

Linear Stability Theory as an early warning sign for transitions in High Dimensional Complex Systems

Duccio Piovani^{1,2}

E-mail: duccio.piovani@gmail.com

Jelena Grujić^{1,3,4}

E-mail: jelenagr@gmail.com

Henrik Jeldtoft Jensen¹

E-mail: h.jensen@imperial.ac.uk

¹Centre for Complexity Science and Department of Mathematics, Imperial College London, South Kensington Campus, SW7 2AZ, UK

²Centre for Advanced Spatial Analysis, UCL London, 90, Tottenham Court Road, W1T 4TJ, UK

³AI-LAB Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

⁴MLG, Département d'Informatique, Université Libre de Bruxelles, Boulevard du Triomphe - CP 212 1050 Brussels, Belgium

Abstract. We analyse in detail a new approach to the monitoring and forecasting of the onset of transitions in high dimensional complex systems by application to the Tangled Nature model of evolutionary ecology and high dimensional replicator systems with a stochastic element. A high dimensional stability matrix is derived in the mean field approximation to the stochastic dynamics. This allows us to determine the stability spectrum about the observed quasi-stable configurations. From overlap of the instantaneous configuration vector of the full stochastic system with the eigenvectors of the unstable directions of the deterministic mean field approximation we are able to construct a good early-warning indicator of the transitions occurring intermittently.

PACS numbers: 00.00, 20.00, 42.10

Keywords: Forecasting Critical Transitions, Adaptive Systems,

Submitted to: *J. Phys. A: Math. Gen.*

1. Introduction

High dimensional complex systems, both physical and biological, exhibit intermittent dynamics, consisting of stretches of relatively little change interrupted by often sudden and dramatic transitions to a new meta-stable configuration [1]. Such transitions can have crucial consequences when they occur in, say, ecosystems or financial markets and it is therefore important to develop methods that are able to identify precursors, warning signals and ideally techniques to forecast the transitions before they take place. We will expect that the mechanisms behind the rapid rearrangement may be different in different systems.

The literature in this field is very extensive and it is difficult to produce a comprehensive review. Here we relate to the literature most relevant to our paper. An often used approach to forecasting in complex systems was introduced by Scheffer and collaborators [2, 3]. They suggest that critical slowing down and enhanced fluctuations can be used as a precursor of approaching systemic change. The method can work when the high dimensional dynamics of the complex systems can be captured by some few macroscopic collective degrees of freedom. It is expected to be of particular relevance when a slow change in some external parameter which drives the system towards the bifurcation point. In this case transitions are called *critical transitions* and the mathematical framework used comes from dynamical systems theory [4, 5, 6]. This method is pertinent to systems that are dynamically effectively low dimensional in which the transition takes the form of a bifurcation captured by a robust macroscopic variable, which emerges from the micro dynamics. This approach has been applied for many years in many different fields, in climate change [7, 8], in population dynamics [9], in ecosystems [10] and very recently in financial markets [11] just to mention some. Furthermore exploiting the same mathematical framework *flickering* between two stationary states has been found to be a useful indicator [12, 13].

More recently in [14] the authors describe transitions in a different way. In this new interpretation transitions are named *saddle-escape transitions* and metastable states are interpreted as high dimensional saddle points. Transitions are not induced as a result of a change in the external parameters, as is the case in the bifurcation interpretation, but happen due to a rare perturbation which pushes the system towards an unstable direction. An early warning sign is then captured by inferring the value of the largest eigenvalue of the Jacobian through the log difference of a conveniently chosen macroscopic time series.

Here we give the same interpretation of metastable states though we develop a different approach. As we have suggested recently in [15] transitions are induced by intrinsic fluctuations at the level of the individual components which propagate to the macroscopic systemic level and thereby trigger a change in the overall configuration. Our approach is relevant to systems in which the available configuration space evolves as a consequence of the dynamics. One may think of a new and more virulent virus being created through a mutation of an existing strain (e.g. the SARS virus in 2003),

or a new economic agent arriving in the market (e.g. the dot-com bubble in 1997-2000). Contrary to [14], to build our indicator we do not make use of a rigid macroscopic variable to monitor the system; instead we analyse the interactions between the single microscopic components.

In the current paper we elaborate the method we have presented in [15], we discuss the mathematical details both of the mean field approximation and of a Linear Stability Analysis (LSA) and explore in greater detail its forecasting power. As a first test case we consider the Tangled Nature (TaNa) model of evolutionary ecology [16], which has had considerable success in reproducing both macro-evolutionary aspects such as the intermittent mode of extinctions [17] and ecological aspects such as species abundance distributions [18] and species area laws [19]. A much more succinct and schematic discussion was given in [15].

Furthermore, as a new test case, we present results for transitions in a model with a very different type of dynamics, namely a high dimensional replicator with a stochastic element of mutation [20, 21]. This model is based on the replicator-mutator equation which has wide application in many different fields like population genetics [22], evolutionary game theory [23], language evolution [24], etc. Furthermore, it is related to a few different models with wide applications on their own. First it contains the mutation element of the quasispecies equation [25, 26] often used to model the spread of strongly mutating viruses like HIV and Hepatitis C [27] and the frequency dependent element of the replicator equation [28] used in evolutionary game theory [29]. The Lotka-Volterra equation [30] otherwise known as predator-prey model, which was originally developed to describe dynamics in ecological systems and later got wide application in economics [31], has been shown to be equivalent to replicator equation [32]. The Price equation [33], which Hamilton used in his work on kin selection [34] and was later used to describe numerous biological systems, is also equivalent to the replicator equation and the expanded Price equation is equivalent to the replicator-mutation equation [35]. Given the broad relevance of replicator-mutation equation (population dynamics, virus spreading, game theory, financial dynamics, social dynamics etc.), success in forecasting transitions in this model may indicate that our method may be relevant to a range of situations. We begin by demonstrating that the high dimensional replicator system with mutations exhibits intermittent behaviour. Without mutations, the replicator equation will typically not exhibit intermittent dynamics.

Despite their different general mechanisms, the TaNa model and the replicator-mutator system share similar properties. Their stochastic dynamics is characterised by a huge number of metastable states. When the system randomly falls into one of them it enters a quiescent period of little change. Eventually the intrinsic stochastic fluctuations lead to the occupancy of hitherto empty parts of the configuration space which may serve as a random kick able to drive the system away from the metastable configuration and towards the chaotic regime where the system undergoes a high dimensional adaptive walk searching for another metastable point. The two systems studied here do not exhibit the characteristic bifurcation captured by some rigid macrovariable, nor can the

transitions be forecasted by inferring the largest eigenvalue. They cannot be forecasted through the observation of systemic properties but require the analysis of microscopic interactions.

Nevertheless, through a mean field description of the stochastic dynamics we can infer the Jacobian and interpret the metastable states as high dimensional saddle points formed of a vast majority of stable directions and typically a few unstable ones. If the mean field description was accurate, we would be able to make deterministic predictions. This is not the case, but we demonstrate that we are able, in both models, to understand which kind of intrinsic stochastic fluctuation will be able to push the system out of its stable configuration.

As said in this paper we start by elaborating on the procedure presented in [15] checking the performance of an alarm threshold built on the stability indicator. This procedure is to be considered a starting point, its weakness consists in the need of full information on the system (one needs to know the structure of the whole configuration space). To overcome this problem we have developed a new methodology, described at the end of the paper, where we have reduced the amount of dynamical details needed to produce forecasts.

The remaining of the paper is structured as follows: in Section 2 we describe the procedure which combines observational data (in our case from simulations) with a LSA of the mean field dynamics, in Section 3 we will go through the details of the models used as test cases and their mean field description. In Section 4 we analyse the results coming from the forecasting procedure first introduced in [15] applied to both models. In Section 5A to study the robustness of the method, we introduce a level of error in the mean field interaction matrix. Finally in Section 5B we introduce and develop the methodology to make the method more applicable to real world problems.

