small mutation probabilities, such that $2\mu N \ll 1$, there is extinction of one species and fixation. The effective potential is symmetric around the centre (where both species are in equal numbers) and the branches of the effective potential are down. This allows for the system to quickly slip into the state where only one of the species is present. Otherwise, both species coexist forever in the high mutation regime, i.e. when $2\mu N \gg 1$. In that regime, the effective potential is symmetric around the centre point, but with branches upwards, which means that the centre point is a minimum potential point. The system then remains in the vicinity of that point for very long times. The effective potential "flips" from "branches up" to "branches down" at the point where $2\mu N = 1$. The transition is second-order, and critical behaviour is observed. The system behaves similarly when migrations are present, instead of mutations. One aspect of migrations in a four-species system has been treated recently by Togashi and Kaneko [100].

Since the rate equations have a stable solution, we employ the van Kampen expansion [35]. The idea of this expansion is to split the variables of the problem into a non-fluctuating part, and a fluctuating one, i.e. deal separately with the mean-field solutions and the fluctuations. In this approach, the numbers of the individual populations would be written as:

$$
\begin{aligned}
A &= N\phi_1 + \sqrt{N}x_1 \\
B &= N\phi_2 + \sqrt{N}x_2 \\
C &= N\phi_3 + \sqrt{N}x_3
\end{aligned}
\tag{4.5}
$$

Here the ϕ_i are the concentrations of A, B, and C species respectively (which only depend on time), and the x_i are the fluctuations (proportional to the square root of system size). The system size (total population) N is conserved for the system with mutations, but not for that with migrations. This conservation rule will require special attention in the case of the system with mutations. Using the van Kampen Ansatz, the probability distribution $P(A, B, C, t)$ is transformed into $\Pi(\{x_i\}, t)$, and the following relations are true:

$$
\begin{aligned}
\Pi &= N^{3/2} P(N\{\phi_i + \sqrt{N}x_i\}, t) \qquad (4.6) \\
\frac{\partial P}{\partial t} &= \frac{1}{N^{3/2}} \frac{\partial \Pi}{\partial t} - \frac{1}{N} \sum \frac{d\phi_i}{dt} \frac{\partial \Pi}{\partial x_i}
\end{aligned}
$$

and

$$
\begin{aligned}
\epsilon_i &= 1 + \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_i} + \frac{1}{2N} \frac{\partial^2}{\partial {x_i}^2} + \ldots \qquad (4.7) \\
{\epsilon_i}^{-1} &= 1 - \frac{1}{\sqrt{N}} \frac{\partial}{\partial x_i} + \frac{1}{2N} \frac{\partial^2}{\partial {x_i}^2} + \ldots
\end{aligned}
$$

Next we substitute everything into the master equation, leave only the term $\partial\Pi/\partial t$ on the left hand side, and group the right hand side terms according to powers of $\sqrt{N}$. The first term is of order $N^{1/2}$, and it must be equal to zero, for an expansion in terms of $N^{1/2}$ to make sense. That term for e.g. the cyclic model with mutations is:

$$\sum \frac{\partial \Pi}{\partial x_i}[\frac{d\phi_i}{dt} + \phi_i\phi_{i+1} - \phi_i\phi_{i+2} + \mu(2\phi_1 - \phi_{i+1} - \phi_{i+2})] = 0, \tag{4.8}$$

which reproduces the rate equations in terms of the concentrations ϕ_i, with steady state solution $\phi_i = 1/3$. Similarly, we get the rate equations for the other models.

The terms of order N^0 give a linear Fokker-Planck equation of the form:

$$\frac{\partial \Pi}{\partial t} = \sum[-A_{ik}\frac{\partial}{\partial x_i}(x_k\Pi) + \frac{1}{2}B_{ik}\frac{\partial^2 \Pi}{\partial x_i \partial x_k}] \tag{4.9}$$

4.4.1 The System with Migrations

We are going to solve the system with migrations first, since the absence of the conservation of total population N makes this system easier to deal with.

The A-matrix for the cyclic system with migrations is:

$$\begin{pmatrix} \phi_3 - \phi_2 - 3\mu & -\phi_1 & \phi_1 \\ \phi_2 & \phi_1 - \phi_3 - 3\mu & -\phi_2 \\ -\phi_3 & \phi_3 & \phi_2 - \phi_1 - 3\mu \end{pmatrix}$$

while that for the neutral system is diagonal, with all the diagonal elements equal to -3μ. The B-matrix is the same for both the cyclic and neutral systems:

$$\begin{pmatrix} \mu(1+3\phi_1) + \phi_1\phi_2 + \phi_1\phi_3 & -\phi_1\phi_2 & -\phi_1\phi_3 \\ -\phi_1\phi_2 & \mu(1+3\phi_2) + \phi_1\phi_2 + \phi_2\phi_3 & -\phi_2\phi_3 \\ -\phi_1\phi_3 & -\phi_2\phi_3 & \mu(1+3\phi_3) + \phi_2\phi_3 + \phi_1\phi_3 \end{pmatrix}$$

The Fokker-Planck equation is linear, and the coefficients depend on time through ϕ_i's. This approximation is otherwise known as "linear noise approximation". The solution is known to be a Gaussian; the problem represents itself as an Ornstein-Uhlenbeck process. For our purposes, it suffices to determine the first and second moments of the fluctuations. Following van Kampen [35], we can multiply the Fokker-Planck equation by x_i and integrate by parts to get:

$$\frac{d < x_i >}{dt} = \sum_j A_{ij} < x_j > \tag{4.10}$$

For simplicity, let us limit our attention to fluctuations around the steady state $\phi_i = 1/3$. The A-matrix for the system with migrations (4.4.1) becomes:

$$\begin{pmatrix} -3\mu & -1/3 & 1/3 \\ 1/3 & -3\mu & -1/3 \\ -1/3 & 1/3 & -3\mu \end{pmatrix}$$

The eigenvalues of the A matrix for the cyclic system are of the form -3μ, $-3\mu \pm i/\sqrt{3}$, while those for the neutral system are degenerate and equal to -3μ. The negativity of the eigenvalues guarantees the stability of the zero solutions to the first moments equations. Hence, the average of the fluctuations decays to zero and remains zero. The equations for the second moments can be obtained similarly:

$$\frac{d < x_i x_j >}{dt} = \sum_k A_{ik} < x_k x_j > + \sum_k A_{jk} < x_i x_k > + B_{ij} \tag{4.11}$$

They depend on time through ϕ_i's. Then, the equations for the second moments for e.g. the cyclic system will read:

$$\frac{d < x_1^2 >}{dt} = 2(\phi_3 - \phi_2 - \mu) < x_1^2 > +(\phi_2 - \phi_1) < x_1 x_2 > +$$
$$+(\phi_1 - \phi_3) < x_1 x_3 > +\mu(\phi_1 + \frac{1}{3}) + \phi_1\phi_2 + \phi_1\phi_3 \quad (4.12)$$
$$\frac{d < x_1 x_2 >}{dt} = (\phi_1 - \phi_2 - 2\mu) < x_1 x_2 > +\phi_2 < x_1^2 > -\phi_1 < x_2^2 > +$$
$$+\phi_1 < x_2 x_3 > -\phi_2 < x_1 x_3 > -\phi_1\phi_2$$

By symmetry, in the steady state all the diagonal terms $< x_i^2 >$ are equal, as well as off-diagonal terms (correlations) $< x_i x_j >$. They depend on the migration probability μ alone. Around the steady state the B-matrix (4.4.1) will become:

$$\begin{pmatrix} 2\mu + 2/9 & -1/9 & -1/9 \\ -1/9 & 2\mu + 2/9 & -1/9 \\ -1/9 & -1/9 & 2\mu + 2/9 \end{pmatrix}$$

The steady state solutions are:

$$< x_i^2 >= \frac{9\mu + 1}{27\mu} \quad (4.13)$$
$$< x_i x_j >= -\frac{1}{54\mu}$$

for both the cyclic and neutral systems. The diagonal terms coincide with the variance, since the average of the fluctuations $< x_i >= 0$. These fluctuations were measured "experimentally", i.e. calculated from the results of the simulations. We simulated 1000 copies of the system (size $N = 300$) at different mutation rates, and calculated the mean and variance of the fluctuations, as well as correlations. The results of those simulations are shown in Fig. 4.4 for the system with migrations. They agree with the calculated values.

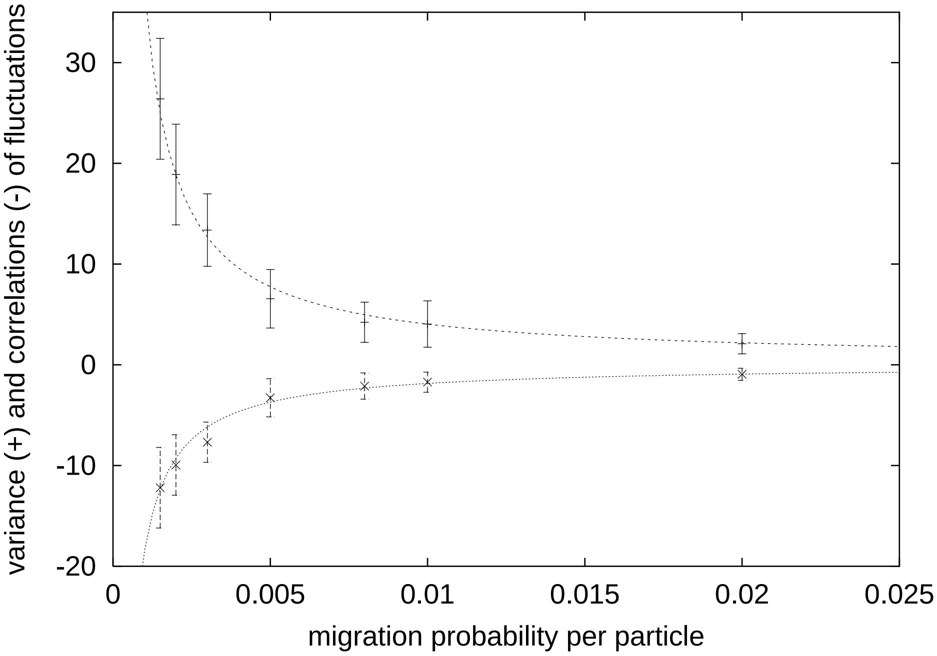

Figure 4.4: The variance and correlations of the fluctuations vs. migration probability per particle for system size $N = 300$, cyclic system. (Dashed lines show the calculated values.) The same behaviour is observed for the system with mutations.

4.4.2 The System with Mutations

In the case of the system with mutations, the total population N is a conserved quantity; this is going to have its consequences in the mathematical treatment of the system. The A-matrix for the cyclic competition system is:

$$\begin{pmatrix} \phi_3 - \phi_2 - 2\mu & \mu - \phi_1 & \phi_1 + \mu \\ \phi_2 + \mu & \phi_1 - \phi_3 - 2\mu & \mu - \phi_2 \\ \mu - \phi_3 & \phi_3 + \mu & \phi_2 - \phi_1 - 2\mu \end{pmatrix}$$

and for the neutral drift one:

$$\begin{pmatrix} -2\mu & \mu & \mu \\ \mu & -2\mu & \mu \\ \mu & \mu & -2\mu \end{pmatrix}$$

The B-matrix is the same for both systems, and it looks like:

$$\begin{pmatrix} \mu(1+\phi_1) + \phi_1\phi_2 + \phi_1\phi_3 & -\mu(1-\phi_3) - \phi_1\phi_2 & -\mu(1-\phi_2) - \phi_1\phi_3 \\ -\mu(1-\phi_3) - \phi_1\phi_2 & \mu(1+\phi_2) + \phi_1\phi_2 + \phi_2\phi_3 & -\mu(1-\phi_1) - \phi_2\phi_3 \\ -\mu(1-\phi_2) - \phi_1\phi_3 & -\mu(1-\phi_1) - \phi_2\phi_3 & \mu(1+\phi_3) + \phi_2\phi_3 + \phi_1\phi_3 \end{pmatrix}$$

Again, we limit our attention to the fluctuations around the steady state $\phi_i = 1/3$. Then the eigenvalues of the A matrix are of the form 0, $-3\mu \pm i/\sqrt{3}$, for the cyclic system with mutations, and 0, -3μ (doubly degenerate) for the neutral system with mutations. They have a zero eigenvalue, which comes from the fact that the total population size in that model is always conserved. Let us again try to determine the first and second moments of the fluctuations. The eigenvalues are zero or negative; this guarantees that, if we start with $< x_i >= 0$, the average of the fluctuations will remain zero.

Using the equations for the second moments (4.11), we can obtain the fluctuations around the steady state. Again, we can substitute $\phi_i = 1/3$ in the equations. Again, by symmetry, all the diagonal terms $< x_i^2 >$ are equal, as well as off-diagonal terms (correlations) $< x_i x_j >$. They depend on the mutation probability μ alone. These equations for the systems with mutations (cyclic or neutral) have the form:

$$\begin{aligned}\frac{d < x_i^2 >}{dt} &= -4\mu < x_i^2 > +4\mu < x_i x_j > +(\frac{4\mu}{3} + \frac{2}{9}) \\ \frac{d < x_i x_j >}{dt} &= 2\mu < x_i^2 > -2\mu < x_i x_j > -(\frac{2\mu}{3} + \frac{1}{9})\end{aligned} \tag{4.14}$$

and clearly:

$$\frac{d(< x_i^2 > +2 < x_i x_j >)}{dt} = 0 \tag{4.15}$$

which must be a result of the conservation law present in the system. It can be seen that the correlations $< x_i x_j >$ are negative, and equal to half the value of the variances $< x_i^2 >$, which could otherwise be derived from the equation $x_1 + x_2 + x_3 = 0$. We can see that the total population N's being a conserved quantity imposes restrictions on allowed fluctuations. We avoid this problem by excluding one of the variables. For that, we transform the problem into one with two variables, which can be done by putting:

$$\begin{aligned}\frac{A}{N} &= \frac{1}{3} - \frac{1}{\sqrt{N}}(\frac{\sqrt{3}}{2}x + \frac{y}{2}) \\ \frac{B}{N} &= \frac{1}{3} + \frac{1}{\sqrt{N}}(\frac{\sqrt{3}}{2}x - \frac{y}{2}) \\ \frac{C}{N} &= \frac{1}{3} + \frac{y}{\sqrt{N}}\end{aligned} \tag{4.16}$$

This corresponds to transforming to cartesian coordinates with the origin placed at the geometrical centre of the equilateral triangle, which constitutes the phase space of our system [30]. The Fokker-Planck equation for the cyclic competition system becomes:

$$\frac{\partial\Pi}{\partial t} = \frac{\partial}{\partial x}(\frac{y}{\sqrt{3}} + 3\mu x)\Pi - \frac{\partial}{\partial y}(\frac{x}{\sqrt{3}} - 3\mu y)\Pi + \frac{1+6\mu}{9}(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})\Pi \quad (4.17)$$

The A-matrix has the form:

$$\begin{pmatrix} -3\mu & -1/\sqrt{3} \\ 1/\sqrt{3} & -3\mu \end{pmatrix}$$

with eigenvalues $-3\mu \pm i/\sqrt{3}$ (the zero eigenvalue has disappeared). The mean value of the fluctuations decays to zero with an oscillating behaviour, and remains zero. The B-matrix is diagonal, with both diagonal elements equal to $\frac{2}{9}(1+6\mu)$. The equations for the second moments become:

$$\begin{aligned}
\frac{d < x^2 >}{dt} &= -6\mu < x^2 > -\frac{2}{\sqrt{3}} < xy > +\frac{2}{9}(1+6\mu) \\
\frac{d < xy >}{dt} &= \frac{2}{\sqrt{3}} < x^2 > -6\mu < xy > -\frac{2}{\sqrt{3}} < y^2 > \\
\frac{d < y^2 >}{dt} &= \frac{2}{\sqrt{3}} < xy > -6\mu < y^2 > +\frac{2}{9}(1+6\mu)
\end{aligned} \quad (4.18)$$

with solutions that tend to

$$< x^2 >=< y^2 >= \frac{1+6\mu}{27\mu}, \quad < xy >= 0 \quad (4.19)$$

as $t \to \infty$. These solutions need to be transformed back in terms of $< x_i^2 >$ and $< x_i x_j >$, and the results are:

$$\begin{aligned}
< x_i^2 > &= \frac{1+6\mu}{27\mu} \\
< x_i x_j > &= -\frac{1+6\mu}{54\mu}
\end{aligned} \quad (4.20)$$

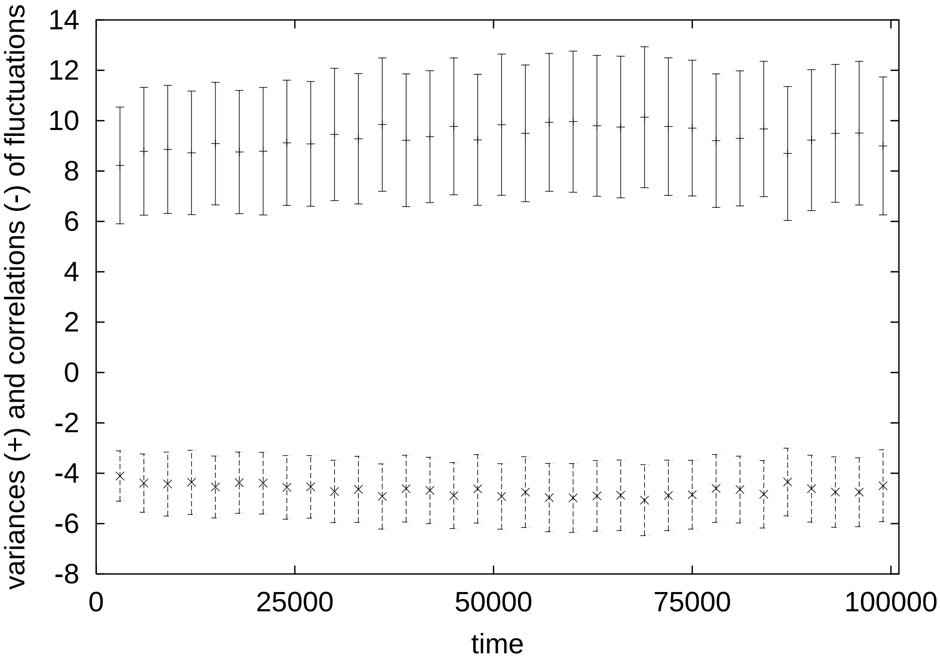

Figure 4.5: Time series of variance and correlations of the fluctuations, mutation probability per particle $\mu = 0.004$, system size $N = 450$, cyclic system.

For the neutral drift case, the Fokker-Planck equations becomes:

$$\frac{\partial \Pi}{\partial t} = \frac{\partial}{\partial x}(6\mu x \Pi) + \frac{\partial}{\partial y}(6\mu y \Pi) + \frac{2(1+6\mu)}{9}(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})\Pi \tag{4.21}$$

with the same steady state distributions as the cyclic competition model. The measured values of the fluctuations for the system(s) with mutations are shown in Fig. 4.5, where we plot their time series for $\mu = 0.004$ and $N = 450$.

Let us now go back to the order parameter, defined as ABC/N^3. If we use the expressions (4.5), the average order parameter becomes:

$$< \eta > = < \phi_1\phi_2\phi_3 > + \frac{1}{9N}(< x_1x_2 > + < x_2x_3 > + < x_1x_3 >) +$$
$$+ \frac{1}{27N^{3/2}} < x_1x_2x_3 > \tag{4.22}$$

where we have used the result $< x_i >= 0$. Substituting the expressions we obtained for the steady state (and neglecting the higher order third term; to calculate that, we would need a higher order expansion of the master equation), we get:

$$< \eta >= \frac{1}{27}(1 - \frac{1 + 6\mu}{6\mu N}) \tag{4.23}$$

which is in good agreement with the values obtained from the simulations, shown in Fig. 4.3.

4.5 The Transition Region

The qualitative change in the behaviour of the system, when the value of the parameter μ is varied, speaks of the presence of a phase transition. This transition from the "fixation" to the "diversity" regime is continuous. The graph that represents the survival probability vs. time decays exponentially at an ever slower rate as the mutation rate is increased. It exhibits a power-law decay at the critical mutation rate, after which it bends upwards in a log-log plot, as shown in Fig. 4.2. All these facts speak of a symmetry-breaking transition as the mutation probability goes through the critical value.

When the mutation/migration probabilities per particle approach zero, the leading term in the variances of the concentrations of individual populations is of order $1/27N\mu$, and it becomes of the same order of magnitude as the macroscopic concentrations, when $3\mu N \propto 1$. This gives us the critical mutation/migration probability dependence on the system size N. The critical μ was verified to be the same for both models with mutations present, and the critical μ for the system with migrations from the simulations is very close to that for the systems with mutations, which supports the calculations above.

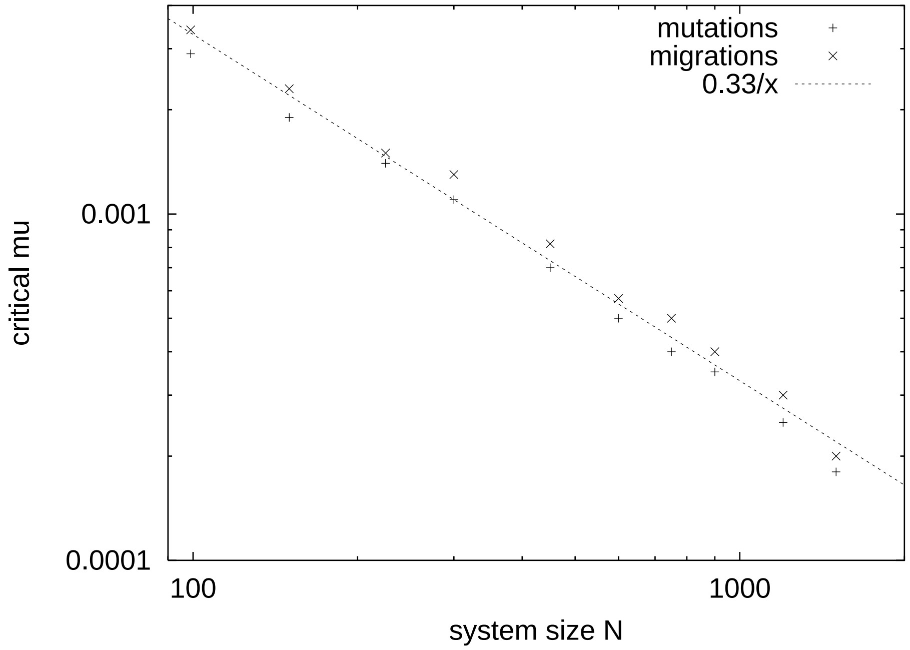

Figure 4.6: Variation of critical mutation and migration probability per particle with system size.

Fig. 4.6 shows plots of critical μ vs. N, as well as a line of slope $1/N$ for comparison. There is in excellent agreement between analytical results and those obtained from computer simulations. It is worth noting that the neutral system essentially behaves like the cyclic one.

The probability that the system has never crossed the boundary, (i.e. none of the populations has ever become zero,) when the system is in the transition regime, still decays to zero, but not exponentially any more. The transition is critical; however, transitions in nonequilibrium (steady-state) systems are different from the thermodynamic phase transitions. If the system size were infinite, one would introduce one infected individual (mutant) and measure the probability of survival of infection/mutation as time goes to infinity. This would constitute the order parameter of our system in the thermodynamic limit. However, our systems are finite, and for finite N the quantity which exhibits critical behaviour is the first passage time for crossing the boundary [1].

In Fig. 4.7 we show plots of the first extinction times (with their cumulative probability in the y-axis) just below, just above, and at the critical μ (system size $N = 210$). (It is worth mentioning that we ran a statistics of 10000 copies of the system, but we are showing only 1 in 20 points in that plot, to prevent figures from becoming cluttered.) The power-law decay behaviour was used as a criterion for determining the critical point. The plot in Fig. 4.7 shows clearly that the survival probability decays as a power law at the critical point. However, once the system is above the critical point, there is considerable probability for the system to have never "gone dead", even at very long times. The extinction times at the critical point scale with the system size. The power-law exponent is -1.03 ± 0.04 for cyclic competition model, -0.99 ± 0.03 for the neutral model with mutations; and -1.05 ± 0.06 for the cyclic system with migrations. This exponent close to 1, found in simulations is compatible with the one expected from branching processes [79], as well as the one obtained for the two-species Kimura-Weiss model [28].

4.5.1 Direct Calculation of the Critical μ

We have so far employed the linear noise approximation to calculate the fluctuations, and, by imposing the requirement that the fluctuations become of the order of the macroscopic concentrations, we have determined that the crit-

[1] Our model differs from the chemical reactions models: in them the onset of the cyclic behaviour is a Hopf bifurcation, in which a stable focus changes into a limit cycle [78]. In our model the limit cycle is absent. The chemical reactions models are dissipative even in the mean-field approximation, while our system has a centre in the mean field treatment. In those models the fluctuations become important only when the system is in the vicinity of the Hopf bifurcation, ours is entirely fluctuations-driven below the critical transition point.