In order to facilitate collaboration we have uploaded all the codes (C/C++) necessary to produce the results of the paper online (both the models and the forecasting procedure). The interested reader can find and download them from H.J. Jensen's home page <https://wwwf.imperial.ac.uk/hjjens/>.

2. Linear Stability Analysis Application

In this section we give a general outline of our approach. We will then describe the application to the two models in detail in the following two sections. The first step is to establish a mean field approximation of the stochastic dynamics in order to obtain a set of deterministic equations. In order to do so we define the state vector $\mathbf{n}(t) = (n_1(t), \dots, n_d(t))$, where $n_i(t)$ represent the occupation of each node (species or strategy). The mean field time evolution is of the form

$$\mathbf{n}(t+1) - \mathbf{n}(t) = \mathbb{T}(\mathbf{n}(t)) \cdot \mathbf{n}(t) \quad (1)$$

where the matrix \mathbb{T} is the mean field evolution matrix, which will contain contributions from the following processes: death, reproduction and mutation. Obviously in this

framework the fixed point configurations \mathbf{n}^* are given as solutions of

$$\mathbb{T}(\mathbf{n}^*) \cdot \mathbf{n}^* = 0 \quad (2)$$

Because of the high dimensionality of the type of systems we have in mind, Eq.(2) will typically not be solvable analytically but can be numerically approximated observing the stochastic dynamics. We now perform a linear stability analysis about \mathbf{n}^* by introducing a small perturbation $\mathbf{n}(t) = \mathbf{n}^* + \delta\mathbf{n}(t)$. By substituting the perturbed vector in Eq.(1) and expanding the right hand side to first order in $\delta\mathbf{n}(t)$ we get

$$\begin{aligned} \delta\mathbf{n}(t+1) - \delta\mathbf{n}(t) &\simeq (\mathbb{T}(\mathbf{n}^*) + \partial_{\mathbf{n}}\mathbb{T}(\mathbf{n}^*) \cdot \mathbf{n}^*) \delta\mathbf{n}(t) \\ &= \mathbb{M}(\mathbf{n}^*) \cdot \delta\mathbf{n}(t) \end{aligned} \quad (3)$$

where we have used Eq.(2). Here the matrix

$$\mathbb{M}(\mathbf{n}^*) = (\mathbb{T}(\mathbf{n}^*) + \partial_{\mathbf{n}}\mathbb{T}(\mathbf{n}^*)\mathbf{n}^*) \quad (4)$$

is the *Jacobian* of the system, or the *stability matrix*. Now exploiting the results of the LSA, we know that the eigenvectors or *generalised* eigenvectors (in case of a non diagonalizable Jacobian) \mathbf{e}_+ associated with λ with $Re(\lambda) > 0$ indicate unstable directions. These can be identified with *dangerous* components n_i of the configuration vector towards which the eigenvectors point. What this means is that if the stochastic fluctuations (mutations) bring the system close to these unstable directions, by activating the *dangerous* components, the system will suffer a repulsive force that will push it away from the fixed (saddle) point \mathbf{n}^* . In other words the *activation* of one of these components corresponds to a perturbation parallel to an unstable direction of the saddle fixed point \mathbf{n}^* . This implies that a sudden growth of these components would indicate the arrival of a transition.

This observation allows us to identify a stability indicator, whose non-zero values are early warning signalling of an approaching transition caused by the system leaving the vicinity of a current fixed point. The details of this indicator will depend on the specific case we are dealing with but will be based on the same general idea. In the following sections we will present the two test case models analysing their basic mechanisms and results, and developing our mean-field stability indicator in both cases.

3. The models

3.1. A. The Tangled Nature Model

In the TaNa, an agent is represented by a sequence of binary variables with fixed length L [?], denoted as $\mathbf{S}^a = (S_1^a, \dots, S_L^a)$, where $S_i^a = \pm 1$. Thus, there are 2^L different sequences, each one represented by a vector in the genotype space: $\mathcal{S} = \{-1, 1\}^L$. In a simplistic picture, each of these sequences represents a genome uniquely determining the phenotype of all individuals of this genotype. We denote by $n(\mathbf{S}^a, t)$ the number of individuals of type \mathbf{S}^a at time t and the total population is $N(t) = \sum_{a=1}^{2^L} n(\mathbf{S}^a, t)$. We define the distance between different genomes \mathbf{S}^a and \mathbf{S}^b as the Hamming distance:

$d_{ab} = \frac{1}{2L} \sum_{i=1}^L |S_i^a - S_i^b|$. A time step is defined as a succession of one annihilation and of one reproduction attempt. During the killing attempt, an individual is chosen randomly from the population and killed with a probability p_{kill} constant in time and independent of the type. During the reproduction process, a different randomly chosen individual \mathbf{S}^a successfully reproduces with probability $p_{off}(\mathbf{S}^a, t) = \frac{\exp(H(\mathbf{S}^a, t))}{1 + \exp(H(\mathbf{S}^a, t))}$, which depends on the occupancy distribution of all the types at time t via the weight function:

$$H(\mathbf{S}^a, t) = \frac{k}{N(t)} \sum_{\mathbf{S}^b \in \mathcal{S}} \mathbf{J}(\mathbf{S}^a, \mathbf{S}^b) n(\mathbf{S}^b, t) - \mu N(t). \quad (5)$$

In Eq. (5), the first term couples the agent \mathbf{S}^a to one of type \mathbf{S}^b by introducing the interaction strength $\mathbf{J}(\mathbf{S}^a, \mathbf{S}^b)$, whose values are randomly distributed in the interval $[-1, +1]$. For simplicity, and to emphasise interactions, we here assume: $\mathbf{J}(\mathbf{S}^a, \mathbf{S}^a) = 0$. The parameter k scales the interaction strengths and μ can be thought of as the carrying capacity of the environment. An increase (decrease) in μ corresponds to harsher (more favourable) external conditions. The reproduction is asexual: the reproducing agent is removed from the population and substituted by two copies \mathbf{S}_1^a and \mathbf{S}_2^a , which are subject to mutations. A single mutation changes the sign of one of the genes: $S_i^a \rightarrow -S_i^a$ with probability p_{mut} . Similarly to a Monte Carlo sweep in statistical mechanics, the unit of time of our simulations is a *generation* consisting of $N(t)/p_{kill}$ time steps, i.e. the average time needed to kill all the individuals at time t . These microscopic rules generate intermittent macro dynamics. The system is persistently switching between two different modes: the meta-stable states (denoted quasi-Evolutionary Stable Strategies or qESS) and the transitions separating them. The qESS states are characterised by small amplitude fluctuations of $N(t)$ and stable patterns of occupancies of the types (Fig. 1, respectively left and right panel). However, these states are not perfectly stable and configurational fluctuations may trigger an abrupt transition to a different qESS state. The transitions consist of collective adaptive random walks in configuration space while searching for a new metastable configuration and are related to high amplitude fluctuations of $N(t)$. All the results we will present for this model have been obtained fixing the parameters to $L = 8$, $p^{mut} = 0.2$, $p^{kill} = 0.4$, $K = 40$ and $\mu = 0.07$ and have been chosen for computational reasons. Furthermore one can see from the occupancy plot in the right panel how a qESS configuration only occupies a small portion of the total available genome space (blue dots). With this parameter set typically 20-50 nodes of the genome space are active, 5-10 of which being heavily occupied (wild types) out of the 256 available.

3.2. Mean Field Description

In the TaNa model there are multiple sources of stochasticity, namely reproduction, mutations and deaths. Following the procedure outlined above we average over these sources in order to derive a deterministic mean field equation. At each time step with probability p_{kill} a randomly chosen individual is removed from the system, which implies that the occupation number of the species it belongs to decreases by one ($\Delta n_i = -1$).