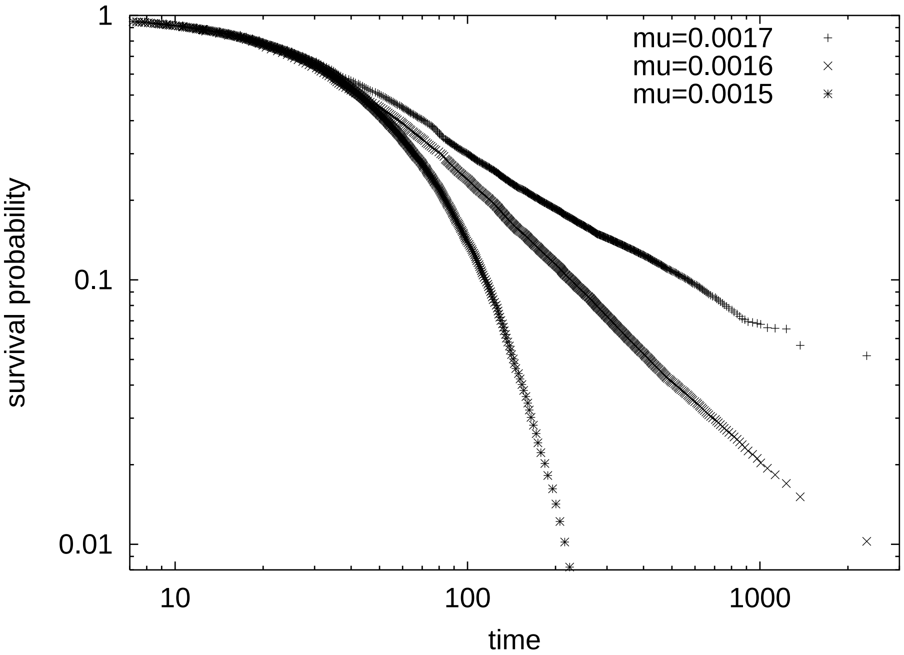

Figure 4.7: Cumulative survival probability vs. time (in units of N) just below, just above, and at the critical point, cyclic system with mutations, system size $N = 210$.

ical μ depends on system size N as N^{-1}. Hence the critical parameter can be posed as μ_0/N. Recall that the first extinction time has an exponential distribution, with a decay exponent that starts at the value -3 for the system with no mutations/migrations, and increases as μ increases, until the distribution becomes a power-law at the critical point. This suggests that we can use the Kramers-Moyal expansion (as employed in Chapter 3 and Ref. [30] for the neutral system with mutations/migrations, pose $\mu = \mu_0/N$, and impose the requirement that the smallest eigenvalue be zero, for criticality (power-law decay). This way we can obtain the value for μ_0, and compare it with the one obtained from the simulations $\mu_{0s} \simeq 0.33$.

Recall that the Kramers-Moyal expansion is an expansion in powers of N. In it we use the intensive quantities, x_A, x_B, x_C, which correspond to the concentrations of the A's, B's, and C's, as in Eq. (3.10). The shift operators now are expanded in powers of N, as in Eq. (3.11), and the probability distribution $P(A, B, C, t)$ is transformed into $W(x_A, x_B, x_C, t)$, according to (3.12). The resulting nonlinear Fokker-Planck equation for the neutral system has no N^0 term; the differential equation we need to solve is:

$$\frac{1}{N} L_{FP} W = -\lambda W \tag{4.24}$$

where the L_{FP} operator has the form

$$\begin{aligned} L_{FP} = \frac{1}{2}[(&\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_C} + \frac{\partial^2}{\partial z^2})x_A x_C + [(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_B} + \\ &+\frac{\partial^2}{\partial y^2})x_A x_B + [(\frac{\partial^2}{\partial x_B^2} - 2\frac{\partial^2}{\partial x_B \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_B x_C] + \\ &+\mu_0[(2\frac{\partial}{\partial x_A} - \frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_C})x_A + (2\frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_A} - \\ &-\frac{\partial}{\partial x_C})x_B + (2\frac{\partial}{\partial x_C} - \frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_A})x_C] \end{aligned} \tag{4.25}$$

for the system with mutations, and

$$L_{FP} = \frac{1}{2}[(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_A x_C + [(\frac{\partial^2}{\partial x_A^2} - 2\frac{\partial^2}{\partial x_A \partial x_B} +$$

$$+\frac{\partial^2}{\partial y^2})x_A x_B + [(\frac{\partial^2}{\partial x_B^2} - 2\frac{\partial^2}{\partial x_B \partial x_C} + \frac{\partial^2}{\partial x_C^2})x_B x_C] +$$

$$+\mu_0[3\frac{\partial}{\partial x_A}x_A + 3\frac{\partial}{\partial x_B}x_B + 3\frac{\partial}{\partial x_C}x_C - \frac{\partial}{\partial x_A} - \frac{\partial}{\partial x_B} - \frac{\partial}{\partial x_C}] \tag{4.26}$$

for the system with migrations. These equations accept solutions of the form:

$$W(x_A, x_B, x_C, t) = e^{-\lambda t} W(x_A, x_B, x_C) \tag{4.27}$$

since L_{FP} does not depend on time. The N from the left hand side denominator can be absorbed into the exponential, resulting in time-rescaling (instead of time t, the time variable is now t/N). We already have a complete set of eigenfunctions for the symmetry of the problem at hand, and those eigenfunctions satisfy the same boundary conditions as the solution W [63, 64]. The additional requirement is that λ be zero, so that the decay obeys a power law. The Galerkin method gives the values for μ_0: they will be obtained from the equation $\lambda \simeq 9 \cdot \mu_0 - 3$, for both the system with mutations and that with migrations. Then the critical regime $\lambda = 0$ is obtained for $\mu_0 \simeq 1/3$. This value agrees very well with the simulations results.

4.6 Conclusions

In this chapter we have considered an ABC model with cyclic competition/neutral drift and mutations (migrations) at a constant probability. The system exhibits a critical transition from a “fixation” regime to one in which biodiversity is preserved over long time. In the “fixation” regime, the number of the

A, B, C species in the cyclic competition version oscillates with an amplitude that drifts with time, while in the neutral drift version the populations numbers drift, until one of the species (and then the second one) goes extinct, i.e. the order parameter, defined as the product of ABC/N^3 goes to zero, and remains zero, except for occasional "bursts". In the "diversity" regime, the number of the A, B, C varieties exhibits Gaussian fluctuations around the centre point, and there are rare extinctions, but the order parameter remains nonzero almost always. The survival probability decays exponentially below the transition point, but the exponent decreases as the mutation (migration) probability per particle increases, until it becomes zero at the critical point. The critical mutation/migration probability depends on system size as N^{-1}, and the models have the same power-law exponent: -1. There is no difference in the behaviours of the neutral system and the cyclic system. Also, there is no qualitative difference between the system with mutations and that with migrations. These simulations results have been verified by expanding the master equation into a nonlinear Fokker-Planck equation, and solving it, to get critical $\mu = \mu_0/N = 1/3N$.

The results we obtained address the concern about the effects of habitat fragmentation on the survival of the species [66]. If the population of a certain species goes extinct in one patch (e.g. a herd, school, swarm) while it still survives in other patches, then it is hoped that the "rescue effect" can prevent global extinction [67–69]. Otherwise, the species is doomed to go extinct altogether. Our results suggest that the number of mutants/migrants (i.e. the product μN) necessary to preserve diversity is independent of system size. However, in most realistic situations the number of mutations that happen would be proportional to system size. This means that a large system size favours the maintenance of biodiversity. On the other hand, if one has in mind epidemiological systems, the important factor is to reduce migration probabilities below critical, which is exactly the purpose quarantines serve. Let's recall here the revival of SARS epidemics in Toronto, once their guard was let down, i.e. infected individuals were allowed to migrate from one community on to others. In that case, a large system size may be a disadvantage. It takes only one bad apple...

The study of three-species ecological or epidemiological systems relates to autocatalytic systems with a loop-like structure [82,100]. Methods of analysis employed here can be expanded to the loop-like autocatalytic systems. We report the results of our work regarding the four- or more-species autocatalytic loops [83] in the next chapter.

Chapter 5

Autocatalytic Systems: Behaviour and Transitions

5.1 Introduction

As we have already discussed, in many ecological, political, epidemiological, economic, chemical, reaction-diffusion, and biological systems, competition plays a very important role. An important sub-class of those is the cyclic competition systems. These systems were our motivation for the models introduced and solved in Chapters 3 and 4. Goodwin [84] introduced a system of interacting biochemical metabolic oscillators, which has an autocatalytic feedback mechanism. Biochemical reactions in a cell support its activities, hence assuring its very existence. Autocatalytic reactions are an important class of reactions within a cell. They are reputed to have made possible the birth and existence of life itself.

The simplest generic example of an autocatalytic reaction is the loop of the type $A_i + A_{i+1} \to 2A_{i+1}$, where $i = 1, \dots, k$; $A_{k+1} = A_1$. The molecules are in a well-stirred container (the cell), which is in contact with a reservoir (the outside environment). They can diffuse into and out of the container, to and from the reservoir. In another (ecological) context, the $\{A_i\}$'s are versions of a biological species, and in the epidemiological one, states of an individual (e.g. susceptible, infected, etc.) In ecological systems it makes sense to study the neutral version of this model, in which $A_i + A_{i+1} \to 2A_{i+1}$ or $2A_i$ with equal probability, corresponding to the Kimura model of neutral genetic drift [17,18].

In two previous works [30, 31] we have considered an ABC model with cyclic competition/neutral drift and mutations (migrations) at a constant probability. The system exhibits a critical transition from a "fixation" regime to one in which biodiversity is preserved over long time. In this chapter we study the system with four or more species, and show that the above described picture holds. We show that the cyclic system is a generalisation of the well-known Lotka-Volterra system, and that the size-induced transition is present in those systems as well.

5.2 The Model

Our system is an autocatalytic loop of the type $A_i + A_{i+1} \to 2A_{i+1}$, where $i = 1, \ldots, k; A_{k+1} = A_1$. The molecules are in a well-stirred container (the cell), which is in contact with a reservoir (the outside environment). They can diffuse into and out of the container, according to the following: a molecule (individual) of species i leaves the container at a rate $D \cdot a_i$, and enters it at a rate $D \cdot s_i$, where s_i is its concentration in the reservoir. (We will assume that the reservoir is large enough, so that the exchanges do not perturb it.) The rate equations then read:

$$\frac{da_i}{dt} = a_{i-1}a_i - a_i a_{i+1} + D(s_i - a_i) \tag{5.1}$$

In the rate equations approximation, the system size is conserved. The above equations (5.1) have a fixed point, and it is a stable solution.

5.2.1 Lotka-Volterra Model

The Lotka-Volterra [85, 86] model of interacting populations was originally introduced by Vito Volterra for the description of competition among species. A very good and pleasant to read review of the model has been written by Goel, Maitra, and Montroll [87].

Volterra [86,88,89] got motivated by discussions with D'Ancona [90,91], who was analysing statistics of fish catches in the Adriatic. The catches of two common species of fish varied with the same period, but a little out of phase. The larger fish (species 2) ate the smaller ones (species 1), and multiplied until the species of the smaller ones diminished to very low levels, to prohibit the survival of the larger ones. With less larger fish present, the population of the smaller fish grew, hence a larger number of larger fish could be supported, etc. Volterra described the situation by the pair of equations:

$$\begin{aligned} \frac{dN_1}{dt} &= \alpha_1 N_1 - \lambda_1 N_1 N_2 \\ \frac{dN_2}{dt} &= -\alpha_2 N_2 + \lambda_2 N_1 N_2 \end{aligned} \tag{5.2}$$

The terms of the form $\lambda_i N_1 N_2$ describe respectively the depletion of the stock of fish 1 and the enrichment of that of fish 2 from eating fish 1, and those of the form $\alpha_i N_i$ the evolution of the species i in absence of the other (in absence of the predator, the population of the prey would grow exponentially, while in absence of the prey, the predators would die of starvation). Lotka had investigated these equations in the theory of autocatalytical chemical reactions [92], as well as that of competing species [85, 93]. Volterra generalized the pair of equations (5.2) to n species:

$$\frac{dN_i}{dt} = k_i N_i + \beta_i^{-1} \sum_{j=1}^{n} a_{ij} N_i N_j \tag{5.3}$$

In absence of other species, the i-th species will grow or die exponentially, depending on the sign of k_i. The constants a_{ij} will be positive, if species i eats j, negative, if it is eaten by j, and zero, if they do not interact at all. This leads to $a_{ij} = -a_{ji}$.

The mean-field (rate equations) behaviour of the system is well-known. It accepts periodic solutions, as shown by Volterra. Statistical mechanics of the Volterra system was first constructed by Kerner [94]. Goel et al. [87] introduced a random function of time (noise) in the growth equation, obtaining a Liouville equation. The model is simple enough to play for the social and biological systems the role the harmonic oscillator or Ising model play in theoretical physics. It is indeed one of the simplest of nonlinear competition models.

5.2.2 Graph Theory Solutions

This paragraph presents some results from an article by Jain and Krishna [82], applicable to the solutions of the rate equations for the sets we are interested in. Autocatalytic sets can be described as directed graphs, i.e. a set of 'nodes' and 'links', where each link is an ordered pair of nodes [95, 96]. A graph with p nodes is specified by its adjacency matrix C, defined in such a way that the $c_{i,j}$ element is one, if the graph contains a link directed from node j to node i and zero otherwise. Then a four-species cyclic system shown in Fig. 5.1 (a)

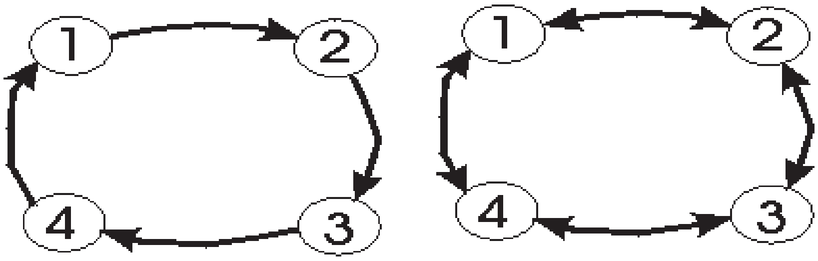

Figure 5.1: The graph describing the autocatalytic loop in the (a) case of the cyclic system, and (b) neutral one.

will be described by the matrix:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

and its neutral counterpart, shown in Fig. 5.1 (b) by the matrix:

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

In other words, the matrix for the neutral system will be the sum of the matrix for the cyclic system and its transpose. For any non-negative matrix, the Perron-Frobenius theorem [97, 98] states that there exists an eigenvalue which is real and larger than or equal to all the other eigenvalues in magnitude. It has been shown [99] that when a graph has a closed loop, the Perron-Frobenius eigenvalue of its adjacency matrix is exactly 1. The set of Perron-Frobenius eigenvectors is of interest to us, since it provides the attractors for the dynamical system whose evolution is described by the set of differential equations:

$$\dot{a_i} = \sum_{i=1}^{s} c_{i,j} a_j - \phi a_i \tag{5.4}$$

which, as shown by Jain and Krishna [82], is equivalent to the generalized Volterra equation:

$$\dot{a_i} = \sum_{i} c_{i,j} a_j - a_i \sum_{k,j} c_{kj} a_j \tag{5.5}$$

when the catalysed reaction is much faster than the one described by the first term (the spontaneous one). The system will then converge towards the fixed point that is a Perron-Frobenius eigenvector of the graph's adjacency matrix. In this approach, all the rates are 1, since the elements of the adjacency matrix can be one or zero. In the case of our cyclic systems, the Perron-Frobenius eigenvalue is always one, and the components of the corresponding eigenvector are all equal. This means that, in the rate equations approximation, the system will approach the centre (all a_i's are equal), and remain there forever.

We would, however, want to study the system beyond the rate equations approximation. So, we take into account the birth and death processes (intrinsic noise), which we take to be Poisson-distributed. The master equation for, say, the four-species cyclic system is:

$$\frac{\partial P(A_1, A_2, A_3, A_4, t)}{\partial t} = \{\frac{1}{N}[(\epsilon_{A_4}\epsilon_{A_1}^{-1} - 1)A_1 A_4 + (\epsilon_{A_1}\epsilon_{A_2}^{-1} - 1)A_1 A_2 +$$

$$+(\epsilon_{A_2}\epsilon_{A_3}^{-1} - 1)A_2 A_3 + (\epsilon_{A_3}\epsilon_{A_4}^{-1} - 1)A_3 A_4] P(A_1, A_2, A_3, A_4, t) \tag{5.6}$$

where we have used the "shift" operators notation:

$$\epsilon_{A_1} f(A_1, A_2, A_3, A_4) = f(A_1 + 1, A_2, A_3, A_4)$$
$$\epsilon_{A_1}^{-1} f(A_1, A_2, A_3, A_4) = f(A_1 - 1, A_2, A_3, A_4) \tag{5.7}$$

5.3 System-Size Expansion of the Master Equation

In a previous study [31] we have shown that for systems like the one above, where the rate equations have a stable solution, the Ω-expansion of van Kampen [35] works exceptionally well. Using the "shift" operators notation (5.7), the master equation for the cyclic competition system with migrations reads:

$$\frac{\partial P(A_1,A_2,A_3,A_4,t)}{\partial t} = \{\frac{1}{N}[(\epsilon_{A_1}\epsilon_{A_1}^{-1}-1)A_1A_4+(\epsilon_{A_1}\epsilon_{A_2}^{-1}-1)A_1A_2+(\epsilon_{A_2}\epsilon_{A_3}^{-1}-1)A_2A_3+$$
$$+(\epsilon_{A_3}\epsilon_{A_4}^{-1}-1)A_3A_4]+D[(\epsilon_{A_1}-1)A_1+(\epsilon_{A_2}-1)A_2+(\epsilon_{A_3}-1)A_3+$$
$$+(\epsilon_{A_4}-1)A_4+\frac{N}{4}(\epsilon_{A_1}^{-1}+\epsilon_{A_2}^{-1}+\epsilon_{A_3}^{-1}+\epsilon_{A_4}^{-1}-4)]\}P(A_1,A_2,A_3,A_4,t) \quad (5.8)$$

The idea of the van Kampen expansion [35] is to split the variables of the problem into a non-fluctuating part, and a fluctuating one, i.e. to deal separately with the mean-field solutions and the fluctuations (which are taken to be of the order $\sqrt{N}$). In this approach, the numbers of the individual populations would be written similar to Eq. (4.5):

$$\begin{aligned} A_1 &= N\phi_1+\sqrt{N}x_1 \\ A_2 &= N\phi_2+\sqrt{N}x_2 \\ A_3 &= N\phi_3+\sqrt{N}x_3 \\ A_4 &= N\phi_4+\sqrt{N}x_4 \end{aligned} \quad (5.9)$$

Here the ϕ_i are again the steady-state (non-fluctuating) concentrations of the i-th species respectively, and the x_i are the fluctuations (proportional to the square root of the system size). Then the probability distribution $P(A_i,t)$ is transformed into $\Pi(\{x_i\},t)$, and relations similar to (4.6) will apply:

$$\Pi = N^2 P(N\{\phi_i + \sqrt{N}x_i\}, t) \tag{5.10}$$
$$\frac{\partial P}{\partial t} = \frac{1}{N^2}\frac{\partial \Pi}{\partial t} - \frac{1}{N}\sum \frac{d\phi_i}{dt}\frac{\partial \Pi}{\partial x_i}$$

Again we substitute everything into the master equation, leave only the term $\partial\Pi/\partial t$ on the left hand side, and group the right hand side terms according to powers of $\sqrt{N}$. The first term is of order $N^{1/2}$, and it must be equal to zero, for an expansion in terms of $N^{1/2}$ to make sense. That term reproduces the rate equations in terms of the concentrations ϕ_i, with steady state solution $\phi_i = 1/4$.

The terms of order N^0 give a linear Fokker-Planck equation of the form (4.9), where the A-matrix for the cyclic system is:

$$\begin{pmatrix} \phi_4 - \phi_2 - D & -\phi_1 & 0 & \phi_1 \\ \phi_2 & \phi_1 - \phi_3 - D & -\phi_2 & 0 \\ 0 & -\phi_3 & \phi_2 - \phi_4 - D & -\phi_3 \\ -\phi_4 & 0 & \phi_4 & \phi_3 - \phi_1 - D \end{pmatrix}$$

and for the neutral system:

$$\begin{pmatrix} -D & 0 & 0 & 0 \\ 0 & -D & 0 & 0 \\ 0 & 0 & -D & 0 \\ 0 & 0 & 0 & -D \end{pmatrix}$$

The B-matrix is the same for both systems. Its diagonal elements are $B_{ii} = D(s_i + \phi_i) + \phi_i(\phi_{i-1} + \phi_{i+1})$, and the off-diagonal ones: $B_{ij} = -\phi_i\phi_j$. The solution of the resulting Fokker-Planck equation is known to be a Gaussian. Here we can use Eqs. (4.10) and (4.11) to determine the first and second moments of the fluctuations. In the special case that all the coefficients $s_i = s$

are the same, all ϕ_i are equal to one-another ($= \phi$). The eigenvalues of the A matrix are $-D$ (doubly-degenerate), $-D \pm i\phi\sqrt{2}$ for the cyclic system, and $-D$ (quadruply degenerate) for the neutral system. The negativity of the eigenvalues guarantees the stability of the zero solutions to the first moments equations. By symmetry, all the diagonal terms $< x_i^2 >$ are equal, as well as off-diagonal terms (correlations) $< x_i x_j >$. They depend on the migration probability D alone. The steady state solutions for the diagonal terms (and also for the variances, since the mean values are zero,) are as follows:

$$< x_i^2 >= \frac{(2\phi^2 + D(\phi + s))}{2D} \tag{5.11}$$

It is important to point out that from the expression for the A-matrix above, we can tell that when migrations are absent ($D = 0$), the (real part of all) eigenvalues of that matrix become zero, and the equation is of the diffusion type [35]. In that case, the rate equations suggest that the non-fluctuating part will remain constant (at the centre). However, small deviations give rise to large differences, and one would expect the fluctuations to grow, rather than remain limited. The separation into a macroscopic part and small fluctuations is no longer meaningful. In this case, after a transient period (of the order $N^{1/2}$), one would expect P to be a smooth function of the concentrations, and expand in powers of N (Kramers-Moyal), rather than $N^{1/2}$.

Furthermore, the only difference between the three and four species systems expansion is the appearance of an "off-diagonal" of zeroes, in positions (i, j) for which the species A_i and A_j do not react. This means that the above algebra will remain exactly valid when there are more than four species in the system. Hence, our results will hold for any number of species, and any reasonable system size.

If there are only migrations into and out of the container (i.e. no cyclic reactions), the system remains near the centre point, and the product of the concentrations remains considerably above zero; in other words, all species are present in the system at (almost) all times. When both the cyclic/neutral mechanism and migrations are present, one can occasionally observe temporary extinctions. Since the boundary is not absorbing, occasional migrations will return the system to the state with maximal symmetry (diversity) where all species coexist. The migrations then manage to keep the system maximally disordered, since they are stronger than the fluctuations (which try to drive the system toward the boundary, i.e. fixation, and keep it there). The migration rate acts then as some sort of "temperature", and decreasing the diffusion rate would be analogous to annealing the system.

5.4 The Transition Region

In two recent studies, Togashi and Kaneko [100, 101] focus on the four-species autocatalytic system with a very small number of molecules. Their work covers aspects of a transition, which they baptize "discreteness-induced". The idea of their work is that, when the number of molecules in the system is very small, a transition from the state where the numbers of molecules of the individual types are Gaussian-distributed, as derived above, to one in which the A_1 and A_3 species form an alliance against species A_2 and A_4 is observed. In their work they keep the diffusion rate D constant, and vary the system size. As the total number of molecules (system size) decreases and goes through a certain value (which coincides with $1/D$, the inverse of diffusion rate), they observe a transition into the broken-symmetry state. One of the symbiotic pairs eventually wins, but, since there are migrations into and out of the container happening all the time, the induction of a member of the opposite symbiotic pair may bring the victory of that initially (almost) non-existent pair, making the pair that previously was the "winner" disappear, and so on. The probability distribution of the number $z = (x_1 + x_3) - (x_2 + x_4)$ (corresponding to our $(a_1+a_3)-(a_2+a_4)$) is shown in Fig. 2 of their article [100], and the formation of the peaks that indicate the victory of one such symbiotic pair over the other as the system size goes through $1/D$ is quite pronounced. In Fig. 4 of their PRL article [100] Togashi and Kaneko show a plot of the rate of residence of the 1-3 (or 2-4) rich state (i.e. the state in which one of the pairs "rules") as a function of the product DV (V is system size, the parameter we call N). The rate of residence of the symmetry-broken state clearly goes to zero, as $DV \to 1$.