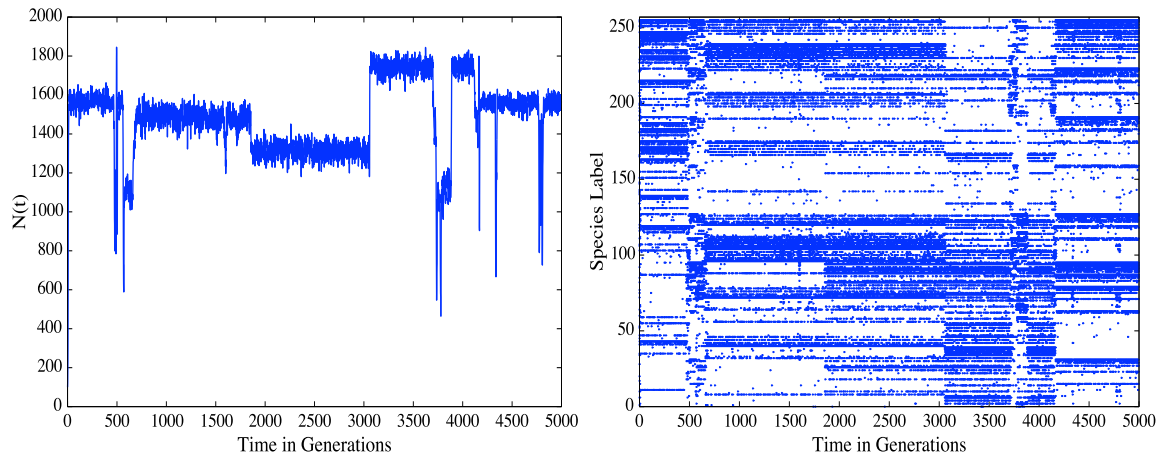


Figure 1: Left panel: total population as a function of time (in generations) for a single realization of the TaNa. The punctuated dynamics is clearly visible: quasi-stable periods alternate with brief periods of hectic transitions, during which $N(t)$ exhibits large amplitude fluctuations. Right panel: occupancy distribution of the types. The genotypes are labelled arbitrarily and a dot indicates a type which is occupied at the time t . These figures are obtained with parameters $L = 8$, $p^{mut} = 0.2$, $p^{kill} = 0.4$, $K = 40$ and $\mu = 0.007$.

Given that the probability of choosing an individual belonging to the i th species is $\rho_i = \frac{n_i}{N}$, the probability that an individual of type i is removed is given by

$$\rho_i \cdot p_{kill} \cdot (-1) \quad (6)$$

The reproduction term is slightly more complicated because mutations need to be treated with some care. A randomly chosen individual is selected for asexual reproduction, which means it is removed from the system while creating two new individuals of the same species. Offsprings can both mutate ($\Delta n_i = -1$), only one can mutate ($\Delta n = 0$), or none mutate ($\Delta n = +1$). Keeping in mind that the probability of reproducing is given by p_i^{off} the average contribution from reproduction of type i including mutations is

$$\rho_i \cdot p_i^{off}(t) [2p_o - 1] = \alpha \cdot \rho_i \cdot p_i^{off}(t) \quad (7)$$

here $p_o = (1 - p^{mut})^L$ is the probability of no mutations and $\alpha = (2p_o - 1)$ is a constant. The third term we have to consider is the *backflow effect*, which describes the event of begin populated by mutations occurring during the reproduction happening elsewhere. This term has the form

$$\sum_j \rho_j(t) p_j^{off} p_{j \rightarrow i}^{mut} \quad (8)$$

Type j will have to mutate a number of genes corresponding to its hamming distance d_{ij} between j and i in order to increase n_i . This will happen with probability

$$p_{i \rightarrow j}^{mut} = p_{mut}^{d_{ij}} \cdot (1 - p_{mut})^{L - d_{ij}} \quad (9)$$

Putting together all these effects we find the expression for Eq.(1) for this model to be

$$n_i(t+1) - n_i(t) = \frac{1}{N} \sum_j \left\{ \left(p_j^{off}(t) (2p_o - 1) - p^{kill} \right) \cdot \delta_{ij} + p_j^{off} \cdot p_{j \rightarrow i}^{mut} \cdot (1 - \delta_{ij}) \right\} n_j(t) \quad (10)$$

where

$$T_{ij} = \left(p_j^{off}(t) (2p_o - 1) - p^{kill} \right) \cdot \delta_{ij} + p_j^{off} \cdot p_{j \rightarrow i}^{mut} \cdot (1 - \delta_{ij}) \quad (11)$$

is the mean-field evolution matrix of the system. By substituting Eq. (11) into Eq. (4) we get the specific form of the stability matrix for the Tangled Nature Model

$$\begin{aligned} \mathbb{M}_{ij} &= (\alpha p_j^{off} - p^{kill}) \delta_{ij} + 2(1 - \delta_{ij}) p_j^{off} p_{j \rightarrow i}^{mut} \\ &+ \sum_k [\alpha \delta_{ik} + (1 - \delta_{ik}) \cdot p_{k \rightarrow i}^{mut}] \frac{\partial p_k^{off}}{\partial n_j} n_k^* \end{aligned} \quad (12)$$

This is the mean field matrix we will use for our linear stability analysis of the stochastic fixed points.

3.3. B. The Replicator Model with Mutations

The replicator equation [28] was introduced in evolutionary game theory in order to capture the frequency dependent nature of the evolution process. Namely, in this model the fitness, or the evolutionary success, of a strategy depends on the frequency of the other strategies in the system. By combining the replicator equation with the quasispecies equation we obtain the replicator-mutation equation, where, apart from the frequency dependence, we also allow for new strategies to enter the system through mutations. As mentioned in the introduction, this model is used to describe numerous high dimensional socio-economic or biological systems.

We are interested in the limit of many strategies. Players may leave the system (say go bankrupt or extinct) or may change their strategy (mutate). This means that the number of players choosing a given strategy and the number of available strategies are in constant evolution. This version of the replicator dynamics set-up was studied by Tokita and Yasutomi in [21]. The authors focused on the emerging network properties. Here we continue this study but with an emphasis on the intermittent nature of the macro-dynamics.

For this model the configuration vector $\mathbf{n}(t)$ contains the *relative frequencies* of all the allowed d different strategies, so the components $n_i(t) \in [0, 1]$ for all $i = 1, 2, \dots, d$. Not all strategies need to be *active* at a given moment, i.e. we can have $n_i(t) = 0$ for some strategy i . We start the simulations by generating the $d \times d$ payoff matrix J of the game that will tell us the payoffs of every pairwise combination. Like for the Tangled Nature model above, the matrix J is a random and fixed interaction network on top of which the replicator dynamics will evolve. Each strategy distinguishes itself from the others in its payoffs or interactions with the rest of the strategy space.

In this chapter we used the same type of uncorrelated interaction matrix as used in the study above of Tangled Nature model. The dimension of the matrix is large, namely

$d \in (10^2, 10^4)$. The qualitative aspects of the behaviour remain the same for other types of payoff matrices. We found that matrices with payoffs uniformly distributed on the interval $(-1, 1)$ or on the set $\{0, 1\}$ exhibit the same behaviour as matrices of the form used for the Tangled Nature model. However, if the payoffs are drawn from a power law distribution with no second moment, the dynamics becomes different and the intermittent behaviour is not so distinct any more.

In the initial configuration, $N_o < d$ strategies start with the same frequency $n_i = \frac{1}{N_o}$. All the other possible strategies are non active, i.e. the corresponding $d - N_o$ components in $\mathbf{n}(0)$ are $n_i(0) = 0$. The empty strategies can only become populated by one of the *active* strategies mutating into them. Once this happens their frequency will evolve according to the replicator equation in which these newly occupied strategies interact with the active strategies which they are linked to through the matrix J .

A time step of the replicator dynamics consists of calculating the *fitness*, $h_i(t) = \sum_j J_{ij}n_j(t)$ of each active strategy and compare it with the average fitness $\bar{h}(t) = \sum_{ij} J_{ij}n_i(t)n_j(t)$, exactly as expected in a replicator dynamics. Each frequency is then updated according to

$$n_i(t+1) = n_i(t) + \left(\sum_j J_{ij}n_j(t) - \sum_{kj} J_{kj}n_k(t)n_j(t) \right) n_i(t) \quad (13)$$

The stochastic element, of the otherwise deterministic dynamics, consists in the following updates. With probability p^{mut} each strategy mutates into another one, this is done by transferring a fraction α_{mut} of the frequency from the considered strategy to another strategy. The label of the latter strategy is chosen in the vicinity of the first by use of a normal distribution $N(i, \Delta)$ centred on label $i \in \{1, 2, \dots, d\}$ with variance Δ with periodic boundary conditions, i.e. label $d+1 = 1$. The closer the labels of two strategies are the more likely it is for one to mutate into the other.

It should be noted that as long as the payoff matrix is random and uncorrelated in its indices, strategies having a small difference between their indices are not necessarily similar since the two strategies may interact with the other strategies in completely different ways. The Δ parameter is introduced to control the level of stochasticity in the system. The higher the values of Δ the greater the range of nodes j that can be activated from a mutation happening in i .

When the frequency of a strategy i goes below a preset extinction threshold $n_i(t) < n^{ext}$, the strategy is considered extinct and its frequency is set to zero $n_i(t+1) = 0$. Right after an extinction event the system is immediately renormalised in order to maintain the condition $\sum_i n_i(t) = 1$.

The dynamics at the systemic level is captured by the time evolution of the occupancy vector $\mathbf{n}(t)$ and is shown in fig.(2), where we present the *occupancy* plot (left panel) and the evolution of the frequencies of the single strategies (right panel).

All the results for this model have been obtained with the same parameter set, namely: $d = 256$, $n_{ext} = 0.001$, $\alpha_{mut} = 0.01$, $p^{mut} = 0.2$ and $\Delta = 15$ that once again have been chosen for reasons of computational performance. In this model the qESS

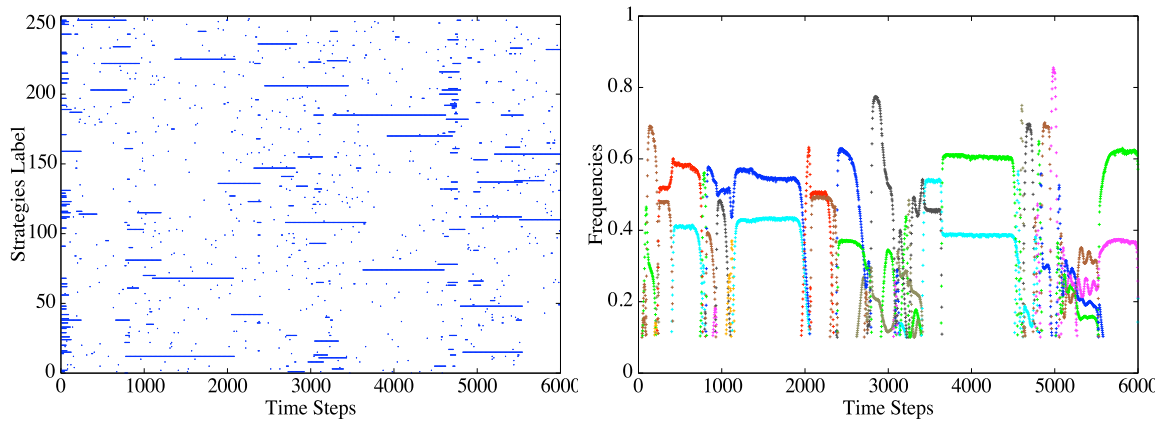


Figure 2: Left panel: occupancy distribution of the types. The genotypes are labelled arbitrarily and a dot indicates a type which is occupied at the time t . The punctuated dynamics is clearly visible: quasi-stable periods alternate with brief periods of hectic transitions. Right panel: the frequencies of the strategies. Each colour belongs to a different strategy. Once again the transitions from one meta stable configuration (approximate fixed point) to another is clear.

are typically characterised by two strongly occupied strategies which are surrounded by 7 to 8 *cloud* strategies that are populated by mutations and quickly die out. So once again the stable configurations occupied only a small part of the entire strategy space.

3.4. Mean Field Description

The random mutations are the only source of stochasticity in the model's dynamics. To account for these stochastic events one has to consider the possibility that a strategy loses part of its frequency by mutating into other strategies or gains frequency as a result of mutations happening elsewhere. As a result a given strategy may lose a fraction of players α_{mut} , which happens with probability p_{mut} or gain $\alpha_{mut} \cdot n_j(t+1)$ which happens with probability $p_{mut} \sum_{j \in N_a} p_{j \rightarrow i}$, where N_a is the number of active strategies and

$$p_{j \rightarrow i} = \frac{e^{-\frac{|i-j|^2}{2\Delta^2}}}{\sqrt{2\pi\Delta^2}} \quad (14)$$

is the probability of i mutating into j (or viceversa). This second effect describes the probability of being populated by a mutation. We therefore get the mean field description as

$$\begin{aligned} n_i(t+1) \simeq n_i(t) + \left(\sum_j J_{ij} n_j(t) - \sum_{jk} J_{ik} n_i(t) n_k(t) \right) \cdot n_i(t) \\ + p_{mut} \alpha_{mut} \left(\sum_j n_j(t) p_{j \rightarrow i} - n_i(t) \right) \end{aligned} \quad (15)$$

which can be expressed, in compact form as

$$n_i(t+1) - n_i(t) \simeq \sum_j \mathbb{T}_{ij} n_j(t) \quad (16)$$

where

$$\begin{aligned} \mathbb{T}_{ij} = & \left(\sum_j J_{ij} n_j(t) - \sum_{jk} J_{jk} n_k(t) n_j(t) - p_{mut} \alpha_{mut} \right) \cdot \delta_{ij} \\ & + p_{mut} \alpha_{mut} p_{j \rightarrow i} \cdot (1 - \delta_{ij}) \end{aligned} \quad (17)$$

The stability matrix is obtained by substituting Eq.(17) in Eq.(4)

$$\mathbb{M}_{ij} = \mathbb{T}_{ij}(\mathbf{n}^*) + \left[J_{ij} - \sum_k (J_{ik} + J_{ki}) \mathbf{n}_k^* \right] \mathbf{n}_i^* \quad (18)$$

4. Procedure and Results

We described in the previous sections how the dynamics of the two models consists in intermittent swift transitions between quasi-metastable configurations. As already stated it is not possible to analytically solve Eq. (2) but we can approximate the fixed points of the mean field dynamics by local time averages over successive configurations in the quasi-stable phases of the full stochastic dynamics, namely: $\bar{\mathbf{n}}^{\text{stoc}} = \frac{1}{T} \sum_{t=0}^T \mathbf{n}(t)$. If the mean field description of the dynamics describes sufficiently well the underlying stochastic dynamics, by substituting the averaged configuration in Eq. (2) we should get $\mathbb{T}(\bar{\mathbf{n}}^{\text{stoc}}) \cdot \bar{\mathbf{n}}^{\text{stoc}} \simeq 0$. We find this to be the case for both models. We have therefore treated $\bar{\mathbf{n}}^{\text{stoc}}$ as our fixed points.