Looking at the expression for the variance of the fluctuations above (5.11), one can observe that when the migration probabilities per particle (diffusion rate) approach zero, the variance of the concentrations of individual populations are of the order $\frac{(2\phi^2+D(\phi+s))}{2D}$, and it becomes of order 1 (i.e. the order of macroscopic concentrations,) when $D \sim 1/N$ (here we are using the values of parameters as chosen by Togashi and Kaneko [100], who assume all $\phi_i = s_i = 1$). This gives us the critical value for the migration probability. The critical D thus obtained is in excellent agreement with the one Togashi and Kaneko observe and show in Fig. 2 of their letter [100], when they see the system go through a transition for $V(N) = 1/D$. Also, this verifies the other result of their simulations, shown in Fig. 4 of their letter [100], where the rate of residence of the symmetry-broken state approaches zero, as $DV \to 1$. In our work on three-species systems in Chapter 4 we obtained similar results using a van Kampen expansion, and verified them in the diffusive (fixation) regime, using a Kramers-Moyal expansion, and imposing the condition that the system be near a critical transition, i.e. the smallest eigenvalue of the Fokker-Planck equation be zero. The agreement between the two is excellent.

5.5 Steady-State Solutions of the Rate Equations and Simulations Results in the Fixation Regime

It is important to know the behaviour of the system, when it is closed, i.e. no migrations into and out of the container are possible, since the final state in the fixation (broken-symmetry) regime is the same. The rate equations for this system, written for the concentrations of individual species, will be:

$$\frac{da_i}{dt} = a_{i-1}a_i - a_i a_{i+1} \tag{5.12}$$

and similarly for the other species. If we consider the case when $A_i + A_{i+1} \rightarrow 2A_i$ or $2A_{i+1}$ with equal probability, the right-hand side of the rate equations will be identically zero.

One can directly find the steady-state solutions of the rate equations (5.12), i.e. the concentrations that reduce the left-hand side to zero. For any number of species in the system, the centre, where all the concentrations are equal, is a solution of the rate equations (5.12). When three species are present, the other solutions are those for which one of the species is alive, and the others are extinct. In the case of four and five species, the solutions are of the form $a_i = a_{i+2}$ (modulo 4 or 5) with all the other concentrations equal to zero. The system with six species, except for the centre, has two other kinds of sets of solutions: one with $a_i = a_{i+2} = a_{i+4}$ (modulo 6) and all the others zero (i.e.

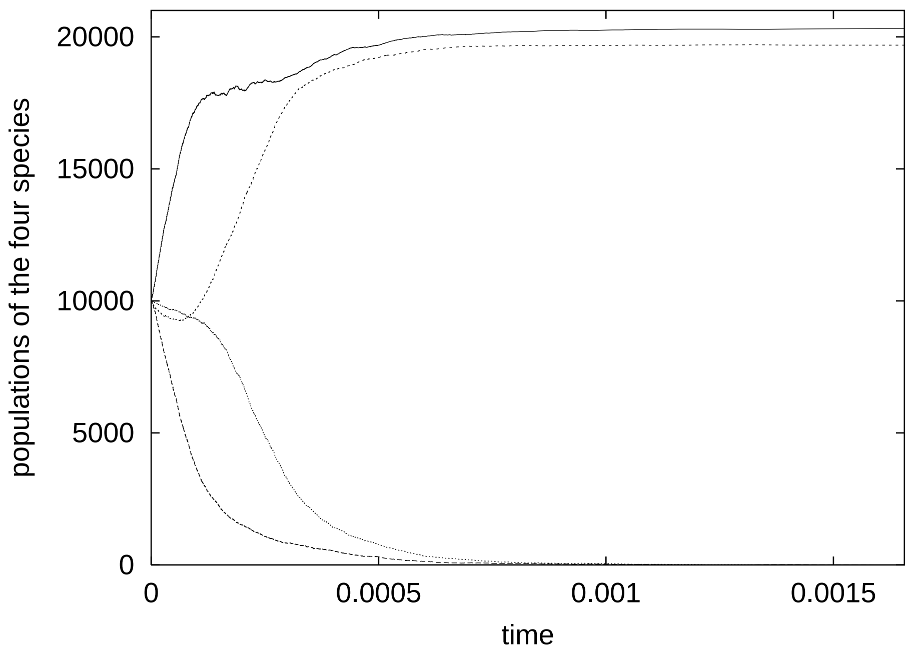

Figure 5.2: The time series for the number of A_1, A_2, A_3, and A_4 species in the "fixation" regime (here $D = 0$, and system size $N = 40\ 000$). In this particular realization the A_1A_3 symbiotic pair wins over the A_2A_4 one.

three species alive and three dead), and one set with $a_i = a_{i+3}$ (modulo 6) with the rest of concentrations equal to zero (i.e. two "opposite" species alive and four dead). For any numbers of species in the system, the product of concentrations is conserved, as well as the total number of individuals. This means that any trajectory with the product of concentrations constant will be a neutrally stable one. In the case of the neutral systems, the right-hand side is identically zero, which means that any state is a stable one.

We simulated copies of the four- or more-species cyclic system, for different system sizes. The simulations started from the symmetric state, with equal individual populations of A_1, A_2, A_3, and A_4. We observed the symmetry-breaking phenomenon for any system size, no matter how large it is.

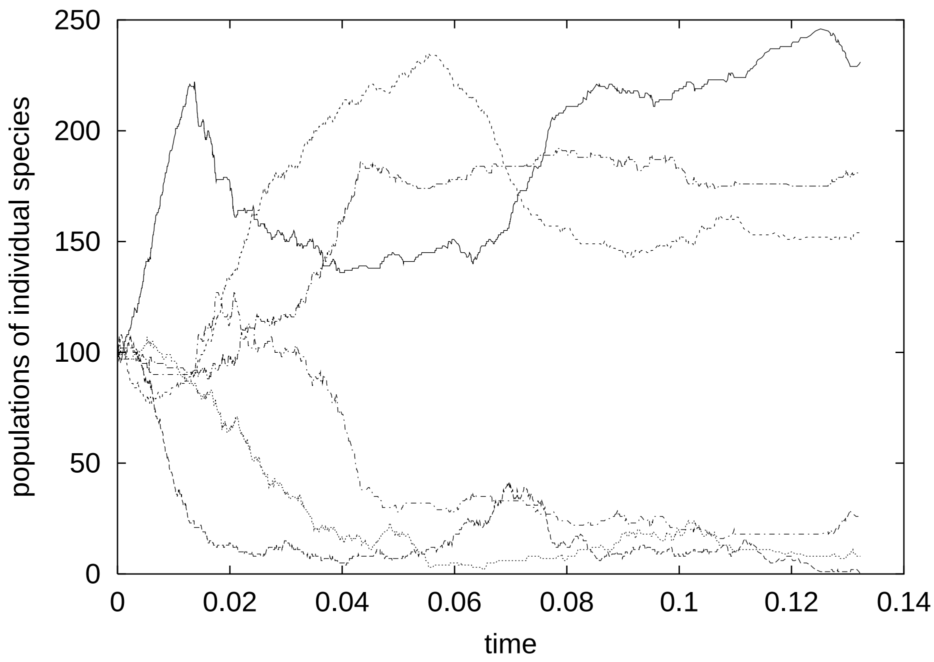

Figure 5.3: The time series for the six-species system ($N = 600$). All the copies result in a state in which three species (rather than two) survive.

In Fig. 5.2 we show the time series for the four-species cyclic system, for $N = 40\ 000$. Contrary to the prediction from the solutions to the rate equations, for the six-species systems, out of two thousand realisations of the system, we never obtained a final state with only two species present. All the copies of the system we simulated ended up with three species present, as the time series shown in Fig. 5.3. In all the time series, time is measured in units of the system size N.

When more species are present in the system, the picture gets more complicated. However, the general feature is that the system experiences "fluctuations-generated forces", which push the system towards the boundary. There the history of the system ends, since the boundary is absorbing. When the number of species is odd, the population sizes of individual species

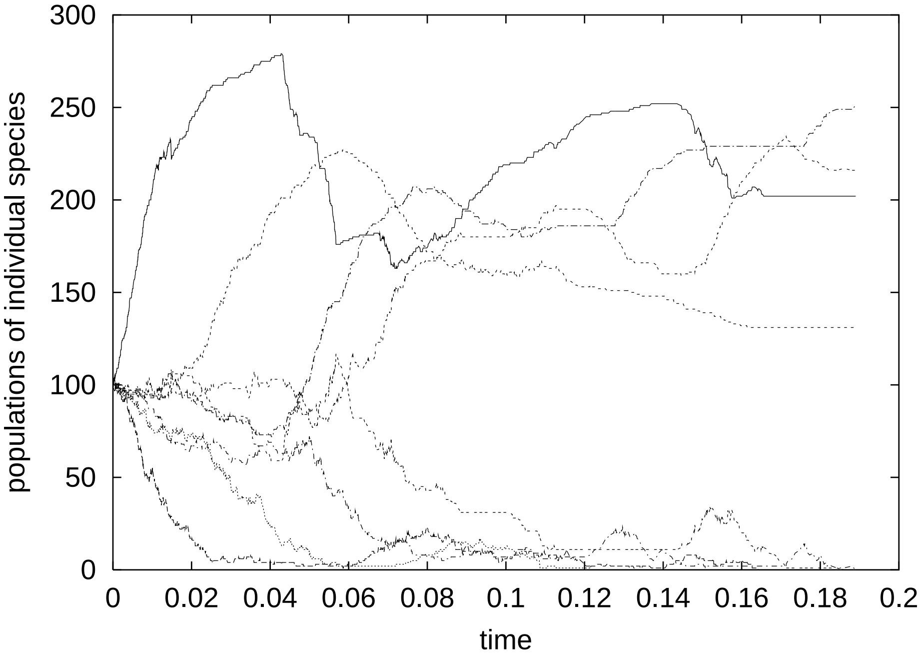

Figure 5.4: The time series for the eight-species system ($N = 800$). The history of the system ends in a state in which four species survive.

oscillate with an amplitude that drifts with time (as in the three-species system), until they reach the boundary, one after another. On the other hand, when the number of species is even, "alliances" do not take long to form, and the history of the system always ends with half of the species winning over the other half. A typical time series for an eight species system is shown in Fig. 5.4. The fate of the system remains similar throughout the "fixation" regime, i.e. when the diffusion rate $D < 1/N$. In that regime, the system exhibits diffusive behaviour toward the boundary, and the migrations in and out are not frequent enough to bring it back to the centre.

The fluctuations push the neutral system towards the boundary, too, but the history of the neutral systems ends either with only one species present, or with disconnected species, in the sense that they have no reason to coexist in symbiosis, since they have no common enemies. On the other hand, the history of the cyclic systems ends with two or more surviving species present, which do not "react" with one-another directly, but have common "enemies". This is quite reminiscent of symbiosis (an alga + a fungus = a lichen, or a tree and mushrooms live together and help each-other fight common enemies), and yet another example of competition giving rise to cooperation.

5.6 The Case of Generalized Volterra Equations

Now let us go back to the Volterra equations (5.3). They can be generalised further, resulting in the following rate equations:

$$\frac{dN_i}{dt} = \sum_j k_{ij} N_j + \frac{1}{N} \sum_{j=1}^{n} a_{ij} N_i N_j \tag{5.13}$$

The diagonal elements in both sums correspond to a Malthus-Verhulst equation. The first sum now contains off-diagonal elements, which correspond to mutations of individuals from species i to species j (or the other way around, depending on the sign of the rate k_{ij}. These mutation rates will be very small.) Hence, the k_{ij}'s and k_{ji} have opposite signs (and also a_{ij} and a_{ji}). The terms of the form $a_{ij} N_i N_j$ describe what happens when an i individual runs into a j individual: when $a_{ij} > 0$ the i individual "eats" a j individual and reproduces, and the other way around when $a_{ij} < 0$. The master equation for the generalised Volterra system, written in terms of the "shift" operators (5.7):

$$\frac{\partial P(\{N_i\}, t)}{\partial t} = \sum_i \{\beta_{ii}(\epsilon_i^{-1} - 1)N_i + \sum_{j \neq i} \beta_{ij}(\epsilon_j \epsilon_i^{-1} - 1)N_j +$$
$$+ \frac{\alpha_{ii}}{N}(\epsilon_i^{-1} - 1)N_i^2 + \sum_{j \neq i} \frac{\alpha_{ij}}{N}(\epsilon_j \epsilon_i^{-1} - 1)N_i N_j\} \tag{5.14}$$

where the way the master equation is written imposes that the coefficients β_{ij} be chosen as equal to k_{ij} if $k_{ij} > 0$, and zero otherwise. Similarly, $\alpha_{ij} = a_{ij}$ if $a_{ij} > 0$, and zero otherwise.

Now we are ready to attempt the van Kampen expansion of the above master equation. As before, the variables N_i are split into a non-fluctuating and a fluctuating part, as in Eqs. (5.9), which transforms the probability density function similarly to (5.10). After substituting everything into the master equation, we obtain a Fokker-Planck equation as in (4.9), where the A and B matrices have dimensions $n \times n$. The diagonal elements of the A-matrix will be given by $A_{ii} = \beta_{ii} - \sum_{j \neq i} \beta_{ji} + \sum_j (\alpha_{ij} - \alpha_{ji})\phi_j$, and the off-diagonal ones by $A_{ij} = \beta_{ij} - (\alpha_{ij} - \alpha_{ji})\phi_i$. The existence and stability of the solutions to the Volterra equations (5.3) have been studied by Goel et al. [87]. When the system is stable, the A-matrix will have negative eigenvalues, which, using the equation (4.10), will mean that the average value of the fluctuations will decay to zero. The B-matrix has diagonal elements of the form $B_{ii} = \beta_{ii}\phi_i + \sum_{j \neq i}(\beta_{ij}\phi_j + \beta_{ji}\phi_i) + \alpha_{ii}\phi_i^2 + \sum_{j \neq i}(\alpha_{ij} + \alpha_{ji})\phi_i\phi_j$ and off-diagonal ones: $B_{ij} = -\beta_{ij}\phi_j - \beta_{ji}\phi_i - (\alpha_{ij} + \alpha_{ji})\phi_i\phi_j$. They are linear in coefficients β_{ij} and α_{ij}, which means that the equations for the second moments of the fluctuations (obtained from their general expression (4.11)) will also be linear in the rate coefficients. This will lead to a size-induced transition as the one calculated for the autocatalytic loops, and observed in simulations by us for the three-species systems [31], and Togashi and Kaneko for four-species systems [100, 101].

5.7 Conclusions

The transition observed by Togashi and Kaneko [100, 101] in the four-species autocatalytic loops is not specific to systems with a very small number of particles/individuals. It is the same critical transition we have previously [31] observed for three-species systems. This transition corresponds to a crossover from a fully symmetric ("neutral") state in the high-migration regime, to a "fixation" state for low-migration rates, in which the symmetry is broken in favour of one or more species. The "fixation" regime in the four-species system exhibits a symbiotic effect, when species A_1 and A_3 join their efforts against species A_2 and A_4, and the final state of the system is one in which one of the pairs has completely "eaten up" the other. This transition is present for any finite system size. In the high-migration regime the system allows for a linear noise approximation, exhibiting itself as an Ornstein-Uhlenbeck process.

Since the system size is always finite (no matter how large it is), there is a value of the migration probability per particle (or diffusion rate) for which the fluctuations of the concentrations become of order one, and the system undergoes a critical transition. The critical diffusion rate varies with system size as $1/N$, and the product $DN \sim 1$, i.e. the number of migrants per unit time, necessary for the symmetry to be preserved in the system (all species to survive) is of the order 1. This result is a bit counterintuitive, since it does not depend on system size. The analytical calculations are in excellent agreement with the simulation results, obtained for three- and four-species systems [31, 100]. Those analytical calculations suggest that all the loop-like autocatalytic systems will exhibit the same critical transition. The form of the equations for the moments of the fluctuations suggests that the generalised Lotka-Volterra systems will also exhibit a similar transition.

This system, known as a diffusion-limited reaction, has manifested itself and been observed in physical systems of low dimensions, when diffusion is not efficient in mixing the reactants. A physical example is the Ovchinnikov-Zeldovich segregation phenomenon [102].

Chapter 6

Why Do Computer Viruses Survive?

6.1 Introduction

So far we have considered a non-spatial version of three-species models in their cyclic competition and neutral genetic drift versions, and studied the long-term behaviour of such a model. When no mutations or migrations are allowed, the final state of the system is that with one "victorious" species, while the other ones have gone extinct. The Fokker-Planck equation for the system has a diffusive character, and the number of survivors vs. time plots show an exponential decay. When mutations or migrations are allowed, three regimes are observed: one in which there is fixation, as described above; one regime in which the population sizes have a Gaussian distribution around the centre (where each species' populations are equal); and a critical regime, where the mutation/migration probability per unit time per particle is inversely proportional to system size, or, in other words, the number of mutations or migrations per unit time is of the order 1 [31]. This generic pattern is preserved for many-species systems, but now in the cyclic version, in the “fixation” regime, the final state is one where symbiotic pairs appear. The species in these symbiotic ensembles help one-another fight common “enemies” [83]. The mathematical formalism is exactly the same as the one that describes autocatalytic loops, whose behaviour has been observed to follow the above described picture [100]. The picture thus obtained, holds for a generalized Volterra system, as shown by analytical calculations [83].

Wright was the first to quantitatively examine the effects of genetic drift for spatially structured populations with his ‘Island’ model. In this model time is discrete, and migration happens between an infinite number of habitat patches, which have an internal dynamics described by the Wright-Fisher model. Kimura and Weiss [26, 27] introduced the “Stepping-Stone” model, in which the patches are arranged on a lattice, and there is migration between the neighbouring patches.

Many studies of the cyclic competition system, or cyclic Lotka-Volterra model, put the ensemble on a lattice. In one dimension, spatial inhomogeneities develop spontaneously in initially homogeneous systems. There is fixation, if the number of species is greater than or equal to 5; and the state of every site changes at arbitrarily large times, if there are less than 5 species present [12, 103, 104]. For predator-prey lattice systems, a crossover between a pure prey phase and a coexistence phase is observed [105, 106], reminiscent of the transition between fixation and neutral regimes in our islands, or the Ovchinnikov-Zeldovich segregation phenomenon. In the case of a cyclic system in a two-dimensional lattice, Frachebourg and Krapivsky [104] have shown that the system tends toward an inhomogeneous state, if the number of species is less than 14. Domains of different species are reported by Szabó et al. [10, 11, 107]. Recently, Kerr et al. obtained laboratory results for a system of three competing species that play rock-paper-scissors, when placed in a lattice-like spatial structure, and showed that there is agreement of the simulation results to the laboratory ones [109]. Sinervo and Clobert in a recent paper in Science, report an interesting aspect of spatial organization in populations of side-blotched lizards [110]. In lattice models [80, 81] small amounts of migration bring about "phase synchronization". Peak population abundances, however, are observed to be largely uncorrelated. Blasius et al. [67] use a Langevin type system of equations to introduce noise in the three-species spatially structured model, and obtain complex chaotic travelling-wave structures.

However, humans and animals alike are mobile. Furthermore, when studying the epidemiological systems, we have to take into account that many diseases are not airborne, and spread through direct contact. In such situations, the lattice models do not correctly represent the real spatial structure of the interaction and migration networks [29]. In recent years the science of networks has made enormous advances, and the structure of networks is by now relatively well researched. In the following paragraphs we will provide a concise description of the results of these new studies, and then continue by "putting our model on a network". This way we built a model that relates to the spread and survival of e-mail viruses in the Internet.

6.2 The New Science of Networks

The Internet is a complex network of computers and routers, connected by wires or wireless links, the World Wide Web is a network of web pages connected by hyperlinks (virtual), the social networks are composed of human beings (nodes) and social connections (links), the transportation network is a set of links (flights) between nodes (cities, towns), the cell is a network of chemicals connected by chemical reactions, etc. These are some of the examples of complex networks, whose structure (topology) has interested scientists, particularly physicists, in recent years.

This interest has been prompted by several, almost simultaneous, developments. The computerization of data acquisition made available large databases of the topology of various networks. These were open for study to researchers of other fields, who could compare their properties and find the common features of networks of very different objects. The fast increasing computing power allows us to investigate the properties of these huge networks. These properties are result of their topologies, which need to be understood. Three concepts are central to the modern understanding of complex networks:

Small worlds: It describes the fact that despite the large size of the network, there is always a relatively short path between any two nodes. The best known example is the "six degrees of separation" idea, discovered by Stanley Milgram [111] who stated that there is a path of acquaintances with typical length of six between any pair of people in the United States. The actors in Hollywood are within three co-stars from one-another, and the chemicals in a cell are about three reactions away.

Clustering: In social networks, particularly, we observe the formation of "cliques", representing circles of friends or acquaintances in which everybody knows everybody. This attribute is characterized by the clustering coefficient, according to Watts and Strogatz [112]. One way of defining this quantity for a node i with $k_i > 1$ edges is

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{6.1}$$

where E_i is the number of links between the k_i nodes connected to the i-th node. The average of these coefficients for the whole network gives the clustering coefficient of the network.

Degree distribution: The number of edges (links) for a particular node is known as the degree of that node. The degree distribution $P(k)$ or the probability that a degree have exactly k edges for the real networks (including the World Wide Web [113], Internet [114], and metabolic networks [115]) has a power-law tail:

$$P(k) \sim k^{-\gamma} \tag{6.2}$$

Networks with the above degree distribution are called scale-free [116], since the degree of a node does not depend on the size of the network.

6.2.1 Empirical Results of the Topology of the Real Networks

The World Wide Web is the largest network whose topology has been explored. The nodes are documents (web pages) and edges are the (hyper)links from one page to another. At the end of 1999 it counted close to one billion nodes [117, 118]. In 1999 it was made known that the degree distribution of the pages follows a power-law over at least six orders of magnitude [113,119]. The edges of the WWW are directed out of a document, or into a document. For both kinds of links we observe power-law degree distribution. Different authors report slightly different exponents, but for all of them γ_{in} is around 2.1 and γ_{out} in the range $2.4 - 2.7$ [113,119,120]. It displays the small world property, with the average path length predicted to be around 19 for the whole WWW [113], and measured to be around 16 for a sample of 200 million nodes [120].

The Internet is the network of computers, connected by wires or other wireless devices. Its topology is studied at the router level, with routers as nodes and the physical connections between them as edges, and at the interdomain level, with domains as nodes, and routes that connect them as edges. Faloutsos et al. [114] have studied the degree distribution at both levels, coming up with power-law distributions with exponents around 2.2 at the interdomain level, and 2.48 at the router level. It displays clustering and small world properties as well. The clustering coefficient has been measured by Yook et al. [121] and Pastor-Satorras et al. [122] to be between 0.18 and 0.3. The average path length is about 3.7 at the interdomain level [121,122], and around 9 at the router level [121].

Movie actor collaboration, science collaboration, phone-call, citations, human sexual contacts networks (as some examples of social/human contacts networks) have also been found to follow power-law degree distributions. For the movie actor collaboration network Newman, Strogatz, and Watts [123] report an unusually high clustering coefficient, and the exponent of the power-law degree distribution is around 2.3 [116, 124, 125]. For the science collaboration networks, the power-law behaviour has been observed in networks of physicists, mathematicians, neuroscientists, and biomedical sciences [126–129]. In the phone-call network the nodes are phone numbers and a long-distance call is an edge. The degree distribution for the phone calls follows a power-law with exponent 2.1 [130, 131]. An exponent of 3 has been observed for the incoming degree distribution of the science citation networks [129]. The network of human sexual contacts was studied based on an extensive survey conducted in Sweden in 1996. The distribution of partners over one year exhibits a power-law behaviour with exponents about 3.5 for females and 3.3 for males [132].

Features of scale-free networks have also been found among cellular networks, ecological (food web/chain) networks, linguistic networks, and protein folding.

All these examples have recently motivated scientists to find models that can generate the topology of these networks. This brought birth of the scale-free and related models. A very good review has been written by Reka Albert and Albert-László Barabási [29]. A good popular science book is "Linked" by Albert-László Barabási [133].

6.2.2 The Random Network Model

A network is represented by a graph, i.e. a set of nodes and edges (links) between them. Graph Theory has its roots in the work of Euler, who studied mostly small graphs. In the twentieth century graph theory became more statistical and algorithmic. The theory of random graphs was first used to study complex networks. It was founded by Erdös and Rényi [134–136]. Bollobás [137] wrote a detailed review of the field. Also useful is the review of the parallels between phase transitions and random graph theory of Cohen [138].

A random graph is defined as N nodes connected by n edges, randomly chosen among all possible edges [134]. By definition almost every graph has a property Q if the probability of having Q goes to 1 as $N \to \infty$. Erdös and Rényi discovered that many properties of random graphs appear suddenly, i.e. at a given probability (defined as $p = n/N$) the property is either present in almost every graph, or in almost none of them. This critical probability is the one familiar to physicists from the percolation phenomenon (for a primer on percolation, see, e.g. Stauffer [139]). We are familiar with percolation in finite dimensions (e.g. lattice), but the networks are of infinite dimensions, since the number of neighbours of a node increases linearly with the network size. For many properties we have to then define a threshold that depends on system size.