Through our procedure we want to study the stability in the neighbourhood of $\bar{\mathbf{n}}^{\text{stoc}}$, in order to predict the system's reaction to stochastic perturbations. To the extent that the mean field matrix correctly describes the system the metastable states will become unstable along directions in configuration space given by the eigenvectors \mathbf{e}_+ corresponding to eigenvalues with a positive real part $Re(\lambda) > 0$.

Once we know the form of the eigenspace we can monitor two important scalar quantities: the instantaneous distance from the fixed point

$$\delta n(t) = \|\delta \mathbf{n}(t)\| = \|\mathbf{n}(t) - \bar{\mathbf{n}}^{\text{stoc}}\| \quad (19)$$

and the maximum overlap between the perturbation and the eigenvectors $\{\mathbf{e}_+\}$ of the unstable subspace

$$Q(t) = \max \|\delta \mathbf{n}(t) \cdot \mathbf{e}_i\| \quad \forall \mathbf{e}_i \in \{\mathbf{e}^+\} \quad (20)$$

The quantity in Eq.(19) tells us how far away the system is from the fixed point while the overlap in Eq.(20) tells us to what extent a deviation $\mathbf{n}(t) - \bar{\mathbf{n}}^{\text{stoc}}$ is within an unstable sub space. We expect $\delta n(t)$ to fluctuate around a low constant while $Q(t)$ is zero, since this would mean that the perturbations happen in the stable subspace, while a transition would induce a sudden increase in both $\delta n(t)$ and $Q(t)$.

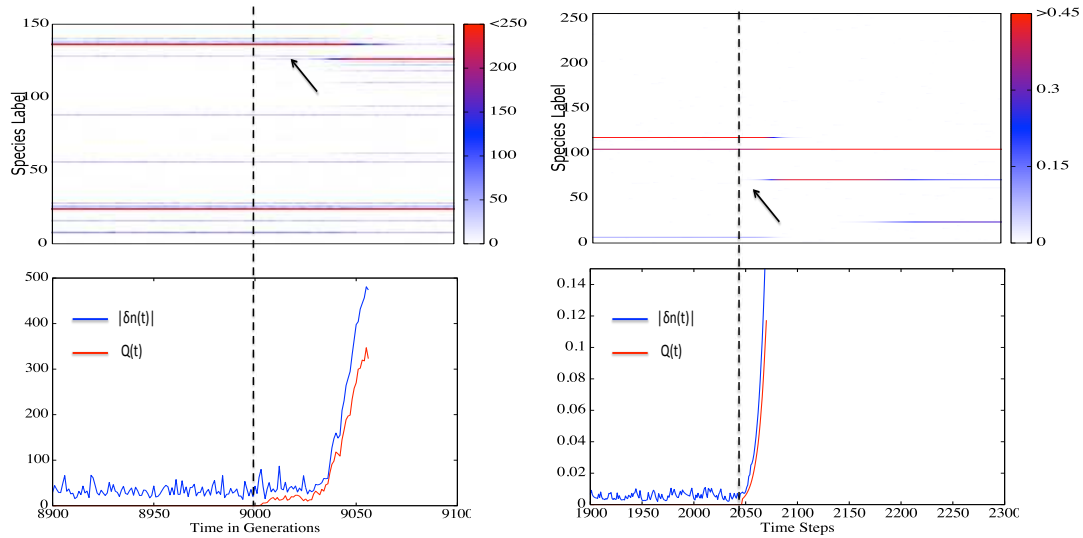


Figure 3: In the bottom panel of both figures we show the behaviour of $\delta n(t)$ (blue curve) and $Q(t)$ (red curve) while approaching the transition in the Tangled Nature (left) and the Replicator Model with Stochasticity (right). In the top panel, a weighted occupation plot is presented. We can see how the beginning of the transitions (dashed vertical black line) is triggered by a new mutant (black arrow) that quickly gains population. The arrival of the new dangerous mutant is singled by a peak in the $Q(t)$.

Another way of picturing $Q(t)$ is as a measure of the activity of the occupancy on dangerous nodes. Indeed every non zero component of the unstable eigenvectors $\{e_+\}$ will tell us which nodes of the interaction network are capable of pushing the system out of its metastable configuration. Namely if $e_+^j \neq 0$, where j indicates the component of the unstable eigenvector, this means the j th node is dangerous. The $Q(t)$ monitors the activity of such nodes. If one of these nodes were to become activated by mutations this would result in a rapid growth of $Q(t)$ and can be considered as a warning of a successive transition. In Fig.(3) of [15] it was discussed how these two quantities behave in the TaNa model and we demonstrated the forecasting power of the indicator $Q(t)$ and we gave an explanation on why we missed some of the transitions. Here we illustrate in Fig.(3) the temporal behaviour of $Q(t)$ and $\delta n(t)$ for both the Tangled Nature Model and the stochastic replicator system. The top panels contain weighted occupation plots while the bottom figures show the behaviour of the two quantities in $Q(t)$ and $\delta n(t)$. The arrow points at the new dangerous mutant that has entered the system, while the dashed vertical line indicates the moment it happens. Before the dashed line we can see how fluctuations in $\delta n(t)$ are bounded and $Q(t)$ is essentially equal to zero. After the dashed line, when the new mutant has entered the system, we see an explosion of both quantities.

We denote t^* the time at which the transition begins, which is set by the $\delta n(t)$ crossing a reasonably chosen threshold T_δ and staying consistently above this threshold (we have used $T_\delta = 150$ for the TaNa and $T_\delta = 0.05$ for the Replicator model). Given the

sharp increase of $\delta n(t)$ when approaching the transition, t^* does not depend strongly on the precise choice of the threshold as long as its is chosen larger than the characteristic fluctuations of $\delta n(t)$ during the metastable configurations.

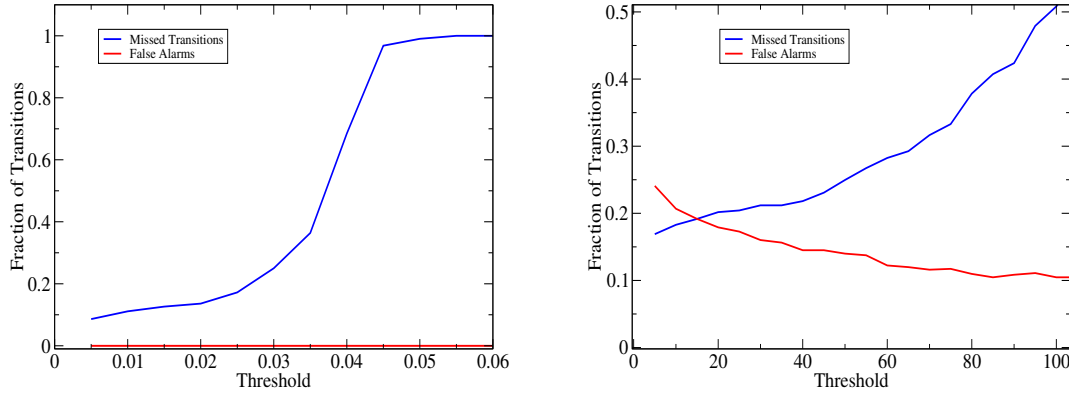


Figure 4: The behaviour of the fraction of false alarms and missed transitions for different values of alarm threshold A_Q in the Replicator Model (Left Panel) and the Tangled Nature (Right Panel). One can see how the procedure produced no false alarms in the Replicator Model which is consistent with what one might expect given the Langevin nature of the model.

To define an alarm we determine an appropriate threshold A_Q on $Q(t)$. To do so we compare the number of false alarms with the number of missed transitions generated by different values of the chosen threshold A_Q . We define a false alarm when the $Q(t)$ crosses A_Q but then goes back under it before any transition occurs. On the other hand a missed transition corresponds to situations where $Q(t)$ remained below A_Q even though the given metastable configuration did become unstable and therefore a transition did occur.