Erdös and Rényi [135] showed that there is a sudden change in the cluster structure as the average degree (i.e. the number of edges) per node $< k >$ goes through 1: from a loose collection of small clusters, a large cluster is formed, which for $< k >_c = 1$ has approximately $N^{2/3}$ of the nodes, and for larger $< k >$ has a number of nodes proportional to network size N. For large N the degree distribution of a random network is Poisson distributed with mean $< k >$. The diameter of these graphs, defined as the maximal distance between any pair of nodes in it, is almost constant, if p is not too small [140].

6.2.3 Small-World Networks

The above described random networks have clustering coefficients that are much smaller than those observed in real-world networks. Furthermore, these clustering coefficients are independent of the network size, an attribute characteristic of ordered structures. The first successful attempt to generate graphs with high clustering coefficients and small diameter was by Watts and Strogatz [112]. In their model, one starts with an ordered ring, where each node is connected to its neighbours, then rewire each edge with probability p. If $p = 0$ we have perfect order, and if $p = 1$, we get a random graph. This model exhibits high clustering, and small average path length, which scales logarithmically with the network size. A well-studied variant of the Watts-Strogatz model was proposed by Newman and Watts [141, 142]. The shape of degree distribution is similar to that of random graphs: it has a pronounced peak at $< k >$ and decays exponentially for large k. The topology of the network is essentially homogeneous, and all the nodes have approximately the same number of edges.

6.2.4 Scale-free Network Model

It is important to understand the mechanism responsible for the emergence of scale-free networks, in order to be able to find an algorithm to model them, and then put ensembles of different nature on such topologies. This requires a shift from modelling network topologies to modelling network assembly and evolution. The root cause is that if we understand the processes that assembled the networks as we see them accomplished, we have obtained their topology correctly. At this point the dynamics becomes the leading factor, with topology becoming simply a by-product.

Definition of the scale-free model. The origin of the scale-free behaviour in networks was first addressed by Barabási and Albert [113]. They attributed the scale-free nature of the networks to two mechanisms present in many networks. Real networks **grow** by continuously adding new nodes. The growth process involves the so-called **preferential attachment** of new nodes to old ones, i.e. the likelihood of a new node connecting to an existing one depends on the degree of that node. These two ingredients, growth and preferential attachment, form the core of the scale-free model of Ref. [113], which exhibits a power-law degree distribution. The Albert-Barabási algorithm for the scale-free model is the following:

(i) *Growth*: Start with a small (m_0) number of nodes, then every step add a new node with $m(\leq m_0)$ edges which link the newly added node to m

nodes already in the network.

(ii) *Preferential attachment*: The probability that the new node will be connected to node i depends on the degree k_i of node i:

$$P(k_i) = \frac{k_i}{\sum_j k_j} \tag{6.3}$$

This algorithm produces a scale-free network with an exponent equal to 3 for the power-law degree distribution. Its dynamical properties have been studied using various analytic approaches. Barabási and Albert [116] proposed a continuum theory; Dorogovtsev, Mendes and Samukhin [143] have used a master equation Ansatz; while Krapivsky, Redner and Leyvraz [144] introduced a rate equation approach. They all obtain a power-law degree exponent of 3. The average path length of the scale-free model follows a logarithmic dependence on the number of nodes in the network N.

This model is a minimal model that captures the mechanisms responsible for the power-law degree distribution. However, the exponent of this model is limited to 3, while the real networks have exponents that vary between 1 and 3. Measurements on real networks resulted in preferential attachment with probability $P(k)$ that a new node attaches to a node with degree k that is generally non-linear in k. In each case the $P(k)$ follows a power-law: $P(k) \sim k^{\alpha}$ with $\alpha \cong 1$ for the Internet [122, 146], citation network [146], Medline and Los Alamos archive [147], and $\alpha = 0.8 \pm 0.1$ for the neuroscience coauthorship and the movie actor collaboration network [146]. This has led to studies with nonlinear preferential attachment [144, 145]. Capocci et al. [186] introduce changes to the Albert-Barabási model, which make it a good description of the World Wide Web. In their model, the nodes are assigned relevances at random, and a new node establishes a link to an existing one (according to

Albert-Barabási rules), only if its relevance is higher. Amaral et al. [124] paid attention to the fact that many networks do show deviations from the power-law behaviour; however they are far from being random networks. Examples include the power grid networks, the neural network of the worm *C. elegans*, as well as the movie actor collaboration network. All these deviations can be explained by aging, i.e. the fact that actors have a finite active period, or cost for the electrical power grid or neural networks. When these factors were included in the models, for large k an exponential cut-off was observed.

Error and attack tolerance: The tolerance of many systems against errors is quite surprising. Examples of this class of systems are abundant: in today's environment, drastically changed from its natural state, organisms still persist and reproduce, due to the robustness of their metabolic and genetic networks [115, 148]. Communication networks also display a very high degree of robustness. This robustness is usually attributed to the redundancy of wiring, but a large part of it must be attributed to the topology of the network. The issue has been studied for the scale-free networks, and compared to the results obtained for other topologies, e.g. random graphs [149]. They considered two types of node removal: randomly selected nodes (failure tolerance) and most highly connected nodes (attack tolerance.) The error (or failure) tolerance of the scale-free networks is admirable: even with 45 percent of the nodes removed, the network still is one giant cluster with almost all the nodes in it. However, under attack the scale-free network performs very badly: when only 18 percent of the nodes are removed, the network ceases to exist.

Dynamics on networks: The advances in understanding the nature and topology of complex networks have opened a whole field of study. Often the topology plays a crucial role in determining the dynamical features of a system, as we had discovered earlier, with several different spatial structures introduced in our models in physics, ecology, epidemiology, population biology, etc. Many studies have been done already, and much more is to be done, since this is a huge area that is developing rapidly. Some of the most influential studies have been those of Watts [150], on the impact of clustering on several processes, such as games, cooperation, the Prisoner's Dilemma paradigm, cellular automata, and synchronization. Another widely investigated topic is search and random walks on complex networks [151–156]. It is important to keep in mind that modelling dynamics on a fixed topology is legitimate when the timescales of the network dynamics and those characteristic of the processes we are trying to model on the network topology differ widely. In the Internet the processes related to traffic have timescales of the order up to a day, while significant topological changes require months to develop.

The spread of ideas, innovations, and computer viruses is influenced immensely by the network structure [157, 158]. The dynamics is influenced heavily by the fact that these networks have a large clustering coefficient. Sometimes a different definition from Eq. (6.1) for the clustering coefficient is used, namely it is defined as three times the ratio of the number of triangles on the graph to the number of connected triples of nodes [112]. Spreading and diffusion has been studied also for the scale-free networks [156, 159, 160]. Pastor-Satorras and Vespignani [161, 162] studied the effect of topology on the spread of diseases. They showed that for the scale-free topology a local infection spreads over the whole network for any infection rate, that is there is no critical infection rate for the scale-free topology, a result that is quite different from what has been believed for a long time. Also, they show that the random immunisation of individuals does not lead to the eradication of infection in scale-free networks [163]. Ahmed et al. [164] employ a modified $SIRS$ model for the foot-and-mouth disease outbreaks, to conclude that ring vaccination can eradicate foot-and-mouth disease, provided that the probability of infection is high enough. Blanchard et al. [165] show that knowledge of the exponent of the degree distribution is not enough to determine epidemic threshold properties. The absence of epidemic threshold in a scale-free network with a diverging second moment has been proven exactly by Boguña, Pastor-Satorras, and Vespignani [166]. Properties of highly clustered networks, as well as social and biological networks have been studied, linked with the spread of disease (according to the SIR model) in them [167–171]. This area of study is a rapidly growing one.

6.3 The ABC Cyclic Model on a Scale-free Network

The ABC model we have studied so far in the non-spatial version has relevance to epidemiological studies, since, as pointed out in Chapter 3, it resembles the case when the immunity to a particular infectious disease is only temporary, otherwise known in epidemiology as the $SIRS$ model. (The similarity is not present at all the steps within the $SIRS$ model.) This is particularly relevant when one is interested in the spread of computer viruses in the Internet. We would like to be able to predict, just as for ordinary infectious diseases, whether they will flare up and die out, or remain endemic. Hence, it is interesting to investigate the behaviour of our cyclic model when placed on networks, while keeping in mind its limitations. Several questions arise, the answer to which will be determined by the topology of the system. First, let us discuss the relevance of the SIRS model to the computer viruses' spread.

6.3.1 Computer and E-mail Virus Features

A good review in the field of computer/network/e-mail viruses has been written by N. N. Bezrukov [172]. The very name of computer viruses comes from the obvious resemblance to natural viruses: their ability to self-replicate, high speed of spreading, the ability to select the target system (i.e. each virus only targets certain systems, or groups of systems), the ability to infect non-infected systems, our difficulties in fighting them, etc. Most recently, apart from these features, characteristic for both computer and natural viruses, one can also add a permanently increasing rate of mutation and the appearance of new generations of viruses. While for the natural viruses the mutation rate can be explained by the virtually infinite fluctuations in nature, in the case of computer viruses the rate of appearance of new mutations results exclusively from negligence, or evil intentions of people with a certain constitution of mind. Once a virus somehow enters a computer system, it will at least self-replicate in various places of the system, and then, or simultaneously with that, will cause such changes to the system, that can lead to catastrophic consequences. Most viruses parasite on particular programs or files.

Computer infections are usually classified into trojans and viruses. The trojans are programs that tell the user, “I am cool; open me”, e.g. the “I love you” bug, and then cause damage to the system. The viruses, on the other hand, are just parasites: they attach to a file, and only activate themselves if the user executes (opens) the file, otherwise they remain dormant. Apart from viruses, there exists thus far a relatively small group of programs, which exhibit features characteristic to viruses, otherwise called worms. While an ordinary virus is activated through the execution of an infected application or through the execution of a certain sequence of events by the operating system, the worm can independently control the execution of its own copies. While an ordinary virus resides on more or less definite places of the system, the residing place of a worm in the system is extremely difficult to predict, for it does not care about where it positions itself, but only about the existence of a system of links. Worms do not attack specific programs or files, but systems of specific architecture. They are generally more dangerous than viruses, exactly because of this relative autonomy, i.e. their independence of some unique environmental conditions, e.g. existence of certain files. The appearance, development, and then fast-spreading of computing networks caused the appearance of a sub-class of worms, so-called network worms. Net worms, which are more complex than ordinary worms, spread through the network using typical means of communication, including e-mail, network utilities, etc. They use the nodes of the network for their own purposes. Such activities can promptly shut down whole network systems. Hence, their spreading is directly related to the spreading of modern technologies. In most cases, a computer parasite combines in itself features of trojans/viruses and worms.

The global distribution of the viruses, which has become more and more obvious in the recent years, in a large degree is determined by the mass-production and popularity of personal computers, which are particularly vulnerable to infection, due to their standardised architecture and software. Large systems, in contrast, tend to have unique architecture and well-developed subsystems for access and security, which is an extra obstacle for viruses and helps the detection and isolation of the source of infection at very early stages. Microsystems, as a rule, are not unique. It is their standardisation, which by all means is to their advantage, that simultaneously becomes their curse. This is because a virus that has penetrated one system can, with no difficulty, infect a neighbouring system, since their architectures are similar like two drops of water. It is remarkable how fast viruses can spread, especially with today's telecommunications technologies. As a result they can rapidly block the whole network, and/or its individual subsystems [174].

In the notorious Morris case of 1988, within five hours the virus infected between 435 and 800 systems, altogether 6000 computers. Analysis of that infection showed that the virus spread through ARPANET, MILNET, Science Internet, and NSFnet. As a result, the whole defence system of the USA was shut down for two whole days, in addition to scientific and research works. Some nodes remained immune, due to the initiative of their system administrators. The reaction was swift, and within 24 hours the remedy (antivirus) was available, and it started spreading from three nodes, which were FBI, DCA, and NCSC. Apparently this was a worm attack, because no programs or files were changed. The "New York Times" reported that the whole event was result of an experiment gone wrong, and of a minor programming error. Another similar event took place about a month later, which was due to the existence of another mechanism of spreading of the worm. The spreading processes overlapped, but were not synchronised, so they seemed as two different worms. Furthermore, nothing unusual appeared in the system monitoring windows: the worm made itself look like another normally functioning task, while it was replicating itself and spreading to the neighbouring systems [173].

Once an e-mail virus installs itself on to an infected node, it checks for the addresses of the nodes to which the given node is linked, and via the SMTP (mail agent in general) sends seemingly legitimate mail, containing a small malicious code, written in a language understandable by the system targeted (C for UNIX and Linux, VB for Windows, or Word Macro for Microsoft Office). This message is executed on the target node, and starts working autonomously there, having avoided the requirement to go through the standard access procedures of the system. Then it starts executing files, becomes autonomous, and breaks the link to the parent process. Many viruses try to replicate and spread as much as they can, before being detected. Others do not reveal their existence before they produce a certain number of copies of themselves. Furthermore, many viruses are able to mutate by modifying their own code randomly, so that it is difficult to establish certain fingerprinting characteristics of a given virus. Compare this to the ability of natural viruses to mutate. It makes a reappearing virus unrecognisable to the immune system in living organisms, or the antivirus software in computer systems, causing reinfection. This circumstance justifies the use of a $SIRS$ model for this class of viruses.

Viruses which only replicate and spread will very soon exhaust all the resources of the network. So, more sophisticated viruses first check whether the node was already infected by another copy of the virus, and if so, don't re-infect it. Instead, they look for another victim-node. This way, they don't use too big a fraction of the system's resources, so they can keep thriving on it unnoticed for a long time. (One more aspect in which they resemble the natural viruses!). Soon the people who write viruses came up with self-encoding viruses, which require specialized antiviruses, designed to identify and destroy these specific viruses. The effect of this is you either have a set of specialized antiviruses, which are not universal, or you have something universal, which is slower and bigger in size, since it requires libraries, e.g. McAfee, Norton [175]. Originally the appearance of self-encoding viruses did not seem to bother anybody, until a new generation of viruses appeared: polymorphic ones. These use diversified self-encoding and decoding algorithms, which may not repeat. New principles of detection are needed. The new detection tools do not rely on the structure of the virus, but rather on its activity within the system.

Furthermore, viruses do not necessarily spread through the operating system (by using standard system procedures). Programs like Word or Excel have their internal macro language to execute a sequence of operations. All you need to get infected is to open a document or spreadsheet. Microsoft Outlook 97 can be configured to use Microsoft Word as its e-mail editor. Outlook 98 goes even farther: it uses Word by default, blocking any possible user's attempt to do otherwise [176]. This endangers the system just as much as opening a compiled executable. Outlook Express allows the execution of files attached to messages without the authorization of the user, or just in the background, while the message is open. This is the mechanism that makes possible the spread of viruses among PC users [177]. Even if the operating system can trace the operations executed by the macro, they all look legal (normal) to the system, so the system does not react. The defence strategies in this case are based exclusively on user's being careful, and not opening files when the source is unknown. Unfortunately, after a while, most users become "brave" and let their guard down, becoming once again vulnerable to exactly the same macro viruses (and creating another situation that calls for using a $SIRS$ model).

Needless to say that standardisation only fosters the spread of these viruses, by providing an environment in which these viruses can thrive. Some viruses have embedded in themselves quite sophisticated methods of guaranteeing their survival: they can encrypt the information stored on the hard disk, or they can move the FAT into another sector and fill it with some sort of garbage, according to a pattern, known only to themselves. In such a situation the structure of the virus is not to the user's advantage, for they would not even be able to access their data. In order to decrypt their data, users must knowingly execute the virus. The virus does not need to look for a loophole in the system in order to replicate itself: its replication will be guaranteed by the fact that it becomes indispensable for the user, who wants to access their data [177]. There is also “Live and Die” viruses, which infect programs only for a certain period of time [178]. After that time has expired, they either remove themselves from the infected program, or just terminate their activities. As a rule, programs retain their useability after the virus’ self-destruction. “Hide and Seek” viruses reside in memory for a certain period of time. Many viruses simply mutate: ShareFun sends its own copies not to all the addresses in the list, but to some of them, selected randomly from the list (under Outlook Express) [179].

The natural question is whether defence from viruses can be achieved through another “supervirus”? If we know the pattern of a virus, we can develop a supervirus which is designed to detect and destroy that very virus. Such a supervirus could be then introduced to the system; it would not have embedded infectious properties; all it would do is find and destroy the malicious viruses. Another scenario would be a supervaccine that would travel the net and inocculate all the applicable systems/files. The recent case of the LoveSan (Blaster) infection was particularly interesting. This was a network worm, which did not need any mediation to spread. An antiworm, named Welchia, started spreading soon after the Blaster. Welchia is another example of a benign virus, the first one being CodeBlue, which in September 2001 searched the Internet for computers infected by CodeRed and cleaned them. Welchia exhibits the same infection pattern as Blaster: through DCOM RPC bug in Windows, as well as the WebDAV bug embedded in IIS 5.0. It neutralizes the Blaster, installs the patch, and resets the computer. Finally, it removes itself from the system. (Epidemiologically, this corresponds to an SIR model.) The head of a leading antivirus research lab (Kaspersky Lab) information service, Denis Zenkin, was quoted as saying: “Apparently we are facing a new phenomenon in the Internet, which applies the well-known principle of driving a wedge by a wedge. It looks like the best remedy for a virus is another virus. The latest epidemics clearly demonstrates that most users are completely indifferent towards the security of their own computers, which has put the Internet on the edge of collapse. However, we do not advocate the idea of turning the Internet into an arena of zealous contest among viruses infecting computers with impunity, having the silent consent of their ignorant users." [180]

Fighting a virus with a virus does seem tempting, but the legal ramifications may make it non-feasible [178]. Also, existing means of antivirus protection are as efficient as the superviruses. Furthermore, there is always a chance that the supervirus might get out of control by unlimited replication and/or mutation. The Welchia story for one did turn sour. As reported by CBC and several news agencies, in August 2003 it spread out of control, clogging many computer networks, infecting over 500 000 computers worldwide, including the Air Canada check-in, the US Navy and Marine Corps, the Canadian Forces, all the 5000 computers of the Winnipeg Public Library, and many others. The Blaster/Welchia outbreak coincided in time with the revival of an old e-mail virus, called SoBig.F, which was responsible for 70% of the mail sent on August 20th. Vincent Weafer, senior director of Symantec Corp.'s Security Response centre was quoted by the BBC: "He [*the virus*-my comment] wants to build up a (robot) net by creating zombie machines he can control." The disaster was spectacular.

The spread of computing networks caused concerns that it would create an environment on which the viruses would thrive. Although this is true, the contrary is also true: computing networks create the environment in which the antiviruses can spread very quickly [172]. Once one of the users of the network is in possession of the means to combat a virus, he can share it with all other users in the network who are interested within minimal time. Particularly the case of the Morris virus demonstrated that even when the virus attack aims at the network itself, regular functioning of the latter becomes an important factor of fast and complete solution of the problem. The vendors of antivirus software discourage sharing of new antiviruses, since they want the users to buy the software, rather than obtain it for free. Even when the antivirus software is free, some users do not share its copies, because they consider the exclusive possession of such programs somehow advantageous. The logic of such policy of primitive egoism regarding antivirus programs is certainly erroneous, because sharing antivirus programs with other users creates an additional security buffer zone in which viruses can be detected and isolated before infecting their own computer. Therefore selfless sharing of antivirus programs is in fact a policy of sensible egoism. For your safety, you give away programs you have acquired for free, plus you gain some respect for the seemingly altruistic motives of such a deed.

As discussed in this section, the spread of some sorts of computer infections (but not all of them) can be epidemiologically modelled by a $SIRS$ model. Our ABC cyclic model is a reasonably good realisation of the epidemiological $SIRS$ model.

6.3.2 A Model for Social and E-mail Networks

When modelling the fate of viral infections, we have to keep in mind that the process of spreading will take place in a social network, for the natural viruses, and in an e-mail network, in the case of the e-mail viruses. Acquaintance networks have been studied extensively [111, 124]. Scale-free properties have been observed in some of them [127, 128, 183]. A model for the emergence of small world properties from local interactions has recently been introduced by Davidsen, Ebel, and Bornholdt [184, 185]. The model is based on "transitive linking", i.e. "my friend's friend is my friend", and a finite lifetime of the nodes. The dynamics is then defined as follows:

(i) A randomly chosen node picks two random "acquaintances", i.e. nodes it is already connected to, and "introduces them", i.e. a link is established between those two nodes. If the picked node has less than two "acquaintances", then a link is established between it and another, randomly picked, node.

(ii) With probability p, a randomly chosen node is removed from the network, i.e. "dies", and is replaced by a node with only one randomly chosen link.

The number of nodes remain constant. Since the nodes have a finite lifetime, the network reaches a stationary state. The death probability p is bound to be very small, since the time separation between meetings with new acquaintances is a lot shorter than that between major events, like somebody leaving the network (dying). The model described above exhibits large clustering, which only depends on the parameter p, and the (large) clustering coefficient can be tuned by varying the death rate p (this is the main reason we chose this model). When p is small enough, the network shows scale-free degree distribution; above that the degree distribution is exponential. This degree distribution is also observed in real-world social networks.

6.4 Simulations of ABC Model in Networks with Different Topologies

It seems reasonable to try to simulate the behaviour of the ABC model in a network, generated according to the transitive linking and finite lifetime assumptions [184, 185], since it correctly describes the social networks. The social networks are a good candidate, since they are the medium through which most viruses spread. Before Internet became a massive phenomenon, computer users exchanged (mostly DOS) software on diskettes. If I had some infected files, and had performed a backup before noticing that my computer was infected, and then made a favour to a friend by giving him a copy of my new "cool" software, the virus also jumped from my computer to his in the process. Nowadays, as we discussed above, viruses do not spread through disk exchanges, but rather in a more direct way: via e-mail. This is the prevalent way, at least before the Blaster attack (which is a virus-worm combination that spreads through the Internet without the "assistance" of computer users). Both the pre-Internet era computer user networks and e-mail networks appear as some "diluted" version of a social network (e.g. not all one's friends use e-mail). Hence, the above model does not completely describe them [182], but it is the best available model, so we decided to use it. In each case, the random networks, as well as scale-free networks (generated using the Albert-Barabási algorithm, or its modified versions), were used for comparison.

6.4.1 Real-Time Computer Simulations

The attempts that have been made at the study of the spread of infectious diseases in a network employ algorithms in which time is discrete, i.e. at each step one tries to change the configuration of the system according to some "rules". Using this approach, events cannot happen simultaneously. In our approach to the simulation of these systems, we aim at making the simulation fit the theoretical description, which means that we have to simulate continuous time Markov processes. In these processes the event times are exponentially distributed, and the characteristic time depends on the current configuration of the system. Then the main idea for the simulation engine is to keep track of the next event for each site. First, we initialize the simulation by iterating through the network, node by node, and picking an allowed event from the set of fundamental processes that specify the model. We record the event times and sort the times of the events for all the nodes in the network. After the initialization is completed, the simulation begins by executing the event with the earliest time. This event changes the configuration of the nodes associated with the one that had the earliest event time, together with changing the state of that very node, depending on the nature of the event. Then we generate a new event time for this node, and the ones coupled to it. The times are again sorted and the procedure is iterated. For this kind of simulation we need several node variables. The reordering of times is the most time-costly process, since the time cost when sorting an unsorted ensemble of N elements grows as fast as N^2. The solution we came up with is to use a "priority queue" to keep in it the elements in such a way that the earliest time is the first in the queue. This queue is a special type of a binary tree that maintains an ensemble of elements under the operations of inserting new elements and extracting the smallest element [181]. The time cost in initially sorting an ensemble is of the order of $N \log N$. The most important part is that the operation of removing the smallest, inserting a new element and reordering the queue requires a time cost of the order of $\log N$. This provides quite an efficient way to implement the continuous time simulations. All our simulation codes were written using C++ and compiled with the GNU compiler for the Linux operating system (under Windows, the Linux emulator program, known as Cygwin, and distributed by Red Hat, was used.)

The simulation will proceed as follows: the nodes will be assigned an initial state at random, i.e. A, B, or C with probability 1/3, and then they will be assigned a "wake up" time with exponential distribution, as described above. The node with the smallest time will be activated: if its state is, say, B, it will convert the nodes at state A among the ones it is connected to (emulating infection at contact), and analogously for other states. The process is then iterated.

For the purpose of generating the networks, starting with the random, and continuing with the scale-free and its many various modifications, as well as the social networks, we wrote a software we called Netgen. The Netgen software, whose source code is available on the website www.physics.ubc.ca/~birger, gives the user the opportunity to choose between different models for generating the network, and saves the network in a .net file. The format in which the network was to be stored posed a problem. The obvious choice would be to store the network's adjacency matrix, i.e. a matrix that contains a 1 in the (i, j) position, if there is a link from node j to node i, and 0 otherwise. However, if we calculate the memory we need to dedicate to storing the information about the network alone for, say, a network with 10 000 nodes, or 100 million elements, considering that we need at least 2 bytes per element, it is clear that the adjacency matrix is not a feasible choice. Furthermore, the time cost for detecting which nodes are linked to the one at hand would be enormous, if we proceeded with the adjacency matrix: at every step we would have to go through the whole 10 000-elements column, to sort the (typically) few ones among a majority of zeroes. This process would slow down the simulation, making it practically impossible.

We then opted for a “dynamic list”, which is technically a vector of memory addresses allocated for data storage. One can see a dynamic list as an array whose elements are also arrays of undefined length. Embedded functions allow us to search through the elements of the list and return the relevant information. The trouble is that this information is in specific terms, such as memory addresses. A person who is not a programming professional might not be comfortable handling this sort of information, therefore Netgen provides a number of embedded functions, which transform this specific information into a more familiar array-like output format. The structure of the network is written out into .net files, which contain the following blocks: (*i*) a generic network header, which contains information such as the number of nodes, algorithm, etc.; (*ii*) the block that contains the network itself. This is organized into lines, each one of which contains information about one individual node. The first element of each line is the number of links to that node, and the rest are the identifying numbers of the nodes that particular node links to. The third and fourth blocks are optional: (*iii*) the third block contains model-specific information, such as e.g. the exponent of the scale-free degree distribution. (*iv*) The fourth block contains model-specific nodes properties, such as e.g. relevances for the Capocci et al. model [186] (modified Albert-Barabási algorithm). This information is coded on four bytes per long integer, which provides 50% compression rate, as opposed to eight bytes per number character data. This way, a network of 10 000 nodes with a large average degree of, say, 200, occupies only 8 Mb disk space, as opposed to the 200 Mb that the adjacency matrix would occupy.

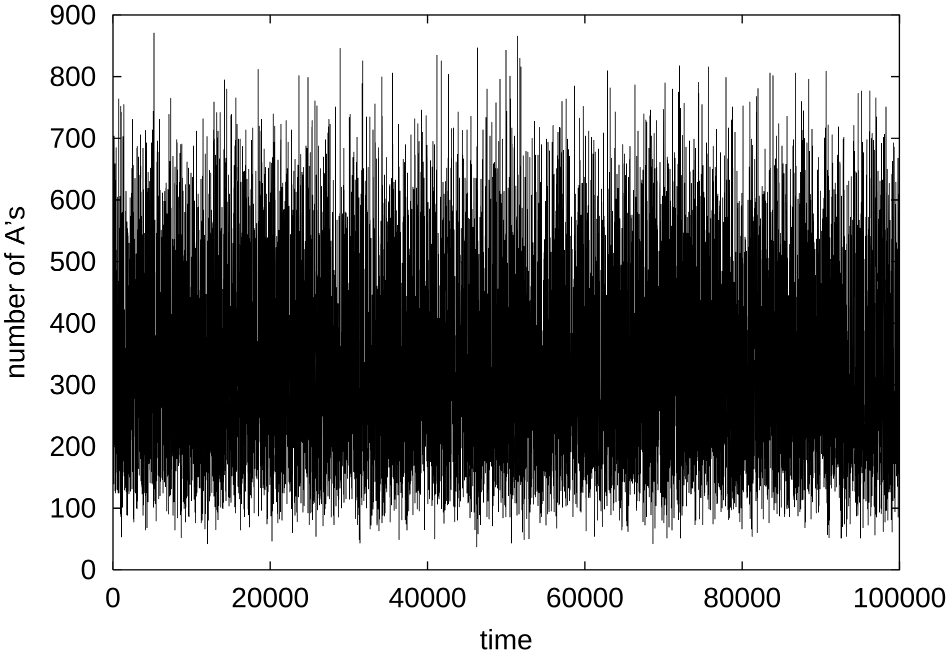

Figure 6.1: A typical time series for the ABC model in a scale-free network (Albert-Barabási model).

6.4.2 No Pandemics in the Internet!

We started our simulations with the Albert-Barabási model, and its modified version, after Capocci et al. [186]. The number of A, B, C individuals drifts with time, but none of them becomes zero. This situation is observed in network sizes up to 50 000. A typical time series is shown in Fig. 6.1. This situation is very applicable to network viruses. In the recent Blaster attack, e.g. a user who appeared on TV, complained that he had so far got infected three times. But does this mean that we cannot have pandemics in small-world and scale-free networks?

The model for the social networks, introduced by Davidsen et al. [184, 185], seems to be a good testing ground for this, since its average degree, as well as clustering coefficient, can be tuned by varying the death parameter.

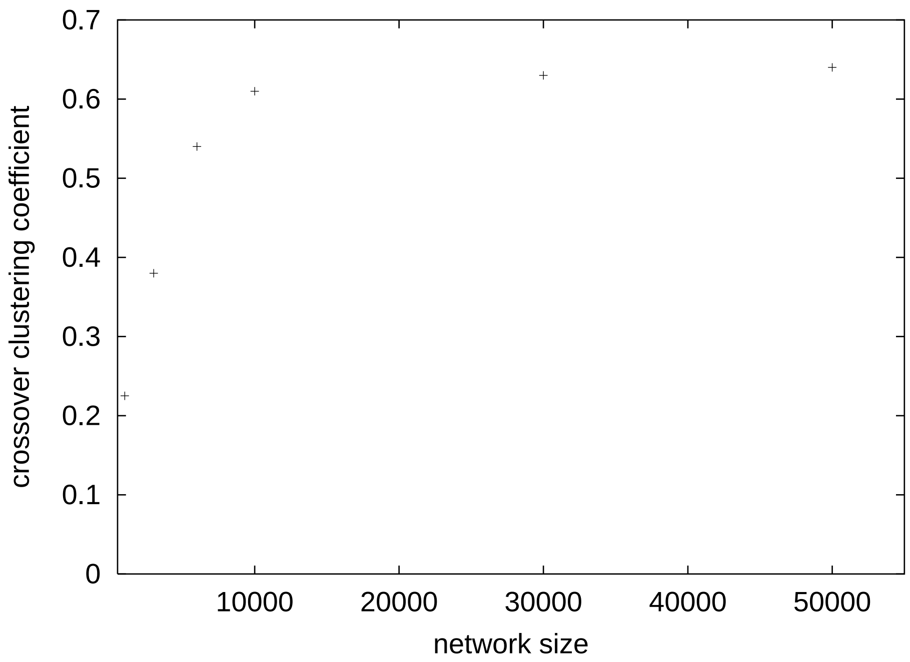

Figure 6.2: Critical clustering coefficient vs. network size for the social networks, modelled using the Davidsen et al. algorithm.

As we increase the clustering coefficient, the system undergoes a transition from the "diversity" regime to the "fixation" one, much like the one observed for the non-spatial system in Chapter 4 [31]. The variation of the critical clustering coefficient with the network size is given in Fig. 6.2. (We calculate the clustering coefficient according to Eq. (6.1).) It was possible to obtain an approximate value of the critical clustering coefficient for network sizes up to 50 000. Beyond that, the runs last for weeks. It can be seen that the critical value of clustering coefficient stabilizes around 0.65.

The computer users and e-mail networks appear as "diluted" versions of the social networks, so one would expect their clustering coefficient to be lower, perhaps close to those obtained with Albert-Barabási model, for which the ABC model does not reach fixation. A study of the e-mail networks was done recently by Ebel et al. [182], based on the log files of the e-mail server at the University of Kiel, recording the sender and receiver of every message from or to a student account over a period of 112 days. The network that emerged from the log files has 59 912 nodes (including 5 165 student accounts), and it has an average degree $< k >= 2.88$. The large cluster comprises 56 969 nodes and has an average degree of 2.96. The power law exponent for the degree distribution is 1.81. The network shows "small world" properties: its clustering coefficient is large, $C = 0.156$. It is one order of magnitude larger than the clustering coefficient one would obtain from a random network with the same degree distribution $C' = 0.0187$, and it is enormous compared to the clustering coefficient of a random network with the same average degree $C_{rand} = 4.82 \cdot 10^{-5}$. Its topology cannot be obtained from the preferential linking model. However, a model that is based on transitive linking [184], i.e. on the assumption that two nodes are more likely to link, if they have a common neighbour, can reproduce the small world properties and the degree distribution, but it also produces too high a clustering coefficient, and does not yield a power-law degree distribution for this network. As far as clustering goes, one end of the spectrum (when networks are considered) are the social networks, while the other one are the random networks. Runs performed on random networks were quite surprising: one would require enormous connectivities (clustering coefficients) to reach fixation on them! A plot of critical connectivity vs. network size for the random networks is shown in Fig. 6.3. The critical connectivity reaches values close to 0.4, which is enormous.

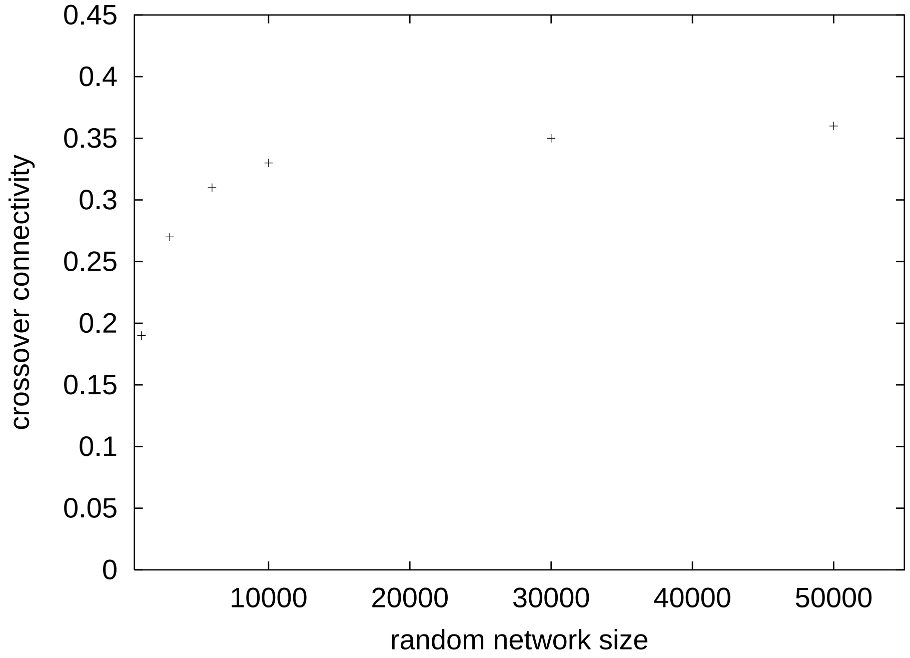

Figure 6.3: Critical connectivity vs. network size for random networks.

6.5 Conclusions

The modern computer and e-mail viruses, with the ability to mutate (some of them), and the absence of a reliable antivirus (for others) call for $SIRS$ epidemiologic models, rather than simply SIR. Simulations of the spread and evolution of such systems in different scale-free topologies, such as Albert-Barabási model, or the network model modified following Capocci et al. [186], show that all three types, (A, B, or C) are present in the network at all times. An attempt made in social network models, shows the presence of a transition to the fixation state for clustering coefficients close to 0.65. This result would relate to the observed peak and disappearance of epidemics in human communities. A test run performed on random networks shows that the critical connectivities approach 0.4, which is quite an enormous connectivity. Apparently, the computer viral epidemics do not exhibit the same characteristics as human epidemics, because the clustering coefficients for the computer users/ e-mail networks cannot achieve the critical values. The clustering coefficients of e-mail networks, as measured, do not seem to get high enough to approach the critical values obtained by our simulations. It seems that the pandemics of a virus in the Internet is an unlikely event, unlike pandemics in human social networks. The Internet is threatened more by the side-effect of clogging from the uncontrolled multiplication of viruses.

Only recently humans started to realize to which catastrophic state they have brought the natural environment. Right before our eyes the world is transforming into a global trash bin. Similarly, viruses, trojans, worms, and their combinations are becoming the polluters of the information environment. That is why the struggle with the computer viruses can be seen as part of the ecological movement; just like the problem of polluted nature, the problems of the computer viruses go beyond national borders. Local pollution of some place almost always turns first into a national, and then into an international problem. Transnational spread of viruses brings economical harm, which can be commensurated with the transnational spread of greenhouse gases or acid rain. It is interesting to point out that the 1988 attack already uncovered the vulnerabilities of computer systems. While many specialists emphasized that the lessons learnt could help in preventing and defending against future virus attacks, there were people who argued that the development of means of computer security, harms and slows down economic performance(!) [173]

Chapter 7

Variable Investment, the Continuous Prisoner's Dilemma and the Origin and Death of Cooperation

7.1 Introduction

The origin of cooperation is a fundamental problem in evolutionary biology. Cooperation is essential in the functioning of almost every known biological system [187–190]. For example, according to Eigen and Schuster [191], Michod [192], Maynard Smith and Szathmáry [193], the early replicating molecules have cooperated to form larger entities which can encode more information. Also, passing from free-living single-cell protists to multicellular organisms seems to have depended on cooperation [193, 194].

In Chapter 5 we visited many-species cyclic competition models, and studied their behaviour. In the no/low migration regime, those systems end up in a "ground state" in which two (or more) species coexist in symbiosis: they help one-another fight and defeat common enemies. But in general it is difficult to explain why individuals would cooperate, given that defecting individuals always have a higher fitness than cooperators, since they receive the benefits of the contributions of the cooperators without ever paying the costs. If a cooperative mutant appears in a world of defectors, it would be eliminated by natural selection. Even if cooperation had already been established somehow, defectors would soon increase in frequency, and take over the whole population. So, a central theoretical problem is to shed light on the mechanisms of evolution of cooperative behaviour in an essentially selfish world, and out of selfish motives; further one would have to wonder how can it then be maintained in a selfish world.

7.2 Prisoner's Dilemma

The Prisoner's Dilemma is the most widely used paradigm for the problems surrounding the evolution of cooperative behaviour. It traces its origins in the paradigm of two delinquents who have been apprehended for a minor trespassing, but have together committed another, much more serious, crime. Each of them is questioned separately, and has two alternatives facing him: cooperate, i.e. do not tell, or defect, i.e. tell the police of the major crime. If they cooperate, they each get one year in prison for the minor trespassing; if only one of them defects, he gets a minor punishment and the other gets ten years in prison; if both defect, they both go to prison for a long time (but less than ten years). Clearly, the best alternative for both is cooperation; however, each of them is afraid the other will tell the police, and defects. Both lose.

This story has been the basis for the Prisoner's Dilemma game, which has been set up as a game with two players, each of which has two available strategies: cooperate (C) or defect (D). The payoff for mutual cooperation (R) exceeds the payoff for mutual defection (P). The highest payoff (T) goes to the player who defects while the other cooperates, who then gets the lowest payoff (S). The letters T, R, P, and S denote Temptation to defect, Reward for mutual cooperation, Punishment for mutual defection, and the Sucker's payoff for cooperating with a defector. In the Prisoner's Dilemma (PD) the ordering is $T > R > P > S$ [195]. Inevitably, in a one-shot game, the winning strategy is defection for both players, as explained above. The requirement of rationality brings about an outcome that is not the optimal one from both players' perspective. It is not surprising then that the PD is considered *the* fundamental paradigm in social sciences as well as evolutionary biology, since it poses the fundamental question: how could the cooperation evolve in populations whose relations are governed by the Prisoner's Dilemma?

7.2.1 Repeated Games

An important element is *repeated play*, since the social and biological interactions are reasonably stable over a certain period of time. Many studies have involved individuals who play against the same set of individuals (or agents). If one agent plays D in a sequence of encounters with another agent who also plays D, both players end up scoring a lot less than they would by cooperating. Axelrod and Hamilton were pioneers in this area [196]. They assumed that players remember past encounters, i.e they have memory. If the players do not remember their previous engagements, that is not a repeated game. A "strategy" is simply a rule that specifies the agent's behaviour, given the history of encounters with the other agent opposite him. It is important to emphasize that an iterated PD game must be of indeterminate length (number of iterations), otherwise it always regresses to pure defection. Axelrod in his book [197] describes a race in which the participants were asked to submit a strategy. They all played against one-another, and at the end the strategy submitted by Rappaport won. It was the very simple Tit for Tat (TFT) strategy: cooperate in the first move (be nice), then always do what the opponent did in the previous encounter (retaliate). An extensive literature has since developed strategies that support cooperation for the repeated PD [197–200].

7.2.2 Evolutionary Games

Another aspect of the PD-based games is classical evolutionary game theory [201, 202]. In this version the agents do not optimize their strategies, rather they inherit a phenotype (a fixed strategy) and replicate depending on the payoff they receive (their fitness). These *replicator dynamics* have a very interesting attribute: their *evolutionary stable strategies* (ESS) or Nash equilibria can correspond to the strategy that a fully informed, rational agent, playing the game, would adopt. There is a vast literature regarding evolutionary games; see e.g. Binmore [203]. In the PD games, the classical replicator dynamics yields a world of pure defectors, which, as we saw, was the dominant strategy in a one-shot game. Frank [204] puts forth the argument of basic evolutionary reasoning in the context of social science, for a “well-mixed society”, i.e. what the physicists call mean-field approximation.

The basic assumption in his argument is that every agent is “hard wired” to execute one of the two available strategies: cooperate or defect. A cooperator is someone who, perhaps through intensive cultural conditioning, has developed an inheritable capacity to experience a certain moral sentiment that predisposes him to cooperate. A defector is someone who lacks such capacity or has failed to develop it [204]. Now "suppose, for argument's sake, that cooperators and defectors look exactly alike, thus making it impossible to distinguish the two types." The individuals will then pair at random. "The expected payoffs to both defectors and cooperators therefore depend on the likelihood of pairing with a cooperator, which in turn depends on the proportion of cooperators in the population." If the proportion of cooperators in the population is c and individuals are paired at random, the probability of a cooperator interacting with another cooperator is c, and with a defector $(1-c)$. Then the payoff of a cooperator is $E(C) = cR + (1-c)S$ and that of a defector $E(D) = cT + (1-c)P$. Since for PD $T > R > P > S$, it follows that $E(C) < E(D)$. Frank then states: "Since defectors always receive a higher payoff here, their share of the population will grow over time. Cooperators, even if they make up almost all of the population to begin with, are thus destined for extinction. When cooperators and defectors look alike, genuine cooperation cannot emerge." [204] This mean-field treatment motivated the experts in game theory to introduce "tags", which make cooperators and defectors distinguishable at sight [205].

The same argument can be made from the point of view of the classical replicator dynamics [201, 202]. The relative frequency of a strategy i grows according to the rate equation:

$$\frac{dz_i}{dt} = z_i(E(i) - < E >) \tag{7.1}$$

where $E(i)$ is the return to the strategy i, and $< E >$ the average return. In the case of the PD (two strategies), the z vector is $(x, 1-x)$, where x is the relative frequency of defection. This leads to the rate equation:

$$\frac{dx}{dt} = (T - R)x + (P + 2R - S - 2T)x^2 + (T + S - P - R)x^3 \qquad (7.2)$$

The steady state (extremum) solutions are $x = 0$, $x = 1$, and $x = (R-T)/(R-T+P-S)$. The eigenvalue for $x = 0$, i.e. no defection, is $T - R > 0$, thus the solution is unstable, while the eigenvalue for $x = 1$, i.e. universal defection is $S - P < 0$, so the solution is asymptotically stable. Any perturbation from pure cooperation ($x = 0$) will drive the system to pure defection ($x = 1$), in agreement with Frank's argument.

7.2.3 Demographic Games

Nowak and May [206] placed a simple ensemble of permanent cooperators and defectors on a lattice. In their simulations, in each round, each agent plays the PD game with its immediate neighbours; after this each site is taken over by the agent (in the neighbourhood) who scores best in that round. This game generates changing spatial patterns, in which both cooperators and defectors persist indefinitely. Epstein [195] introduced what he called the Demographic Prisoner's Dilemma. In it agents inherit a fixed strategy, and are indistinguishable; the system is not well-stirred, but rather placed in a spatially structured world, in Epstein's case, a lattice. The *demographic games* involve spatial, evolutionary, and population dynamics (including migration). Agents move around in this space, interact with their neighbours, and have offspring according to their fitnesses. Periodic boundary conditions apply. Agents have attributes such as vision, wealth, age, and strategy. They are born with a strategy to cooperate or defect. Their only rule of behaviour is as follows: choose a random unoccupied site within your vision; go there and play your strategy against each neighbour. The payoffs are constructed in such a way that $T > R > 0 > P > S$, and they accumulate as wealth. An agent's wealth can become negative, in which case they "die". If the wealth accumulated by an agent surpasses a certain threshold, they will produce an offspring, which will inherit the parent's attributes, as well as an initial endowment, and will occupy one of the unoccupied sites within parent's neighbourhood. The interesting phenomenon is the appearance of zones of cooperation, and the resulting robustness of cooperation as a phenomenon.

In some versions the model allows for mutations in the process of vertical (parent to child) as well as horizontal (intragenerational) transmission. In another, equivalent version, agents are able to learn, rather than die and reproduce; they also develop their expectations on how their actions will affect others [207]. In all these, and many other models, the maintenance of cooperation in absence of any strategic complexity represents a very important advance in the understanding of cooperation. However, these models consider the invasiveness and stability of fully developed, highly cooperative interactions.

The gradual evolution of cooperation from an initially selfish state represents a more plausible evolutionary scenario. It is then more natural to consider models in which several degrees of cooperation are possible [107, 108, 208–212]. When we take into account the possibility of variable levels of cooperation, we can study the crucial issue of how cooperation can gradually evolve from a non-cooperative initial state. Roberts and Sherratt [210] came up with a "raise-the-stakes" strategy for the iterated PD, and showed that it invades and is stable against a number of alternative strategies. In their model the strategy is fixed and does not evolve, but the extent of cooperation, as well as its frequency between two interacting agents can increase. Doebeli and Knowlton [209] considered interspecific relationships, and concluded that cooperation could increase in extent and frequency, if the populations are spatially structured. In this model, strategies with very low levels of cooperation can gradually evolve to much more cooperative strategies. The end result is a high degree of mutualism between pairs of interacting individuals that belong to different species.

Killingback, Doebeli, and Knowlton [213] extended the classical Prisoner's Dilemma, introducing a model of cooperation which is based on the concept of investment, and develops further the ideas of Doebeli and Knowlton [209]. This evolutionary game is called Continuous Prisoner's Dilemma. They showed that intraspecific cooperation easily evolves from very low levels, and is sustained and fluctuates around relatively high levels, when the game is played in spatially structured populations. This chapter further explores their model, its behaviour in different topologies, and the limits at which cooperation breaks down and defection is the winning strategy.

7.3 The Basic Model: Continuous Prisoner's Dilemma

The Continuous Prisoner's Dilemma (CPD) game between two individuals is based on the assumption that each of them makes an investment I (which can be zero). It has the effect of reducing the fitness of the person who makes it by "the cost" $C(I)$ and increasing the fitness of the beneficiary by "the benefit" $B(I)$. So, if 1 and 2 play against each-other and make investments I_1 and I_2, the payoff of 1 is $B(I_2) - C(I_1)$ and that of 2 is $B(I_1) - C(I_2)$. Possible benefit and cost functions are shown in Fig. 7.1. Cost and benefit functions of this

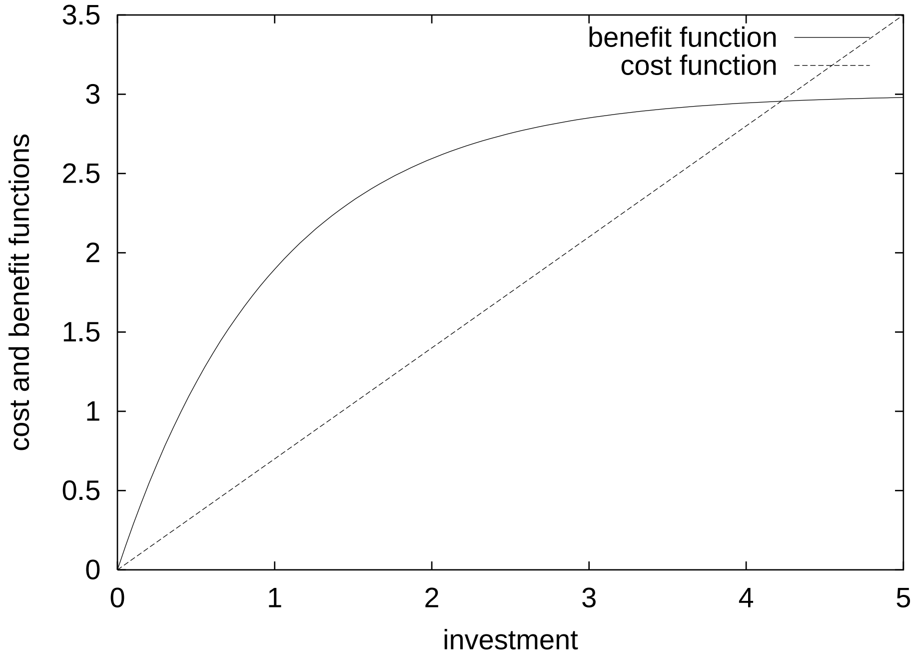

Figure 7.1: Possible benefit and cost functions for the CPD game.

type are typical of what might be expected in a real biological situation, such as those discussed by Hart and Hart [214] and Wilkinson [215]. The common feature of the functions $C(I)$ and $B(I)$ is that $B(I) > C(I)$ for a range of the argument $0 < I < I_{max}$ (for $I > I_{max}$ the cost is higher than the benefit, so it does not make sense for a player to invest an amount higher than I_{max}). Limiting the investment levels to just a pair of investment amounts would bring us back to the standard Prisoner's Dilemma.