In Fig.(4) we show these two quantities for different A_Q . The red curve is the fraction of missed transitions while the blue is the fraction of transitions that have produced false alarms. In the Tangled Nature model, when increasing A_Q the fraction of false alarms decreases, as expected, while the fraction of missed transitions increases. The same figure for the replicator model shows how the procedure, although missing an increasing number of transitions, produce no false alarms at all.

The reason for this, we believe has to do with the Langevin nature of the dynamics in the Replicator Model, i.e. deterministic dynamics + stochastic noise. Within this approach we expand the configuration vector $\mathbf{n}(t)$ in the \mathbb{M} 's eigenspace or generalised eigenspace plus noise. One gets

$$\mathbf{n}(t) = \sum_k [c_k(0)\exp(\lambda_k t) \cdot \mathbf{e}_k + \epsilon_k] \quad (21)$$

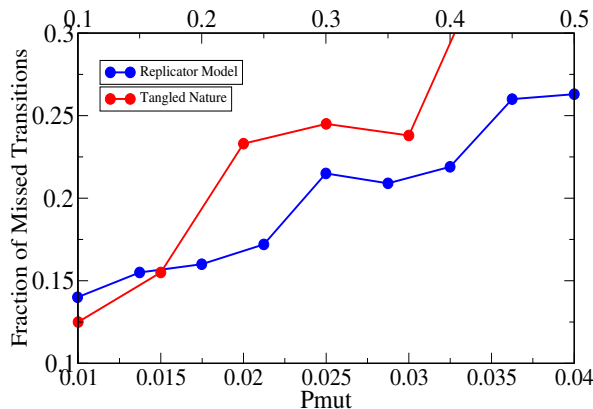


Figure 5: The fraction of missed transitions as a function of the noise in the system for both models. As the stochasticity is increased it becomes harder to forecast transitions in both models. The blue curve is the Replicator Model and it refers to the top x-axis, while the red curve is the Tangled Nature and refers to the bottom x-axis.

where $c_k(0)$ are the coefficients of the expansion and ϵ_k is the noise. This dynamics is clearly dominated by those components for which $\text{Re}(\lambda_k) > 0$, but this is true only if $c_k(0) \neq 0$. When a node is populated by a mutation, in our framework this corresponds to setting $c_k(t) > 0$. From then on the term is suppressed if and only if the ϵ_k points in the opposite direction at all times which is highly unlikely. The same picture is less applicable to the Tangled Nature where all updates are stochastic and hence the separation into a robust deterministic part perturbed by a weak stochastic part is less applicable.

The way to interpret the missed transitions is to think of the fixed points as saddle points of a heterogeneous high dimensional energy landscape. The eigenspace of the mean field matrix tells where the *downhill slopes* and *uphill barriers* are. Although it is far more likely for the system to leave the saddle point through a downhill slope, a stochastic perturbation may be able to push the system over a barrier. This interpretation is confirmed by Fig.(5) where we show that the fraction of missed transitions increases in both models as the degree of stochasticity is increased.

Once the threshold A_Q is fixed, we can determine the time t_{cross} at which $Q(t)$ goes above A_Q and determine the number of time steps $\Delta T = \|t^* - t_{cross}\|$ between the passing of the threshold and the transition as given by the time t^* at which the configuration starts to change significantly. In this way we can check the forecasting power of the indicator. In Fig.(6) we present the distribution of ΔT for $A_Q = 0.01$ and $A_Q = 20$ respectively for the Replicator and the Tangled Nature model. We can see that in the replicator model the crossing times are tenths of time steps before the transition time. This means that the system will go through many cycles of updates before the

transition occurs. In the Tangled Nature in more than 50% of cases $\Delta T \in [2, 5]$. As explained above when introducing the model, one generation corresponds to average number of time steps necessary to remove everyone from the system, i.e. $\frac{N(t)}{p^{kill}}$ individual updates. So even low values of ΔT will involve many individual updates and hence can be considered to correspond to a strong forecasting power.

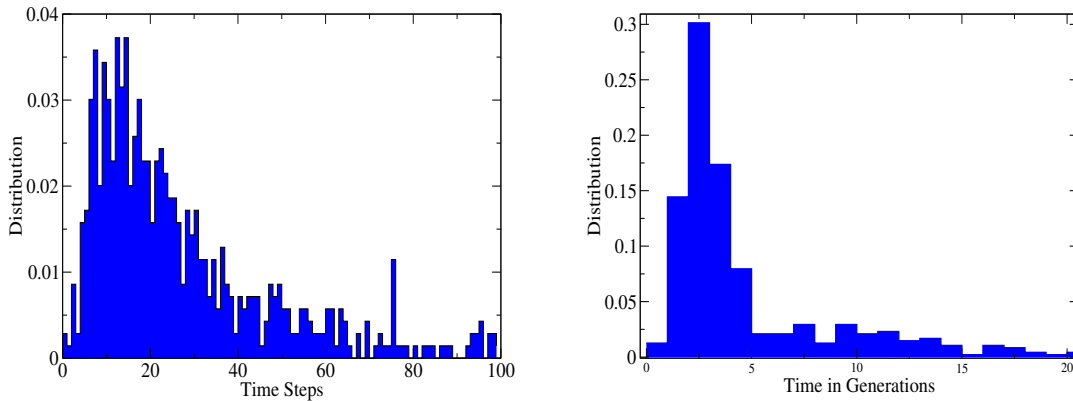


Figure 6: Distribution of the respire of the alarms for a given threshold. The left panel refers to the Replicator model, for which $A_Q = 0.01$ and the right panel to the Tangled Nature model, for which $A_Q = 20$.

5. Incomplete Knowledge

An obvious short coming concerning the real-life application of the forecasting procedure described so far is that we make use of complete knowledge of the entire space of agents and their interactions, i.e we use both the actually realised and the potential part of the space of agents. In this section we first consider how the lack of full knowledge of the interaction strength between agents influences our ability to detect approaching transitions. We next consider a much simpler measure than the overlap function $Q(t)$. This new measure is inspired by the analysis presented above and leading to $Q(t)$ but avoids access to information about the adjacent possible, i.e. information about agents that are not extant in the system at the time of forecasting. Our new measure only makes use of the time evolution of directly observable quantities.

5.1. Error in Interactions

We investigate the effect of lack of complete information concerning the iterations between agents by introducing an error in the interaction matrix used for the mean field treatment. We do this in the following way

$$J_{ij}^e = J_{ij}^{sim} + \chi \quad (22)$$

where χ is $N(0, \sigma)$, i.e. a normally distributed random variable, of mean 0 and variance σ . We then repeat the exact same procedure outlined in the previous section but using \mathbf{J}^e in the forecasting calculations while the simulations evolves according to \mathbf{J}^{sim} .

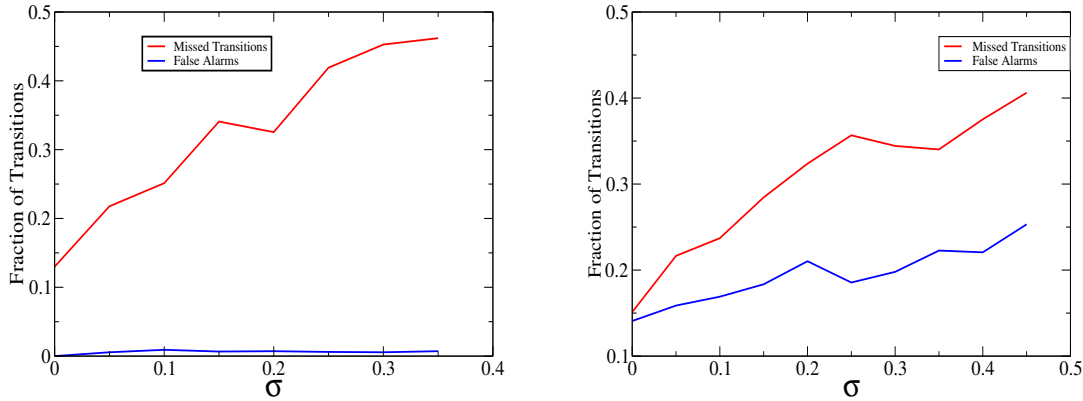


Figure 7: The fraction of the missed transitions and the fractions of false positive as function of the σ of the distribution of the random error in the interactions. Once again we have used $A_Q = 30$ for the Tangled Nature (right panel) and $A_Q = 0.01$ of the Replicator Model (left panel).