In absence of an additional structure, i.e. the mean-field approximation, or the well-mixed system, the zero-investment strategy is again the winning one. Starting at any level, investments will gradually evolve to zero, since the defectors will benefit from the investment of the cooperators without bearing the costs. Using adaptive dynamics, one can see that the zero investment strategy is the evolutionary stable strategy.

7.4 CPD in a Lattice

Killingback et al. [213] introduced spatial structure to the model, following the general approach of spatial evolutionary game theory [197, 206, 216]. The individuals are placed in the cells of a 2D square lattice. Each individual makes an investment, and interacts with the individuals within a Moore neighbourhood. (A Moore neighbourhood of, e.g. size one, includes a cell's eight immediate neighbours: north, northeast, east, southeast, south, southwest, west, northwest; that of size 2 will include two immediate neighbours in all directions, etc.) This individual will get payoffs as prescribed by the rules of the game, when playing against each of the neighbours. The fitness of the individual is then the sum of the payoffs she gets from playing against her neighbours. The fitness of the neighbours is calculated too, as the sum of the payoffs they get from playing against their own neighbours. She then looks around, identifying the neighbour with the highest fitness, and adopts that neighbour's strategy, by changing the amount of her investment to that of the neighbour's. At this stage the investment amount of the player can be mutated at a fixed probability. This corresponds to an economic scenario in which the agents learn from their more successful partners, or to an evolutionary scenario in which the more successful phenotypes replace the less successful ones (with noise included in the model).

We let the system evolve according to these rules. Every few generations we calculate the average investment per individual. When a neighbourhood of size one is chosen, the average investment increases from an almost zero starting value, to a significant final level of investment. Its average value for a cost function $C(I) = 0.7 \cdot I$ and benefit function $B(I) = 8 \cdot (1 - e^{-I})$, in a 70×70 lattice, with periodic boundary conditions, is around 1.05. (We introduce mutations with probability 0.01, which follow a Gaussian curve with centre at the individual's investment value, and width 10 percent of its peak.) The average investment amount is maintained close to 1.05 by the dynamics of the spatial system. It is much smaller than the maximum investment I_{max}, but much larger than the initial maximum investment amount of 0.0001.

The key factor that causes a state with considerable average investment levels to be established is the fact that higher investment individuals benefit from clustering together. One such cluster, when placed in a sea of individuals with lower investment amounts, will (in most cases) expand, since their payoffs get higher from the interactions with individuals who invest comparable amounts. This way, clustering can result in the growth of geometrical structures with higher-investing strategies, which is observed in the simulations by Killingback et al. [213], as well as in our simulations. Furthermore, if one individual undergoes a mutation that increases her investment amount only slightly, for certain configurations of her neighbourhood, she might still have higher payoff (fitness) than her neighbours, in which case she will take over. A cluster of higher investing individuals will be formed, and it will expand, as explained above. These mechanisms that drive up the average investment value apply to any choice of benefit and cost functions which satisfy the basic assumption that the investment benefits are higher than the costs for a range of investment values. This condition is a necessary one for cooperation to evolve in any situation, and is analogous to the ordering of the payoffs for the standard Prisoner's Dilemma. The choice of the benefit function in our case satisfies the condition that for increasing investment the payoffs diminish. This, as we already mentioned, is a realistic choice for many real-life situations.

As reported previously by Killingback et al. [213], cooperation evolves and is maintained quite easily in the CPD game, and this suggests that cooperation might not be so difficult to explain. The above results were originally obtained for deterministic synchronous updating (one full sweep after another), and a perfectly regular lattice. The synchronous updating has been known to produce unrealistic and misleading results: Huberman and Glance report the disappearance of cooperation in the Nowak and May [206] model, once asynchronous updating is introduced [217]. Killingback et al. [213] have investigated the model with stochastic noise in the updating [218] in which the successful neighbours take over a site with only 80% probability: the system evolves in a similar fashion to that with deterministic updating. Asynchronous updating [217,219] not only results in similar levels of cooperation, but the rate at which the investment evolves to the higher level is about ten times faster than that for the synchronous updating (10 000 generations for asynchronous updating, as opposed to 100 000 for the synchronous one). The same investment levels are reached for a random spatial lattice of the type used by Nowak et al. [219], as well as when the mutations are heavily biased toward defection. The general underlying effect is that some higher-investing mutations eventually form basic clusters (which then expand), and the random irregularities seem to even favour this effect.

We introduced "real-time" evolution as follows: every individual gets her own "clock" which tells her when to "wake up", look around, play against the neighbours, compare her payoff to that of the neighbours, and revise her strategy. The "wake up" times are generated as $-\ln(rn)$ where rn is a random number with uniform distribution in $[0,1]$ [62]. (This corresponds to asynchronous updating, but with a different distribution of "wake up" events.) After a sufficiently long time (which is of the order of 3 000 generations), the ensemble again reaches a state with a considerable degree of cooperativity, and the average investment level is about the same as that obtained by Killingback

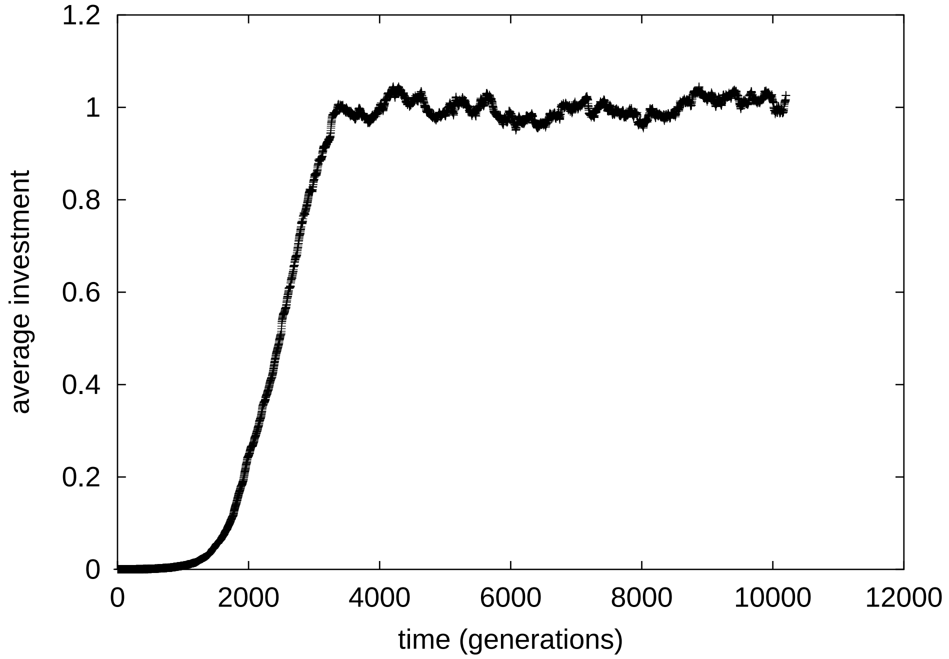

Figure 7.2: The evolution of average investment with time when we only consider the nearest neighbours for interaction-control run.

et al. [213]. A typical evolution picture is given in Fig. 7.2.

In the real world, an individual's vision is not confined to her nearest neighbours! So, we started our exploration of the model by checking how does the picture change, when the neighbourhood size is increased? Killingback et al. [213] in their discussion mention that they expect the relative size of the interaction neighbourhood to that of the whole lattice to be a factor. When studying the PD in lattices, it is customary to have the agents interact with individuals within a certain neighbourhood size, and then compare themselves with the same individuals, within the exact same neighbourhood. We will distinguish between the two processes: interaction ("play against") and learning ("compare with"). The radius of the interaction neighbourhood will be referred to as **the neighbourhood parameter**. The **dispersal pa-**

rameter will be the radius of the neighbourhood, individuals within which she compares herself to, and learns from (or is replaced by, in the evolutionary scenario). Keeping a Moore type of neighbourhood, and the neighbourhood and dispersal parameters equal to each-other (for the moment), we considered the case when an individual plays against everybody within a neighbourhood with radius 2 (24 individuals), 3 (48 individuals), 4 (80 individuals), and so on. The rules of the game, shape and parameters of the cost function, benefit function, mutation rate, width of the Gaussian mutation curve, are kept the same. For each case, we averaged over 500 realisations of the system.

For neighbourhood and dispersal parameters equal to 2 (i.e. a neighbourhood that includes 24 individuals, as opposed to 8 for neighbourhood parameter 1), the final average investment value is very close to that obtained when those parameters are 1, but, curiously enough, the average investment has a "jump" at early evolution times, as shown in Fig, 7.3 (this "jump" persists in different realisations of the model). When we increase the neighbourhood and dispersal parameters to 3, the final average investment drops to around 0.85, and it drops even further to about 0.6 for both parameters' values equal to 4. Some typical runs are shown in Fig. 7.3.

The picture changes qualitatively when the neighbourhood and dispersal parameters become 5. Even though we start with a relatively big initial investment, the final average investment is extremely small (about 0.04, as opposed to 0.6 for a size four neighbourhood), and can occasionally drop to zero under the influence of even moderate fluctuations. This is shown in Fig. 7.4. It can be safely assumed that at this neighbourhood size we have reached the mean-field limit, i.e. the picture we would obtain in absence of a spatial structure altogether (which for the CPD corresponds to a state of pure defection, as discussed above). The value of the neighbourhood and dispersal

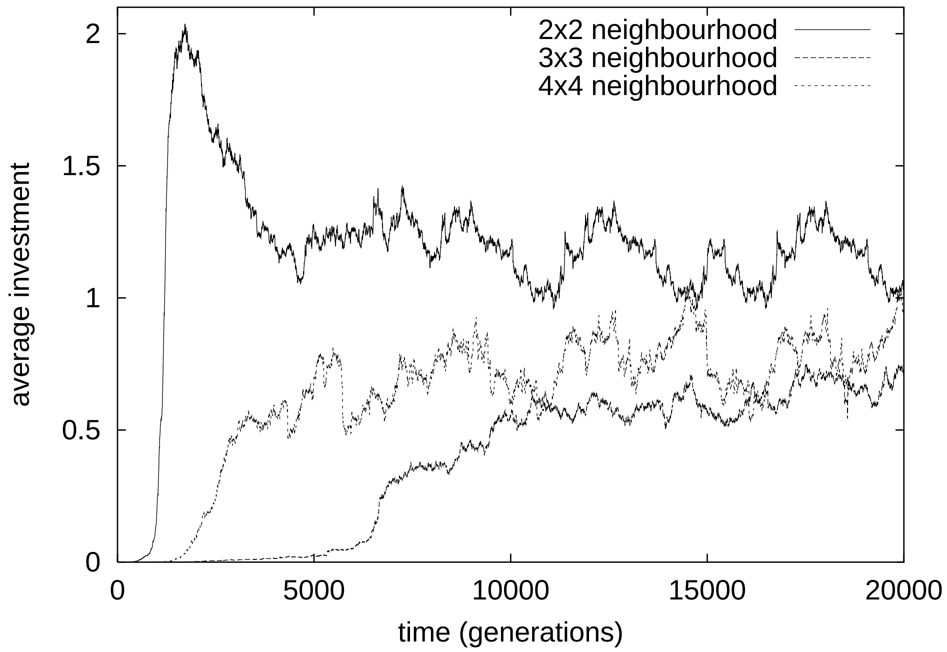

Figure 7.3: Average investment vs. time for the cases when we consider more than one neighbour, here the neighbourhood and dispersal parameters are equal and vary from 2 to 4.

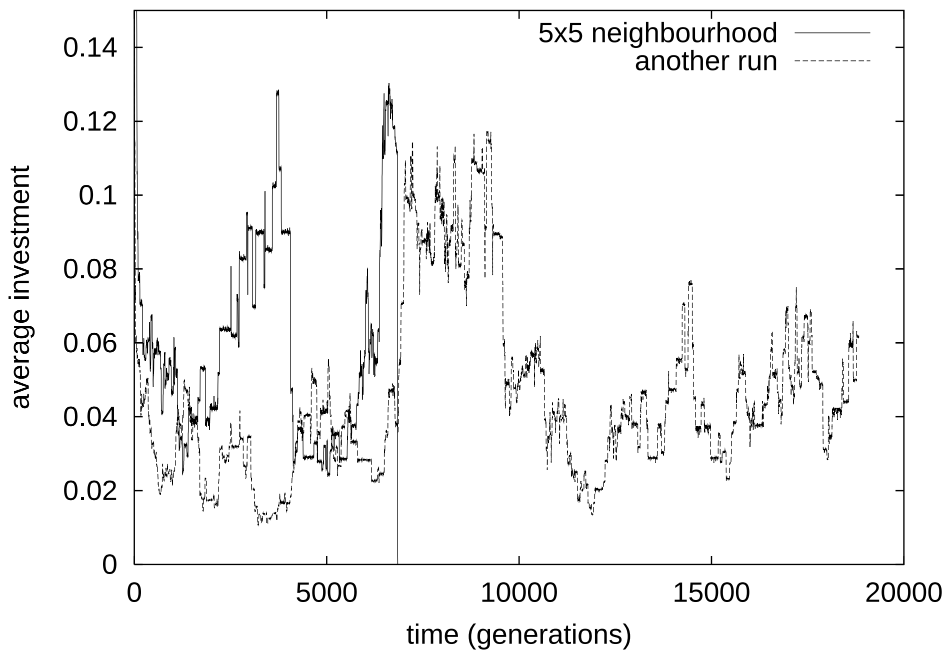

Figure 7.4: In the case when the neighbourhood and dispersal parameter become 5, we are at the mean–field limit.

parameters for which the mean-field limit has been reached was observed to be independent of the lattice size, at least for reasonably large lattices, and the immediate neighbourhood comprises about 100 individuals, while the extended one (the neighbours of the neighbours) about 300. Furthermore, the results were independent of the choice of the boundary conditions; they hold even when periodic boundary conditions are not introduced (and the last row cells have only five neighbours, as opposed to eight the other cells have, for a size one neighbourhood).

Let us now come back and examine the role of the ratio of the neighbourhood size to the lattice size. In Fig. 7.5 we show a time series of average investment for the case when both the neighbourhood and dispersal parameters are 4, and the lattice size is 30×30. It is important to note that for

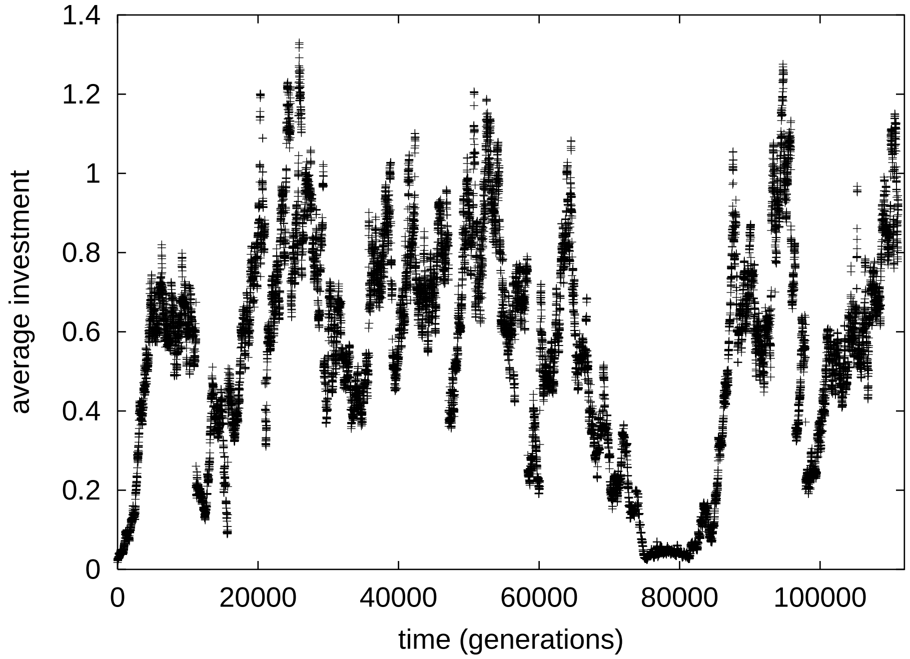

Figure 7.5: The evolution of average investment with time in the case when both the neighbourhood and dispersal parameters are 4, but the lattice size is 30×30. Note the huge fluctuations, which bring the investment very close to zero as a result of mutations.

smaller lattices (or "worlds"), the size of fluctuations can become comparable to the average investment amount, even when the latter is considerably above zero, and often a mutation is enough to result in the final victory of defectors (the average investment does come dangerously close to zero for a considerable interval in Fig. 7.5). We can then conclude that the neighbourhood size for which we achieve the mean-field limit does not depend on the lattize size (at least for reasonably large lattices), contrary to the expectations of Killingback et al. [213].

7.5 Variable Neighbourhood Size, the Role of Information, and Evolution

We mentioned above that the size of the interaction neighbourhood can be different from that of the learning neighbourhood. We then simulated a situation in which our agents play against individuals within a certain neighbourhood, but compare their payoffs to those of individuals within a neighbourhood that is different (larger or smaller) than their interaction neighbourhood. It is observed that, if the neighbourhood parameter is taken to be different (by one or more) from the dispersal parameter (as defined above), the cooperation is no longer sustained, and, independent of whether the neighbourhood parameter is larger than the dispersal one, or the opposite is true, we end up with a final state of pure defection. When the neighbourhood parameter is smaller than the dispersal parameter, the defectors outside the interaction neighbourhood, but within the learning neighbourhood, manage to "tempt" the cooperators. In the case when the neighbourhood parameter is larger than the dispersal one, there are simply too many defectors in the interaction neighbourhood, taking advantage of the cooperators. In the economic context, too little or too much information favours defection, and kills the incentive to cooperate.

In the real life, however, there are no walls to separate neighbourhoods from one-another. Individuals interact not only within a neighbourhood with rigid walls, but rather occasionally include other individuals in their area of relations. This motivated us to find out what is the (fractional) difference between the neighbourhood and dispersal parameters for which cooperation is no longer sustained. This could be emulated in our model by making the neighbourhood and dispersal parameters "continuous". For that we constructed a "fractional size" neighbourhood the following way: suppose the parameter has the form $< m.n >$ with m the integer part and $< 0.n >$ the fractional part. Then for the $(m+1)$-th neighbours we "throw the dice", generating a random number with uniform distribution in $[0, 1]$. If the random number is smaller than $< 0.n >$, the individual plays against this particular neighbour, otherwise the next event in the queue is generated. If the neighbour is at the corner, the random number is compared to $(0.n)^2$ (i.e. the fractional area of the rectangle). The remaining rules of the game are unchanged. Cooperation only persists for differences between parameters up to about 0.5, and breaks down when the difference between the neighbourhood and dispersal parameters is 0.5 or larger (independently of whether the neighbourhood, or the dispersal parameter is larger). Some runs with different neighbourhood and dispersal parameters are shown in Fig. 7.6. In it we can see that the average investment has already droped to zero, when the difference between the neighbourhood and dispersal parameters reaches 0.6.

It is not given that every agent should have the same parameters (e.g. their visions might be different), so we treat them as "local", i.e. every individual has her own neighbourhood and dispersal parameters, n_i and d_i, which are not integers. When her turn comes, she plays against individuals within a neighbourhood with radius n_i and then compares herself to the individuals within the neighbourhood with radius d_i. However, there are cases when she meets a neighbour j, whose dispersal parameter is smaller than hers (i.e. he does not "see" her, even though she "sees" him). In such cases she could include him into her learning neighbourhood, comparing her payoff to his, and then adopting his investment amount, if he is doing better than her (in the evolutionary scenario, he would take over). Or she could leave him out of her learning neighbourhood, since he cannot take over a cell he does not even see.

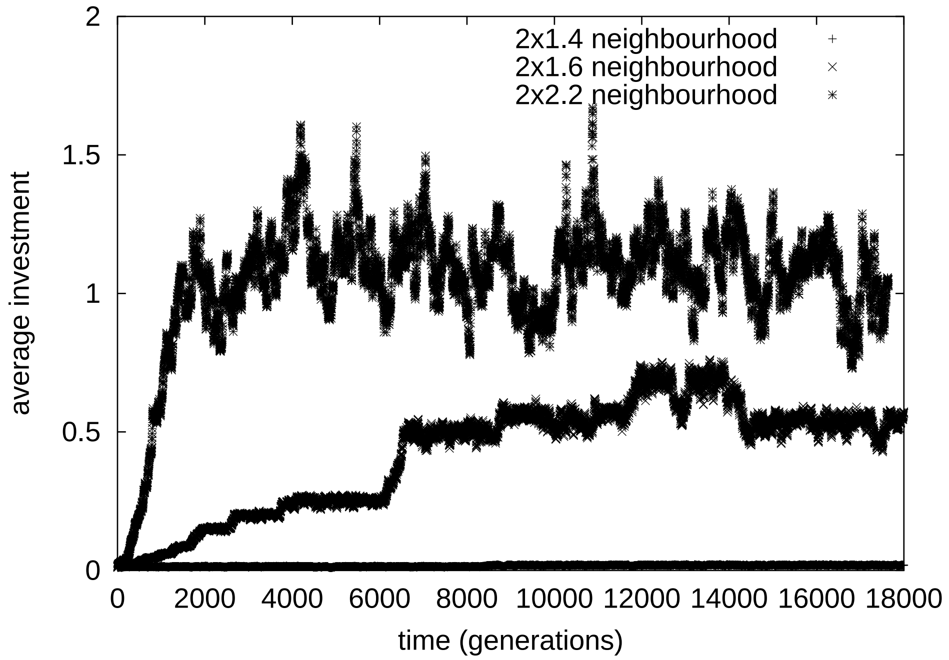

Figure 7.6: When both neighbourhood and dispersal parameters become "continuous", cooperation dies when their difference is about 0.5.

We decided to play the game both ways: first when she does compare her payoff to that of the neighbour with a smaller dispersal parameter, then when she does not.

We initialized both the neighbourhood and dispersal parameters randomly between the values 1 and 4 (which are the interesting values, since for larger parameters there is no sustainable cooperation). The neighbourhood and dispersal parameters are treated as evolutionary phenotypes: the parameters of the neighbour that is doing better are adopted, together with the value of the investment. Both parameters, as well as the investment amount, are allowed to mutate at a rate of 0.01, following a Gaussian curve with a width 10 % of its peak. In both scenarios described above, something interesting happens with the neighbourhood and dispersal parameters: in Fig. 7.7 (a) we plot their initial values, which are uniformly distributed within a square. From the plot obtained at the end of the simulation, we see their values converge closer to the diagonal $n_i = d_i$, as in Fig. 7.7 (b). The convergence of the neighbourhood and dispersal parameters toward each-other corresponds to the development of a cooperative state, after which the average investment continues to sustain relatively high values, almost as high as those initially reported by Killingback et al. [213], but the fluctuations are now huge. A typical time series for the average investment is shown in Fig. 7.8.