In Fig.(7) we present the fractions of transitions we are not able to forecast (missed transition) and the fractions of false alarms we generate as function of the variance σ , i.e. as function of how much the interaction matrix used for the stability analysis differs from the correct set of interactions. For the Tangled Nature (see the right panel) we can notice that for $\sigma < 0.2$ we are still able to forecast around 70% of the transitions and we generate less than 20% of false alarms. This is an encouraging result since a $\sigma = 0.2$ is clearly a significant error given that $J_{ij} \in (-1, 1)$. A very similar result holds for the Replicator Model.

5.2. New Procedure

We now discuss a forecasting procedure that does not need any knowledge about "in potentia" agents. We only need to focus on the highly occupied nodes present in the system. We only know what we see without making any use of the non active part of the interaction network, nor of the poorly occupied nodes. By applying the LSA to the occupied network we can check that, during a stable phase, the configuration corresponds to a situation where the spectrum of the stability matrix \mathbb{M} consists of eigenvalues that all have negative real parts. As the system evolves new mutants appear. As an indicator of approaching transitions we track the growths of the occupancy of these new agents, if their occupancy exceeds a certain threshold we check the spectrum of the updated \mathbb{M} , in which the new agents are included. In case the spectrum now includes positive

eigenvalue we take this as an indicator of, an approaching transition out of the present metastable configuration. This will be our new alarm. That rapidly growing new types may destabilise the configurations of the Tangled Nature Model was also discussed by Becker and Sibani [36]

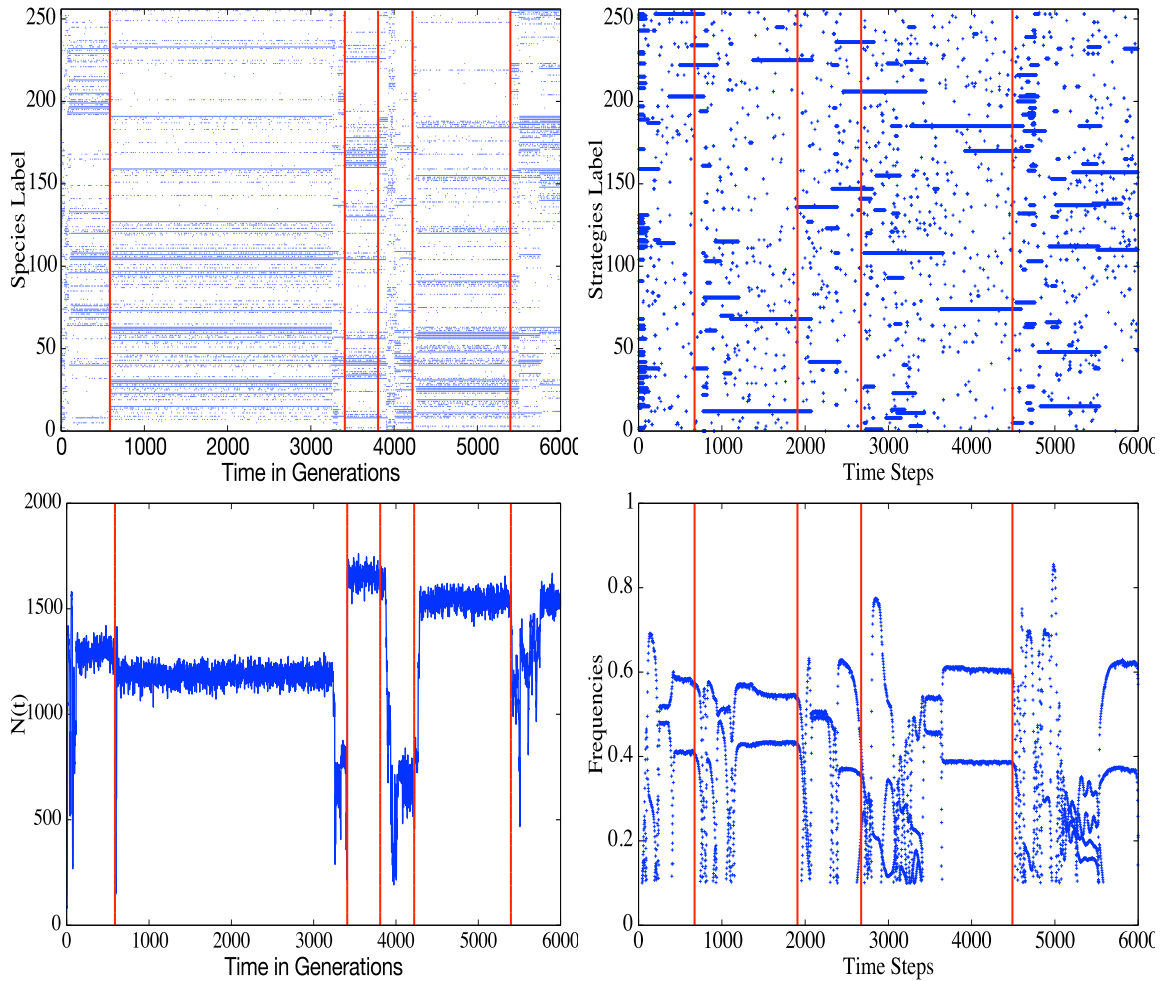


Figure 8: Top left and bottom left respectively occupation plot and total numbers of individual $\sum_j n_j(t) = N(t)$ in the Tangled Nature Model. The vertical red lines represent the alarm times. In the top and bottom right we compare the behaviour of the occupation plot and the frequencies of the most occupied strategies (blue curves) in the Replicator model with the alarms given by our new procedure . One can clearly see how after every alarm the system changes its configuration.

In Fig.(8) we show the results of an application of this new procedure. In both panels the red vertical lines indicate the times of appearance of a species able to change the stability of the system. We can qualitatively see from the figure that just after the alarms the system actually undergoes a transition. In the left panel of the Fig.(8) the blue curves represent the frequencies of the most occupied strategies in the Replicator model. We can see how right after the red lines, the alarm times, a new strategy starts

gaining frequency and eventually puts an end to the stable configuration. It should be stressed that using the growth in the population of the new mutant alone as an early warning signal will not work since it would produce a very high and continuous rate of false alarms. It is the combination of the spectral analysis with the monitoring of the arrival of new mutants that makes the method very powerful.

In the bottom left panel we show the total number of individuals present in the Tangle Nature Model: $N(t) = \sum_j n_j(t)$. A transition to a new metastable configuration is associated with a sudden change of this quantity. We notice that after each alarm $N(t)$ exhibit a significant change induced by the arrival of the new fast growing type. Preliminary analysis indicates that this procedure is able to forecast transitions with an accuracy similar the $Q(t)$ indicator. Further investigation of the efficiency and reliability of using the grows of newcomers as indicators of approaching transitions is underway. Obviously this can make our procedure more readily applicable to real systems since we would then only need directly observable information.