7.6 CPD in a Network

As discussed in Chapter 6, for most systems of interacting individuals, the regular lattices, or even the random networks do not provide an adequate model for the observed topological properties. Watts and Strogatz's "small world networks" [112] are clustered structures. A spatial structure is called

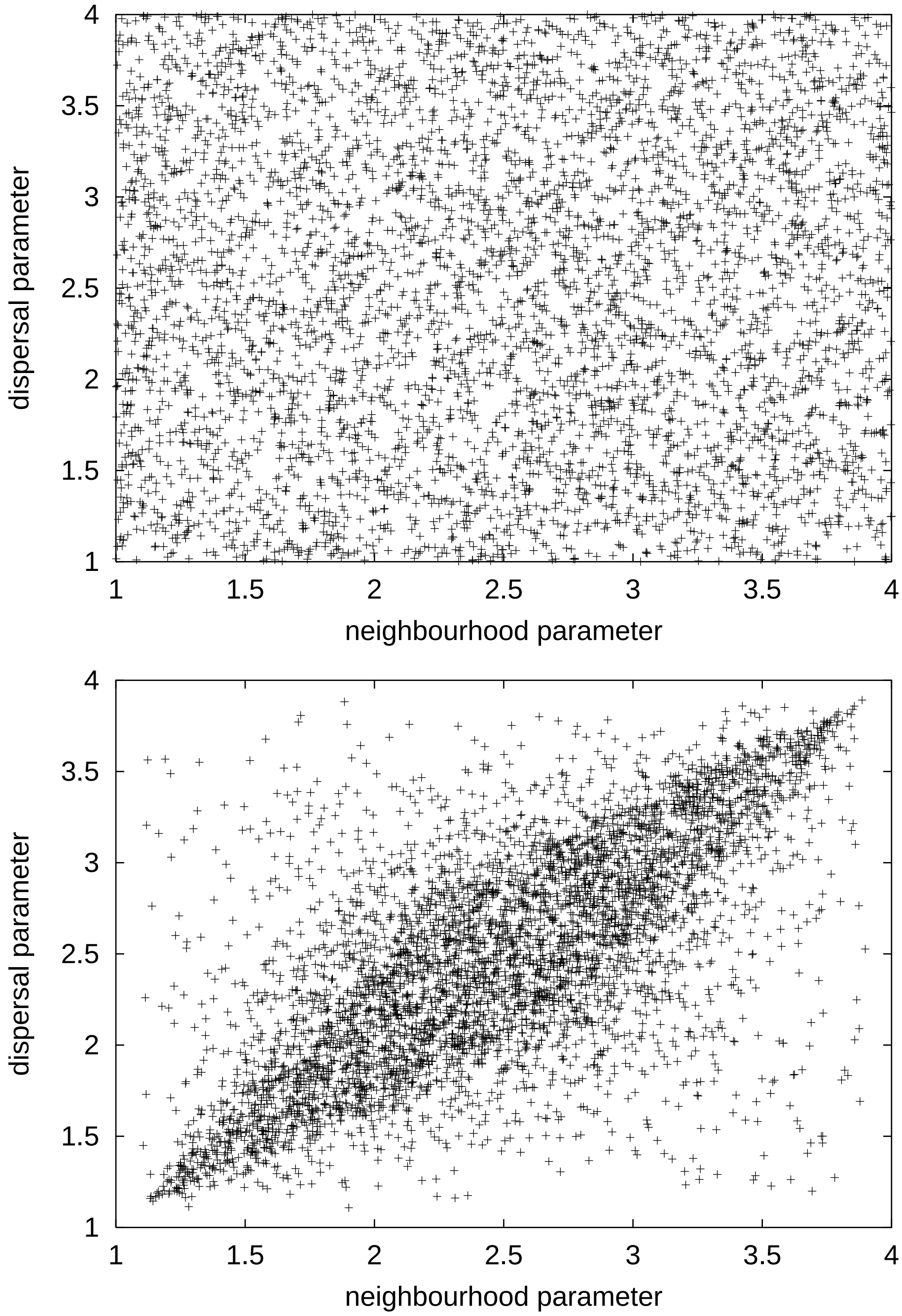

Figure 7.7: (a) At the start of the run the neighbourhood and dispersal parameters are uniformly distributed between the values 1 and 4. (b) At the end, they have converged toward one-another. The distribution of points is closer to the diagonal $n_i = d_i$

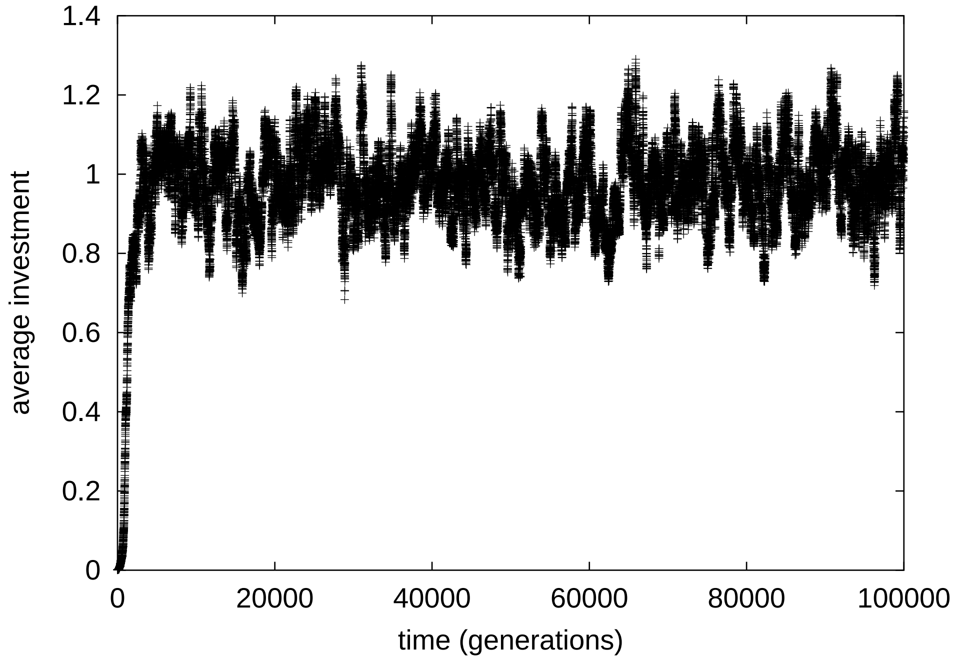

Figure 7.8: The evolution of the average investment when both neighbourhood and dispersal parameters mutate, and are passed on from the more successful individual to the less successful one. The cooperation persists and investment remains in considerable levels.

clustered, if the existence of links between nodes A and B, and B and C, makes it very probable that there be a link between A and C (sociologists call this "transitivity"). The regular lattices are highly clustered structures, but they also exhibit a long path length. The random graphs, in which a node links to any other with a constant probability, have short path lengths between nodes, but exhibit no clustering, i.e. are very distributed structures. A small world has the two properties of high clustering and short average path length between two nodes, which increases with the network size N as the logarithm of N [112, 124]. Social networks are examples of small world networks [29,183,221]. A special class of small world networks are the scale-free ones, in which the degree (i.e. the number of the links a node has) distribution obeys a power law (hence it is independent of network size). Many algorithms for generating scale-free networks have been proposed. This kind of topology is observed in some social networks [29, 127, 128, 182, 183, 221].

The hypothesis that clustering in lattices is the factor that makes possible the establishment and maintenance of cooperation can be checked in networks of different topologies, since some of them exhibit large clustering, and some do not. For that we chose to study the evolution of average investment in social networks, as well as random networks, for comparison. The two topologies were chosen, since they represent extremes when clustering properties in networks are considered. Davidsen, Ebel, and Bornholdt [184, 185] introduced an algorithm that produces small world networks which follow closely the characteristics of the social and acquaintance networks. In their model the connectivity of the network (defined as the ratio of the number of the existing links in the network to the total number of all the possible links) can be tuned by varying the death probability. For the random networks, on the other hand, the connectivity is equal to the ratio of the number of links in the network to the number of all the possible links.

Then the game proceeds as follows: "wake-up" times are generated for each node as $-\ln(rn)$, where rn is a random number with uniform distribution in $[0, 1]$. The picked node then plays against the nodes it can "see", i.e. it is linked to, and calculates her payoff. These neighbours also play against their neighbours, and calculate their respective payoffs. The picked node then adopts the investment amount of the neighbour that is doing best as her new investment amount. The process is iterated, and the average investment is calculated every few "generations". This process is repeated for different network sizes and averaged over 100 different network configurations.

For a fully connected (saturated) network, with connectivity equal to one, the final state will be of pure defection, since this is a well-stirred system, in which "everybody sees everybody". As the connectivity decreases, for each network size there is a certain threshold value, for which the final state ceases to be that of pure defection, but rather a very small average investment. As the connectivity decreases even further, the steady state average investment increases. The variation of average investment with the connectivity for a social network with $N = 6\ 000$ nodes is shown in Fig. 7.9. The transition from the cooperative network to the non-cooperative one is what physicists call a critical transition, i.e. the value of the average investment decays slowly, as opposed to exhibiting a sharp drop at the crossover connectivity value.

For a random network, the same transition is present, but the connectivity for which the transition happens is much smaller: $C_r = 0.003$ (a much smaller value than $C \simeq 0.025$ obtained for the social network). Since the social and random networks differ in their degree of clustering, it seemed logical to check the dependence of the average degree per node at the transition point on the network size for both topologies. The graph is shown in Fig. 7.10.

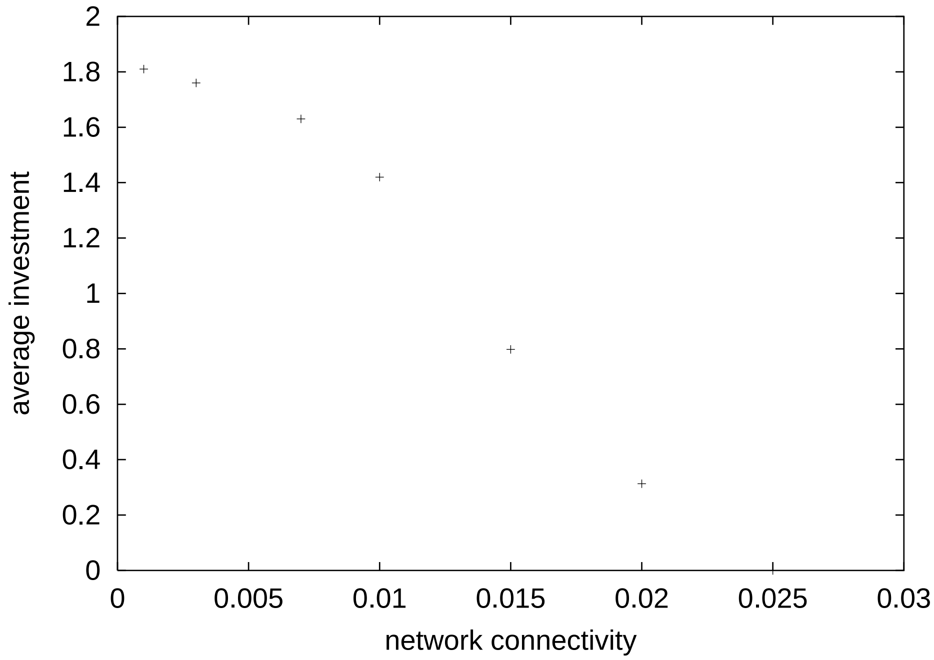

Figure 7.9: The variation of average investment in the steady state with the connectivity for a social network of size $N = 6\ 000$.

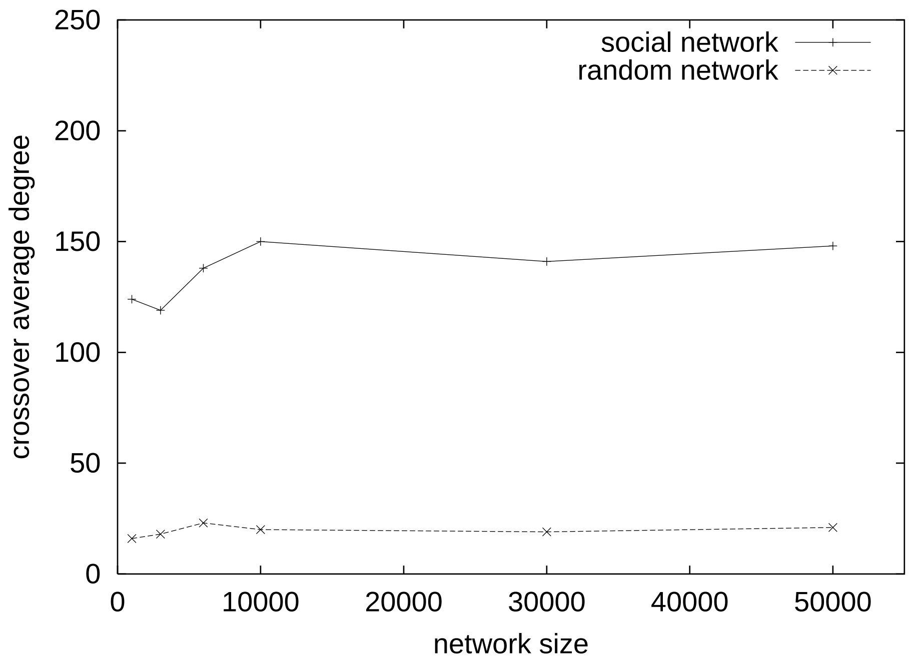

Figure 7.10: The average degree at transition vs. network size. The average degree remains practically constant, but for the social networks it is much higher than for the random networks.

Interestingly enough, for both topologies, the critical average degree remains essentially constant as the network size is varied. It is about 150 for the social network model, and about 20 for the random network. The hypothesis made about clustering as **the** mechanism on which cooperation is first established and then sustained seems to be reinforced by the results obtained on network topologies. Indeed, the social networks, with their pronounced tendency to cluster, help to maintain the cooperative state for a much larger average connectivity (and node degree) than the maximally distributed random networks.

7.7 Conclusions

We have studied the behaviour of the Continuous Prisoner's Dilemma model in different spatial structures, a two-dimensional lattice, as well as networks of different topologies [222]. The Continuous Prisoner's Dilemma is an evolutionary game which is a natural extension of the standard Prisoner's Dilemma. In the Continuous Prisoner's Dilemma each individual makes an investment in favour of herself and the neighbours, and the investment takes on any value, not just two discrete values. In spatial structures like the ones we have considered, the investment levels evolve by getting higher and stabilizing around a given value in a steady state.

The cooperation develops and is maintained relatively easily, which leads us to believe that the cooperation is not such a difficult evolutionary paradox. We have tested this in the case of the square lattice, by increasing the neighbourhood sizes, and found that the neighbourhood size for which the steady state is pure defection is about 100 neighbours. As long as the number of the neighbours the individual plays against (neighbourhood parameter) and that of the neighbours the individual compares herself to (dispersal parameter) remain the same, cooperation is quite robust. The mean–field limit does not depend on the size of the lattice ("world").

When we make the neighbourhoods asymmetric, i.e. the neighbourhood and dispersal parameters differ, we notice that cooperation only develops for differences between them of about 0.5. Too little information favours defection, but so does too much information.

Something interesting happens when we set the neighbourhood and dispersal parameters to mutate, and at the same time, treat them as phenotypes: they are adopted, together with the investment amount of the individual that is doing better. The values of the neighbourhood and dispersal parameters converge toward one-another, i.e. their differences decrease. Cooperation persists, and the average investment amounts fluctuate around relatively high levels in the steady state.

In the network topology, a transition is observed between the cooperative state and the defector one, as the connectivity of the network increases. For both the social networks and the random ones, the average degree at the transition point is practically independent of the network size. However, the average degree at transition for the social networks is about 150, while that for the random networks is much lower (about 20). The two topologies differ in their degree of clustering. The social networks topology, with its pronounced tendency to cluster, helps in the establishment and maintenance of a cooperative state, something the distributed random networks topology cannot do. However, when the connectivity of the network exceeds a certain value (i.e. too many individuals "see" one-another and communicate), cooperation cannot develop or be maintained, and the steady state strategy is pure defection.

In conclusion, our results suggest that defection (selfishness) must be quite rare in spatially structured and localized intraspecific interactions. It is quite reasonable to limit the number of neighbours, as many organisms are spatially distributed and interact much more with the neighbouring individuals, than with those who are far away. This model then would deepen the understanding as to how has cooperation has evolved gradually from lower to higher levels in many spatially structured systems, where the entities involved can vary from replicating molecules to whole organisms [190, 192–194, 223]. It is very encouraging that the cooperation is quite robust in such systems. On the other hand, in the economic world, with the increase in information exchange in modern times, and globalization, the connectivity of the networks becomes quite large, and cooperation will be less likely to persist.

Chapter 8

Concluding Remarks

8.1 Summary of Results and Conclusions

We have considered many-species models in both the cyclic competition and neutral genetic drift versions, and studied the long-term behaviour of such models in the absence and presence of mutations and migrations. This model is relevant, since there exist many ensembles which influence one-another through competition, such as: populations of various biological species, political parties, businesses, countries, coupled chemical reactions, epidemiological systems, components of nervous systems, elementary excitations in matter, etc. Models for these ensembles have been constructed, either from first principles or intuitively. They yield rate equations, which in general are nonlinear, and contain a number of constants, generally unknown ones.

The neutral model is relevant in evolutionary biology. In the 1960s Kimura [17] first combined population genetics theory with molecular evolution data [18] to arrive at a theory using the genetic drift as the main force changing allele frequencies, known as the "Neutral Theory of Molecular Evolution" [19, 20]. In order to compare the evolutionary path due to purely neutral drift to that due to cyclic competition, we would need a minimum model with three alleles present. Our second model is a three-component version of the famous Kimura model of neutral genetic drift [19, 20].

This thesis focused on idealized representation or modelling of the cyclic and neutral version of the three-, four-, and n-species model. We use the Fokker-Planck equation, as well as computer simulations to study the non-spatial model with and without mutations [30, 31, 83], and "experiment" via computer simulation to study different (non-spatial and network) versions of our model. Also, the Continuous Prisoner's Dilemma model has been studied in spatially structured topologies, the mechanisms that influence the development, maintenance, and finally, the breakdown of cooperation in it are explored.

8.1.1 Isolated Three-Species Models

In Chapter 3 we have considered an ABC model in both the cyclic competition and neutral genetic drift versions, and studied the long-term behaviour of both these models. The number of the A, B, C species in the cyclic competition case oscillates with an amplitude that drifts with time, until one of the species (and then the next one in the cycle) goes extinct. In the neutral drift case the number of the A, B, C species drifts, until one of them, and then the other one, become zero. In the cyclic case the amplitude of the oscillations drifts, while in the neutral one the population numbers drift, making the neutral model seem as an adiabatic approximation of the cyclic one.

In both scenarios the number of survivors vs. time plots show an exponential decay, with the same exponent: -3. The extinction time scales as the system size N, at least when the population size is reasonably large. This conclusion is verified by checking the distribution of the product of the concentrations H on the condition that the system be alive. The plots for long enough times are straight lines, indicating a constant probability distribution for the H quantity. This is compatible with the exponential decay. There is no difference in the time scale for the ensemble with cyclic competition and that of neutral genetic drift.

The probability distribution function for the whole system has been reconstructed, based on data from simulations. The eigenfunctions of an equilateral triangle that satisfy the Dirichlet boundary conditions were used as expansion basis, and the expansion coefficients were calculated, using a sample of 3000 copies of the system. The expansion coefficients decay as $1/n$ (with n the index of the eigenfunction), indicating an essentially flat probability distribution function, which was verified by plotting the function. The result of the simulations was verified by writing and solving the Fokker–Planck equation for the neutral model. Finally, the robustness of the exponential decay for the survival probability is checked against variations in the rate of the three different reactions in the system. The slope in this case depends only on the distance of the centre (starting point) from the boundary, i.e. the relationship between the rates of different reactions.

8.1.2 Three-Species Models with Mutations or Migrations

In Chapter 4 we study the model with mutations (migrations) at a constant probability. The system exhibits a critical transition from a “fixation” regime to a “neutral” one, in which biodiversity is preserved over long time. In the "fixation" regime, the system behaves similarly to the isolated system. The order parameter, defined as the product of ABC/N^3 goes to zero, and remains zero, except for occasional "bursts". In the "diversity" regime, the number of the A, B, C varieties exhibits Gaussian fluctuations around the centre point, and there are rare extinctions, but the order parameter remains nonzero almost always. The survival probability decays exponentially below the transition point, but the exponent decreases as the mutation (migration) probability per particle increases, until it becomes zero at the critical point. The critical mutation/migration probability depends on system size as $1/N$, and the models have the same power-law exponent: -1. There is no difference in the behaviours of the neutral system and the cyclic system. Also, there is no qualitative difference between the system with mutations and that with migrations. The simulations results have been verified by expanding the master equation into a nonlinear Fokker-Planck equation, and solving it.

The results thus obtained address the concern about the effects of habitat fragmentation in the survival of the species [66]. If the population of a certain species goes extinct in one patch (e.g. a herd, school, swarm), while it still survives in other patches, then it is hoped that the "rescue effect" can prevent global extinction [67–69]. Otherwise, the species is doomed to go extinct. Our results suggest that the number of mutants/migrants (i.e. the product $\mu \cdot N$) necessary to preserve diversity is independent of system size. On the other hand, in most realistic situations the number of mutations that do happen would be proportional to system size. This means that a large system size favours biodiversity. On the other hand, if one has in mind epidemiological systems, the important factor is to reduce migration probabilities below critical, which is exactly the purpose quarantines serve. The SARS epidemics in Toronto reappeared, once they let their guard down, i.e. infected individuals were allowed to migrate from one community to another. In that case, a large system size may be a disadvantage. It takes only one bad apple...

The study of three-species ecological or epidemiological systems relates to autocatalytic systems with a loop-like structure [82,100]. Methods of analysis employed here can be expanded to the loop-like autocatalytic systems. We report the results of our work regarding the four- or more-species autocatalytic loops [83] in Chapter 5.

8.1.3 Many-Species Models: Autocatalytic Loops

In Chapter 5 we show that the autocatalytic loops are a generalized case of the Lotka-Volterra [85, 86] model. Previously Togashi and Kaneko [100, 101] had observed a transition from the symmetric state to a broken-symmetry state in small size four-species systems. We show that this transition is not specific to systems with a very small number of particles/individuals. It is the same critical transition we have previously [31] observed for three-species systems. The "fixation" regime in the four-species system exhibits a symbiosis effect, when species A_1 and A_3 join their efforts against species A_2 and A_4, and the final state of the system is one in which one of the pairs has completely "eaten up" the other. This is true for any finite system size. In the high-migration regime the system allows for a linear noise approximation, exhibiting itself as an Ornstein-Uhlenbeck process.

Since the system size is always finite (no matter how large it is), there is a value of the migration probability per particle (or diffusion rate), for which the fluctuations of the concentrations become of order one, and the system undergoes a critical transition. The critical diffusion rate varies with system size as $1/N$, and the product $DN \sim 1$, i.e. the number of migrants per unit time, necessary for the symmetry to be preserved in the system (all species to survive) is of the order 1. This result is a bit counterintuitive, since it does not depend on system size. The analytical calculations are in excellent agreement with the simulation results, obtained for four-species systems [100]. Our analytical calculations suggest that the n-species loop-like autocatalytic systems will exhibit the same critical transition. Judging on their form, it is reasonable to expect the generalized Lotka-Volterra systems to exhibit a similar transition.

8.1.4 E-mail Viruses in Internet

Many studies of the cyclic competition system, otherwise known as cyclic Lotka-Volterra model, put the ensemble on a lattice [10–12, 103, 104, 107, 109]. However, the lattice models do not correctly represent the real spatial structure of the interaction networks [29]. In Chapter 6 we "put our model on a network". This way we built a model that relates to the spread of e-mail viruses through the Internet, and also addresses the issue of their survival in the network.

Recently, the rate of mutation appears to have grown, and new generations of viruses have appeared. Hence for many such viruses, a $SIRS$ model, rather than an SIR one, is more appropriate. In our study of the fate of viral infections, we keep in mind that the process of spreading takes place in a social network, for natural viruses, and in an e-mail network, in the case of e-mail viruses. We simulate the behaviour of the ABC model in a network, generated according to the transitive linking and finite lifetime assumptions [184, 185], since it correctly describes the social networks. In each case, the random networks, as well as the scale-free network were used for comparison. For that, a new software (Netgen) was written.

Simulations of the spread and evolution of such systems in different scale-free topologies show that all three types, (A, B, or C) are present in the network at all times. An attempt made in social networks models, shows the presence of a transition to the fixation state for clustering coefficients close to 0.65. This result would relate to the observed peak and disappearance of epidemics in human communities. Test runs performed on random networks shows that the critical connectivities approach 0.4, which is quite an enormous connectivity. Apparently, the computer viral epidemics do not exhibit the same characteristics as human epidemics, because the clustering coefficients for the computer users/e-mail networks cannot reach the critical values. The real e-mail networks are somewhere in between, as far as clustering is concerned, since their clustering coefficients, as measured, do not seem to get high enough to approach the critical values obtained by our simulations. It seems that the pandemics of a viral infection in the Internet is an unlikely event. The Internet is threatened more by the side-effect of clogging from the uncontrolled multiplication of viruses.

8.1.5 Continuous Prisoner's Dilemma and Cooperation

In Chapter 7 we have studied the behaviour of the Continuous Prisoner's Dilemma model in different spatial structures: a two-dimensional lattice, as well as networks of different topologies. The Continuous Prisoner's Dilemma is an evolutionary game which is a natural extension of the standard Prisoner's Dilemma. In the Continuous Prisoner's Dilemma each individual makes an investment in favour of herself and the neighbours, and the investment takes on any value, not just two discrete values. In spatial structures like the ones we have considered, the investment levels evolve by getting higher and stabilizing around a given value in a steady state.

Cooperation develops and is maintained relatively easily under certain conditions. We have tested this, in the case of the square lattice, by increasing the neighbourhood sizes, and found that the neighbourhood size for which the steady state is pure defection is about 100 neighbours. As long as the number of the neighbours the individual plays against (neighbourhood parameter) and that of the neighbours the individual compares herself to (dispersal parameter) remain the same, cooperation is quite robust. The mean–field limit does not depend on the size of the lattice ("world").

When we make the neighbourhoods asymmetric, i.e. the neighbourhood and dispersal parameters are different, we notice that cooperation only develops for differences between them of about 0.5, independent of which one is larger. Too little information favours defection, but so does too much information. Something interesting happens when we set the neighbourhood and dispersal parameters to mutate, and at the same time, treat them as phenotypes: they are adopted, together with the investment amount of the individual that is doing better. Cooperation persists, and fluctuates around relatively high levels in the steady state, even though the local differences between the neighbourhood and dispersal parameters might not be that small. In this case, the parameters converge toward one-another.

In the network topology, a transition is observed between the cooperative state and the defector one, as the connectivity of the network increases. For both the social networks and the random ones, the average degree at the transition point is practically independent of the network size. However, the average degree at transition for the social networks is about 150, while that for the random networks is much lower (about 20). The two topologies differ in their degree of clustering. Social networks, with their pronounced tendency to cluster, help in the establishment and maintenance of a cooperative state, something the distributed random networks apparently cannot do. However, when the connectivity of the network exceeds a certain value (i.e. too many individuals "see" one-another and communicate), cooperation cannot develop or be maintained, and the steady state strategy is pure defection.

8.2 Avenues for further work

One limitation of our study of the many species systems is that we are only considering the effects of birth and death processes, otherwise known as intrinsic noise. However, in the real-world situations, the systems are under the influence of another kind of noise, caused by the environment. The study of the effects of this environmental noise is the next natural step in our study of many-species systems, and that project is already under way.

Another avenue would be to consider the effects of real-world size distribution of habitat patches. Zipf [224] studied the city size distribution, and concluded that it obeys a power-law with exponent $-1/2$. Marsili and Zhang [225] calculated the size distribution one would get when considering different migration reasons (diffusion rates), and concluded that pairwise interaction of individuals is to blame for the Zipf size distribution. It would be interesting to put our many-species system in a world with that particular topology.

Bibliography

[1] Bailey N. T., *The Mathematical Theory of Infectious Diseases*, 2nd edition, Griffin, London (1975).

[2] Anderson R. M. and May R. M., *Infectious Diseases of Humans*, Oxford University Press, London (1991).

[3] Hethcote H. W., SIAM Review **42**, 599 (2000).

[4] Cooke K. L., Calef D. F., and Level E. V., *Nonlinear Systems and its Applications*, Academic Press, New York, (1977).

[5] Longini I. M., Mathematical Biosciences **50**, 85 (1980).

[6] Reeves P., *The Bacteriocins*, Springer Verlag, New York (1972).

[7] James R., Lazdunski C., and Pattus F., (editors) *Bacteriocins, Microcins and Lantibiotics*, Springer Verlag, New York (1991).

[8] Sinervo B. and Lively C., Nature **380**, 240 (1996).

[9] Maynard Smith J., Nature **380**, 198 (1996).

[10] Szabó G., and Czárán T., Phys. Rev. E **63**, 061904 (2001).

[11] Szabó G., Santos M. A., and Mendes J. F. F., Phys. Rev. E **60**, 3776 (1999).

[12] Frachebourg L., Krapivsky P. L., and Ben-Naim E., Phys. Rev. E **54**, 6186 (1996).

[13] Wright S., Genetics **16**, 97 (1931).

[14] Pareto V., *Un applicazione di teorie sociologiche*, Rivista Italiana di Sociologia, 1901, appeared in English as *The Rise and Fall of Elites: An Application of Theoretical Sociology*, Bedminster Press, Totowa (1968).

[15] Pareto V., *Trattato di sociologia generale*, Barbera, Florence (1916), appeared in English as *Treatise on General Sociology*, Dover, New York (1963).

[16] Fisher R. A., Proc. Roy. Soc. Edin. **42**, 321 (1922).

[17] Kimura M., Nature **217**, 624 (1968).

[18] Kimura M., Proc. Natl. Acad. Sci. USA **63**, 1181 (1969).

[19] Kimura M., *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge (1983).

[20] Kimura M., Scientific American **241**, 94 (1979).

[21] King J. L., and Jukes T. H., Science **164**, 788 (1969).

[22] Ohta T., Ann. Rev. Ecol. Syst. **23**, 263 (1992).

[23] Crow J. F., Science **253**, 973 (1991).

[24] Coyne J. A., Barton N. H., and Turelli M., Evolution **51**, 634 (1997).

[25] Rouhani S., and Barton N., Theor. Pop. Biol. **31**, 465 (1987).

[26] Kimura M., and Weiss G. H., Genetics **49**, 561 (1964).

[27] Weiss G. H., and Kimura M., J. Appl. Prob. **2**, 129 (1965).

[28] Duty T., *Ph.D. Thesis*, University of British Columbia (2000).

[29] Albert R., and Barabási A.-L., Rev. Mod. Phys. **74**, 47 (2002).

[30] Ifti M., and Bergersen B., Eur. Phys. J. E **10(3)**, 241 (2003).

[31] Ifti M., and Bergersen B., arXiv:nlin.AO/0307006 (submitted to Eur. Phys. J. B).

[32] Press W. H., Teukolsky S. A., Vetterling W. T., and Flannery B. P., *Numerical Recipes in C: The Art of Scientific Computing* (2nd ed.), Cambridge University Press, Cambridge (1992).

[33] Gammel B. M., Phys. Rev. E **58**, 2586 (1998).

[34] Matsumoto M., and Nishimura T., ACM Transact. Model. Comp. Sim. **8**, 3 (1998).

[35] van Kampen N. G., *Stochastic Processes in Physics and Chemistry* (revised edition), North-Holland (1997).

[36] Risken H., *The Fokker-Planck Equation*, 2nd ed., Springer (1989).

[37] Redner S., *A Guide to First-Passage Processes*, Cambridge University Press, Cambridge (2001).

[38] Kolmogoroff A., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebn. Mathem. Grenzgeb. **2**, no. 3, Springer, Berlin, 1933 == *Foundations of Probability Theory*, Chelsea Press, New York (1950).

[39] Spitzer F., *Principles of Random Walk*, 2nd ed., Springer-Verlag, New York (1976).

[40] Gardiner C. W., *Handbook of Stochastic Methods*, 2nd ed., Springer-Verlag, New York (2002).

[41] Fokker A. D., Ann. Physik **43**, 810 (1914).

[42] Planck M., Sitzber. Preuß Acad. Wiss., 324 (1917).

[43] Gazizov I. Ch., Tulebayev S. D., and Salinas-Martynovskiy Yu. M., Dokl. Akad. Nauk **343(2)**, 179 (1995).

[44] van Kampen N. G., Adv. Chem. Phys. **34**, 245 (1976).

[45] Reichl L., *A Modern Course in Statistical Physics*, Edward Arnold Ltd. (1981).

[46] Kubo R., Matsuo K., and Kitahara K., J. Stat. Phys. **9**, 51 (1973).

[47] Morse P. M., and Feshbach H., *Methods of Theoretical Physics I*, McGraw-Hill, New York (1953).

[48] Horsthemke W., and Brenig L., Zeitschrift für Physik B **27**, 341 (1977).

[49] Horsthemke W., Malke-Mansour M., and Brenig L., Zeitschrift für Physik B **28**, 135 (1977).

[50] Kurtz T. G., J. Appl. Prob. **8**, 344 (1971).

[51] Kurtz T. G., J. Chem. Phys. **57**, 2976 (1972).

[52] Pawula R. F., Phys. Rev. **162**, 186 (1967).

[53] Langevin P. R., C. R. Acad. Sci. (Paris) **146**, 530 (1908).

[54] van Kampen N. G., J. Stat. Phys. **24**, 175 (1981).

[55] Ito K., Proc. Imp. Acad. **20**, 519 (1944).

[56] Stratonovich R. L., *Conditional Markov Processes and Their Application to the Theory of Optimal Control*, Elsevier, New York (1968).

[57] Hethcote H. W., Math. Biosci. **28**, 335 (1976).

[58] Hethcote H. W., *Three basic epidemiological models*, in Applied Mathematical Ecology, L. Gross, T. G. Hallam, and S. A. Levin, eds., Springer-Verlag, Berlin (1989).

[59] Ruijgrok Th., and Ruijgrok M., J. Stat. Phys. **87**, 1145 (1997).

[60] Sato Y., and Crutchfield J. P., arXiv:nlin/0204057, Santa Fe Institute Working Paper 02-04-017.

[61] Ifti M., Li Q., Soukoulis C. M., et al., Mod. Phys. Lett. B **15(21)**, 895 (2001).

[62] Gibson M. A., and Bruck J., J. Phys. Chem. A **104**, 1876 (2000).

[63] Lamé M. G., *Leçons sur la Théorie Mathématique de l'Elasticité des Corps Solides*, Bachelier, Paris (1852).

[64] Pinsky M. A., SIAM J. Math. Anal. **11**, 819 (1980).

[65] Kimura M., Evolution **9**, 419 (1955).

[66] Pimm S. L., Nature **393**, 23 (1998).

[67] Blasius B., Huppert A., and Stone L., Nature **399**, 354 (1999).

[68] Brown J. H., and Kodric-Brown A., Ecology **58**, 445 (1977).

[69] Earn D. J. D., Levin S. A., and Rohani P., Science **290**, 1360 (2000).

[70] Laurance W. F., and Bierregaard R. O. (eds) *Tropical Forest Remnants: Ecology, Management, and Conservation of Fragmented Communities*, Univ. Chicago Press (1997).

[71] Lovejoy T. E. et al. in *Conservation Biology: The Science of Scarcity and Diversity* (ed. Soulé M. E.), Sinauer, Sunderland, MA (1986).

[72] Earn D. J. D., Rohani P., Bolker B. M., and Grenfell B. T., Science **287**, 667 (2000).

[73] Sinervo B, Svensson E., and Comendant T., Nature **406**, 985 (2000).

[74] Earn D. J. D., Rohani P., and Grenfell B. T., Proc. R. Soc. London Ser. B **265**, 7 (1998).

[75] Rohani P., Earn D. J. D., and Grenfell B. T., Science **286**, 968 (1999).

[76] Beier P., and Noss R. F., Conserv. Biol. **12**, 1241 (1998).

[77] Gonzalez A., Lawton J. H., Gilbert F. S., Blackburn T. M., and Evans Freke I., Science **281**, 2045 (1998).

[78] Jackson E. A., *Perspectives of nonlinear dynamics*, Cambridge University Press (1991).

[79] Resnick S., *Adventures in Stochastic Processes*, Springer-Verlag, New York (1994).

[80] Rosenblum M. G., Pikovsky A. S., and Kurths J., Phys. Rev. Lett. **76**, 1804 (1996).

[81] Schafer C., Rosenblum M. G., Kurths J., and Abel H., Nature **392**, 239 (1998).

[82] Jain S., and Krishna S., arXiv:nlin.AO/0210070 (30 Oct 2002).

[83] Ifti M., and Bergersen B., to be published.

[84] Goodwin B. C., *Temporal Organization in Cells*, Academic Press, New York (1963).

[85] Lotka A. J., Proc. Natl. Acad. Sci. USA **6**, 410 (1920).

[86] Volterra V., *Leçon sur la théorie mathématique de la lutte pour la vie*, Gauthier-Villars, Paris (1931).

[87] Goel N. S., Maitra S. C., and Montroll E. W., Rev. Mod. Phys. **43(2)**, 231 (1971).

[88] Volterra V., J. Conseil Permanent Intern. Exploration Mer III; translated in *Animal Ecology*, McGraw-Hill, New York (1928).

[89] Volterra V., Acta Biotheoret. **3**, 1 (1937).

[90] D'Ancona U., R. Comit. Talass. It. Mem. **126**, 95 (1926).

[91] D'Ancona U., *The Struggle for Existence*, Brill, Leiden (1954).

[92] Lotka A. J., J. Phys. Chem. **14**, 271 (1910).

[93] Lotka A. J., *Elements of Mathematical Biology*, Dover, New York (1956).

[94] Kerner E. H., Bull. Math. Biophys. **19**, 121 (1957).

[95] Harary F., *Graph Theory*, Addison Wesley, Reading, MA, USA (1969).

[96] Bang-Jensen J., and Gutin G., *Digraphs: Theory, Algorithms, and Applications*, Springer-Verlag, London (2001).

[97] Seneta E., *Non-Negative Matrices*, George Allen and Unwin, London (1973).

[98] Berman A., and Plemmons R. J., *Non-negative matrices in the mathematical sciences*, SIAM, Philadelphia (1994).

[99] Jain S., and Krishna S., Comp. Phys. Comm. **121-122**, 116 (1999).

[100] Togashi Y., and Kaneko K., Phys. Rev. Lett. **86(11)**, 2459 (2001).

[101] Togashi Y., and Kaneko K., Jour. Phys. Soc. Jpn. **72**, 62 (2003).

[102] Ovchinnikov A. A., and Zeldovich Ya. B., Chem. Phys. **28**, 215 (1978).

[103] Bramson M., and Griffeath D., The Annals of Prob. **17(1)**, 26 (1989).

[104] Frachebourg L., and Krapivsky P. L., J. Phys. A **31**, L287 (1998).

[105] Antal T., and Droz M., Phys. Rev. E **63**, 056119 (2001).

[106] Rozenfeld A. F., and Albano E. V., Phys. Rev. E **63**, 061907 (2001).

[107] Szabó G., and Hauert C., Phys. Rev. E **66**, 062903 (2002).

[108] Szabó G., and Hauert C., Phys. Rev. E **89**, 118101 (2002).

[109] Kerr B., Riley M. A., Feldman M. W., and Bohannan B. J. M., Nature **418**, 171 (2002).

[110] Sinervo B., and Clobert J., Science **300**, 1949 (2003).

[111] Milgram S., Psych. Today **2**, 60 (1967).

[112] Watts D. J., and Strogatz S. H., Nature **393**, 440 (1998).

[113] Albert R., Jeong H., and Barabási A.-L., Nature **401**, 130 (1999).

[114] Faloutsos M., Faloutsos P., and Faloutsos C., Proc. ACM SIGCOMM, Comput. Commun. Rev. **29**, 251 (1999).

[115] Jeong H., Tombor B., Albert R., Oltvai Z. N., and Barabási A.-L., Nature **407**, 651 (2000).

[116] Barabási A.-L., and Albert R., Science **286**, 509 (1999).

[117] Lawrence S., and Giles C. L., Science **280**, 98 (1998).

[118] Lawrence S., and Giles C. L., Nature **400**, 107 (1999).

[119] Kumar R., Raghavan P., Rajalopagan S., and Tomkins A., Proceedings of the 9th ACM Symposium on Principles of Database Systems, 1 (1999).

[120] Broder A., Kumar R., Maghoul F., Raghavan P., Rajalopagan S., Stata R., Tomkins A, and Wiener J., Comput. Netw. **33**, 309 (2000).

[121] Yook S., Jeong H., Barabási A.-L., and Tu Y., arXiv:cond-mat/0101309.

[122] Pastor-Satorras R., Vázquez A., and Vespignani A., Phys. Rev. Lett. **87**, 258701 (2001).

[123] Newman M. E. J., Strogatz S. H., and Watts D. J., Phys. Rev. E **64**, 026118 (2001).

[124] Amaral L. A. N., Scala A., Barthélémy M., and Stanley H. A., Proc. Nat. Acad. Sci. USA **97**, 11149 (2000).

[125] Albert R., and Barabási A.-L., Phys. Rev. Lett. **85**, 5234 (2000).

[126] Newman M. E. J., Proc. Nat. Acad. Sci. USA, **98**, 404 (2001).

[127] Newman M. E. J., Phys. Rev. E **64**, 016131 (2001).

[128] Barabási A.-L., Jeong H., Néda Z., Ravasz E., Schubert A., and Vicsek T., Physica A **311**, 590 (2002).

[129] Redner S., Eur. Phys. J. B **4**, 131 (1998).

[130] Abello J., Pardalos P. M., and Resende M. G. C., DIMACS Series in Disc. Math. and Theor. Comp. Sci. **50**, 119 (1999).

[131] Aiello W., Chung F., and Lu L., in Proc. 32nd ACM Symp. Theor. Comp. (2000).

[132] Liljeros F., Edling C. R., Amaral L. A. N., Stanley H. E., and Aberg Y., Nature **411**, 907 (2001).

[133] Barabási A.-L., *Linked: The New Science of Networks*, Perseus Publishing (2002).

[134] Erdös P., and Rényi A., Publ. Math. Debrecen **6**, 290 (1959).

[135] Erdös P., and Rényi A., Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[136] Erdös P., and Rényi A., Bull. Inst. Int. Stat. **38**, 343 (1961).

[137] Bollobás B., *Random Graphs*, Academic Press, London (1985).

[138] Cohen J. E., Discr. Appl. Math. **19**, 113 (1988).

[139] Stauffer D., and Aharony A., *Introduction to Percolation Theory* (2nd ed.), Taylor & Francis, London (1992).

[140] Chung F., and Lu L., Adv. Appl. Math. (2001).

[141] Newman M. E. J., and Watts D. J., Phys. Rev. A **263**, 341 (1999).

[142] Newman M. E. J., and Watts D. J., Phys. Rev. E **60**, 7332 (1999).

[143] Dorogovtsev S. N., Mendes J. F. F., and Samukhin A. N., Phys. Rev. Lett. **85**, 4633 (2000).

[144] Krapivsky P. L., Redner S., and Leyvraz F., Phys. Rev. Lett. **85**, 4629 (2000).

[145] Krapivsky P. L., and Redner S., Phys. Rev. E **63**, 066123 (2001).

[146] Jeong H., Néda Z., and Barabási A.-L., arXiv:cond-mat/0104131.

[147] Newman M. E. J., Phys. Rev. E **64**, 025102 (2001).

[148] Jeong H., Mason S. P., Oltvai Z. N., and Barabási A.-L., Nature **411**, 41 (2001).

[149] Albert R., Jeong H., and Barabási A.-L., Nature **406**, 378 (2000), Erratum, Nature **409**, 542 (2001).

[150] Watts D. J., *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton (1999).

[151] Huberman B. A., Pirolli P. L. T., Pitkow J. E., and Lukose R. M., Science **280**, 95 (1998).

[152] Kleinberg J. M., Nature **406**, 845 (2000).

[153] Adamic L. A., Lukose R. M., Puniyani A. R., and Huberman B. A., arXiv:cs.NI/0103016.

[154] Burda Z., Correira J. D., and Krzywicki A., arXiv:cond-mat/0104155.

[155] Walsh T., Proceedings of the IJCAI (2001).

[156] Bilke S., and Peterson C., arXiv:cond-mat/0103361.

[157] Coleman J. S., Mendel H., and Katz E., Sociometry **20**, 253 (1957).

[158] Valente T., *Network Models of the Diffusion of Innovations*, Hampton Press, Cresskill (1995).

[159] Johansen A., and Sornette D., Physica A **276**, 338 (2000).

[160] Tadić B., arXiv:cond-mat/0104029.

[161] Pastor-Satorras R., and Vespignani A., Phys. Rev. Lett. **86**, 3200 (2001).

[162] Pastor-Satorras R., and Vespignani A., Phys. Rev. E **63**, 060117 (2001).

[163] Pastor-Satorras R., and Vespignani A., Phys. Rev E **65**, 036104 (2002).

[164] Ahmed E., Hegazi A. S., and Elgazzar A. S., arXiv:nlin.CG/0108013.

[165] Blanchard Ph., Chang C.-H., and Krüger T., arXiv:cond-mat/0207319.

[166] Boguña M., Pastor-Satorras R., and Vespignani A., arXiv:cond-mat/0208163.

[167] Klemm K., and Eguíluz V. M., arXiv:cond-mat/0107606.

[168] Girvan M., and Newman M. E. J., Proc. Natl. Acad. Sci. USA **99**, 8271 (2002).

[169] Newman M. E. J., Phys. Rev E **66**, 016128 (2002).

[170] Watts D. J., Dodds P. S., and Newman M. E. J., Science **296**, 1302 (2002).

[171] Newman M. E. J., arXiv:cond-mat/0303183.

[172] Bezrukov N. N., http://cc.bstu.by/~tsv/bezr/index.htm

[173] Moiseenkov I. E., http://sp.sz.ru/morris_.html

[174] Wolf D., http://emanual.ru/download2/1913.html

[175] Kotov V., http://www.gazeta.ru/flopovod/04-08-1999_virus.htm

[176] Wolf D., http://emanual.ru/download2/2054.html

[177] Portnov E., http://emanual.ru/download2/4748.html

[178] Burger R., http://vx.netlux.org/lib/arb00-2_ru.html

[179] Portnov E., http://emanual.ru/download2/4259.html

[180] Kaspersky Lab Press Release, 19th August, 2003.

[181] Bentley J., Comm. ACM **28**, 245 (1985).

[182] Ebel H., Mielsch L.-I., and Bornholdt S., Phys. Rev E **66**, 035103(R) (2002).

[183] Strogatz S. H., Nature **410**, 268 (2001).

[184] Davidsen J., Ebel H., and Bornholdt S., Phys. Rev. Lett. **88**, 128701 (2002).

[185] Ebel H., Davidsen J., and Bornholdt S., arXiv:cond-mat/0301260, Complexity (to be published).

[186] Capocci A., Caldarelli G., and De Los Rios P., arXiv:cond-mat/0206336.

[187] Hamilton W. D., J. Theor. Biol. **7**, 1 (1964).

[188] Hamilton W. D., J. Theor. Biol. **7**, 17 (1964).

[189] Trivers R., Q. Rev. Biol. **46**, 35 (1971).

[190] Dugatkin L. A., *Cooperation among animals*, Oxford University Press (1997).

[191] Eigen M, and Schuster P., *The hypercycle: a principle of natural self-organization*, Springer, Berlin (1979).

[192] Michod R., Am. Zool. **23**, 5 (1983).

[193] Maynard Smith J., and Szathmáry E., *The major transitions in evolution*, W. H. Freeman & Co., Oxford (1995).

[194] Buss L. W., *The evolution of individuality*, Princeton University Press, Menlo Park, CA. (1997).

[195] Epstein J. M., Complexity **4(2)**, 36 (1998).

[196] Axelrod R., and Hamilton W. D., Science **211**, 1390 (1981).

[197] Axelrod R., *The evolution of cooperation*, Basic Books, New York (1984).

[198] Guttman J. M., J. Econ. Behav. Org. **29(1)**, 27 (1996).

[199] Lindgren K., and Nordahl M. G., Physica D **75**, 292 (1994).

[200] Miller J. H., J. Econ. Behav. Org. **29(1)**, 87 (1996).

[201] Maynard Smith J., *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, MA (1982).

[202] Weibull J. W., *Evolutionary Game Theory*, MIT Press, Cambridge, MA (1995).

[203] Binmore K. G., *Fun and Games: A Text on Game Theory*, D. C. Heath and Co., Lexington, MA (1992).

[204] Frank R. H., *Passions Within Reason: The Strategic Role of the Emotions*, W. W. Norton and Co., New York (1988).

[205] Holland J. H., *Hidden Order: How Adaptation Builds Complexity*, Helix Books, Reading, MA (1995).

[206] Nowak M. A., and May R. M., Nature **359**, 826 (1992).

[207] Huberman B., and Glance N., in *Interdisciplinary Aprroaches to Nonlinear Systems*, Haken H., and Mikhailov A., eds., Springer (1993).

[208] Mar G., and St. Denis P., Int. J. Bifurc. Chaos **4**, 943 (1994).

[209] Doebeli M., and Knowlton N., Proc. Natl. Acad. Sci. USA **95**, 8676 (1998).

[210] Roberts G., and Sherratt T. N., Nature **394**, 175 (1998).

[211] Wahl L. M., and Nowak M. A., J. Theor. Biol. **200(3)**, 307 (1999).

[212] Wahl L. M., and Nowak M. A., J. Theor. Biol. **200(3)**, 323 (1999).

[213] Killingback T., Doebeli M., and Knowlton N., Proc. R. Soc. Lond. B **266(1430)**, 1723 (1999).

[214] Hart B., and Hart L., Anim. Behav. **44**, 1073 (1992).

[215] Wilkinson G., Nature **308**, 181 (1984).

[216] Killingback T., and Doebeli M., Proc. R. Soc. Lond. B **263**, 1135 (1996).

[217] Huberman B. A., and Glance N. S., Proc. Natl. Acad. Sci. USA **90**, 7716 (1993).

[218] Mukherij A., Rajan V., and Slagle J. R., Nature **359**, 125 (1995).

[219] Nowak M., Bonhoeffer S., and May R., Proc. Natl. Acad. Sci. USA **91**, 4877 (1994).

[220] Zimmernann M. G., Eguíluz V. M., and San Miguel M., in *Economics with Heterogeneous Interacting Agents*, Kirman A., and Zimmermann J.-B. (eds), Springer (2001).

[221] Dorogovtsev S. N., and Mendes J. F. F., Adv. Phys. **51**, 1079 (2002).

[222] Ifti M., Doebeli M., and Killingback T., to be published.

[223] Wilson D. S., *The natural selection of populations and communities*, Benjamin Cummings, Menlo Park, CA (1980).

[224] Zipf G. K., *Human Behaviour and the Principle of Least Effort*, Addison-Wesley Press, Inc., Cambridge (1949).

[225] Marsili M., and Zhang Y., Phys. Rev. Lett. **80**, 2741 (1998).