6. Summary and Conclusion

We have describe a new procedure for forecasting transitions in high dimensional systems with stochastic dynamics. Our method is of relevance to systems where the macroscopic dynamics at the systemic level is not adequately captured by a well defined set of essentially deterministic collective variables (e.g. as handled by Langevin equations). Hence we are dealing with situations that are not captured by the application of bifurcation theory such as considered by Scheffer and collaborators [6, 2, 3]. We have in mind complex systems in which the dynamics involves some evolutionary aspects, in particular situations where the dynamics generates new degrees of freedom. E.g. biological evolution, or economical and financial systems, where new agents (organisms, strategies or companies, say) are produced as an intrinsic part of the dynamics. We have demonstrated by use of two models of varying degree of stochasticity (the Tangled Nature Model and the stochastic Replicator Model) that a combination of analytic linear stability analysis and simulation allows one to construct a signal (overlap with unstable directions) which can be used to forecast a very high percentage of all transistons.

The weakness of the procedure is that for real situations of interest (e.g. an ecosystem or a financial market) one may obviously not possess complete information. One will typically not have access to all the information about the interaction amongst the agents. This turns out to be less of a problem, since we can show that even with a 10% inaccuracy in interaction strengths, we are still able to forecast a substantial percentage of transitions. Another short coming is that in real situations it can also be very difficult to know the nature of the new agents that may arrive as the system evolve. Our full mathematical procedure suggests a way to overcome this problem. Namely, the eigenvector analysis showed that transitions are often accompanied by the arrival of new agents, which exhibit a rapid growth in their relative systemic weight. We found that simply monitoring the rapidly growing new agents can enable prediction

of major systemic upheavals. I.e. approaching transitions might not be apparent by focusing on the systemic heavyweights, but rather one should keep a keen eye on the tiny components to monitor whether they suddenly start to flourish. This can often be the signal of upcoming systemic changes.

To address on a real system how crucial incomplete data and limited model accuracy is, our next step is to test our approach on real data streams including high frequency financial time series.

Acknowledgments - This work was supported by the European project CONGAS (Grant FP7-ICT-2011-8-317672). We are grateful for computer time on Imperial College's High Performance Cluster. JG is a Postdoctoral Fellow of the Research Foundation - Flanders.

7. References

- [1] P. Sibani and H.J. Jensen. *Stochastic Dynamics of Complex Systems*. Imperial College Press, 2013.
- [2] Marten Scheffer, Stephen R Carpenter, Timothy M Lenton, Jordi Bascompte, William Brock, Vasilis Dakos, Johan van de Koppel, Ingrid A van de Leemput, Simon A Levin, Egbert H van Nes, et al. Anticipating critical transitions. *Science*, 338(6105):344–348, 2012.
- [3] Marten Scheffer, Jordi Bascompte, William A Brock, Victor Brovkin, Stephen R Carpenter, Vasilis Dakos, Hermann Held, Egbert H Van Nes, Max Rietkerk, and George Sugihara. Early-warning signals for critical transitions. *Nature*, 461(7260):53–59, 2009.
- [4] Christian Kuehn. A mathematical framework for critical transitions: Bifurcations, fast–slow systems and stochastic dynamics. *Physica D: Nonlinear Phenomena*, 240(12):1020–1035, 2011.
- [5] Christian Kuehn. A mathematical framework for critical transitions: Normal forms, variance and applications. *Journal of Nonlinear Science*, 23(3):457–510, 2013.
- [6] Marten Scheffer. *Critical transitions in nature and society*. Princeton University Press, 2009.
- [7] Timothy M Lenton. Early warning of climate tipping points. *Nature Climate Change*, 1(4):201–209, 2011.
- [8] Timothy M Lenton, Hermann Held, Elmar Kriegler, Jim W Hall, Wolfgang Lucht, Stefan Rahmstorf, and Hans Joachim Schellnhuber. Tipping elements in the earth's climate system. *Proceedings of the National Academy of Sciences*, 105(6):1786–1793, 2008.
- [9] Annelies J Veraart, Elisabeth J Faassen, Vasilis Dakos, Egbert H van Nes, Miquel Lüring, and Marten Scheffer. Recovery rates reflect distance to a tipping point in a living system. *Nature*, 481(7381):357–359, 2012.
- [10] SR Carpenter and WA Brock. Rising variance: a leading indicator of ecological transition. *Ecology letters*, 9(3):311–318, 2006.
- [11] Stefano Battiston, J Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. Complexity theory and financial regulation. *Science*, 351(6275):818–819, 2016.
- [12] Rong Wang, John A Dearing, Peter G Langdon, Enlou Zhang, Xiangdong Yang, Vasilis Dakos, and Marten Scheffer. Flickering gives early warning signals of a critical transition to a eutrophic lake state. *Nature*, 492(7429):419–422, 2012.
- [13] Vasilis Dakos, Egbert H van Nes, and Marten Scheffer. Flickering as an early warning signal. *Theoretical ecology*, 6(3):309–317, 2013.
- [14] Christian Kuehn, Gerd Zschaler, and Thilo Gross. Early warning signs for saddle-escape transitions in complex networks. *arXiv preprint arXiv:1401.7125*, 2014.

- [15] A. Caroli, D. Piovani, and H.J. Jensen. Forecasting transitions in systems with high dimensional stochastic complex dynamics: A linear stability analysis of the tangled nature model. *Phys. Rev. Lett.*, 113:264102, 2014.
- [16] K. Christensen, S. A. di Collobiano, M. Hall, and H. J. Jensen. Tangled nature model: A model of evolutionary ecology. *J. Theor. Biol.*, 216:73–84, 2002.
- [17] M. Hall, K. Christensen, S. A. di Collobiano, and H. J. Jensen. Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E*, 66, 2002.
- [18] E. P. Anderson and H. J. Jensen. Network properties, species abundance and evolution in a model of evolutionary ecology. *J. Theor. Biol.*, 232:551–558, 2005.
- [19] D. Lawson and H. J. Jensen. The species-area relationship and evolution. *J. Theor. Biol.*, 241:590–600, 2006.
- [20] Diederich Sigurd and Manfred Opper. Replicators with random interactions: A solvable model. *Physical Review A*, 39:4333, 1989.
- [21] K. Tokita and Yasutomi A. Emergence of a complex and stable network in a model ecosystem with extinction and mutation. *Theoretical population biology*, 63:131–146, 2003.
- [22] KP Hadeler. Stable polymorphisms in a selection model with mutation. *SIAM Journal on Applied Mathematics*, 41(1):1–7, 1981.
- [23] Josef Hofbauer and Karl Sigmund. *Evolutionary games and population dynamics*. Cambridge university press, 1998.
- [24] Martin A Nowak, Natalia L Komarova, and Partha Niyogi. Evolution of universal grammar. *Science*, 291(5501):114–118, 2001.
- [25] M. Eigen. Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58:465, 1971.
- [26] Manfred Eigen and Peter Schuster. A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565, 1977.
- [27] Adam S Lauring and Raul Andino. Quasispecies theory and the behavior of rna viruses. *PLoS Pathog*, 6(7):e1001005, 2010.
- [28] Peter D Taylor and Leo B Jonker. Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1):145–156, 1978.
- [29] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge university press, 1982.
- [30] Alfred J Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3):271–274, 1910.
- [31] Richard M Goodwin. A growth cycle. In *Essays in Economic Dynamics*, pages 165–170. Springer, 1982.
- [32] Josef Hofbauer. On the occurrence of limit cycles in the volterra-lotka equation. *Nonlinear Analysis: Theory, Methods & Applications*, 5(9):1003–1007, 1981.
- [33] George R Price. Selection and covariance. *Nature*, 227:520–21, 1970.
- [34] William D Hamilton. Selfish and spiteful behaviour in an evolutionary model. 1970.
- [35] Karen M Page and Martin A Nowak. Unifying evolutionary dynamics. *Journal of theoretical biology*, 219(1):93–98, 2002.
- [36] N Becker and P Sibani. Evolution and non-equilibrium physics: A study of the Tangled Nature Model. *Europhysics Letters*. 105(1): 18005–11, 2004.