The Tangled Nature Model of evolutionary dynamics reconsidered
structural and dynamical effects of trait inheritance

Andersen, Christian Walther; Sibani, Paolo

Published in:
Physical Review E

DOI:
10.1103/PhysRevE.93.052410

Publication date:
2016

Document version
Early version, also known as pre-print

Citation for published version (APA):
https://doi.org/10.1103/PhysRevE.93.052410

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 09. dec., 2018
The Tangled Nature Model of evolutionary dynamics reconsidered: structural and dynamical effects of trait inheritance.

Christian Walther Andersen and Paolo Sibani
FKF, University of Southern Denmark, DK5230 Odense M, Denmark

The Tangled Nature Model of biological and cultural evolution features interacting agents which compete for limited resources and reproduce in an error prone fashion and at a rate depending on the ‘tangle’ of interactions they maintain with others. Extensive investigations of different versions of this model have shown a number of dynamical features strongly reminiscent of real ecosystem evolution.

The set of interactions linking a TNM individual to others is key to its reproductive success and arguably constitutes its most important property. Yet, in many studies, the interactions of an individual and those of its mutated off-spring are unrelated, a rather unrealistic feature corresponding to a point mutation turning a giraffe into an elephant. The issue has been addressed in the literature and is revisited here to bring out how trait inheritance affects the emergent core structure of the TNM ecology and the aging dynamics of the model.

To this end, we introduce and numerically analyze a family of TNM models where a positive integer $K$ parametrises correlations between the interactions of an agent and those of its mutated off-spring. For $K = 1$ a single point mutation randomizes all the interactions, while increasing $K$ up to the length of the genome ensures an increasing level of trait inheritance. We show that the distribution of the interactions generated by our rule is nearly independent of the value of $K$. Consequently, the structural and dynamical effects of trait inheritance can be studied in isolation, i.e. independently of effects produced by changing the interaction distribution itself. We show that changing $K$ strengthens the core structure of the ecology, leads to population abundance distributions which are better approximated by log-normal probability densities and increases the probability that a species extant at time $t_w$ is also extant at a later time $t$. In particular, survival probabilities are shown to decay as powers of the ratio $t/t_w$, similarity to the pure aging behaviour approximately describing glassy systems of physical origin. Increasing the value of $K$ decreases the numerical value of the decay exponent of the power law, which is a clear quantitative dynamical effect of trait inheritance.

PACS numbers: 87.23.-n, 05.40.-a, 87.23.Kg
I. INTRODUCTION

Several models of biological and social evolution involve networks of interacting agents whose stationary state is interpreted in biological or cultural terms. The Tangled Nature Model, a relatively recent but already well-studied description of evolutionary dynamics, differs as it can either quickly reach a stationary state characterised e.g. by the power-spectrum of population or diversity fluctuations or can enter an aging regime similar in many ways to the dynamics of glassy systems: As pointed out in, a distribution of interactions with finite support leads to stationary behaviour, while aging ensues if ever larger positive interactions become accessible on ever longer time scales. This is the case presently considered, where as seen in, the dynamics is driven through a sequence of macroscopically different metastable states, often called QESS (Quasi Evolutionary Stable States) by a series of non-equilibrium events called quakes.

In many versions of the TNM e.g., two agents whose genomes differ by one bit may have completely different interactions to the rest of the system, even though a successful point mutation would in reality have far less dramatic consequences. To address this issue, Sevim and Rikvold start out with an interaction matrix consisting of Gaussian deviates with variance identical to the uniform distribution previously used by one of these authors. The matrix is then averaged locally over neighborhoods in genome space to produce the desired correlations. These authors find a stationary fluctuation dynamics, resembling that of a model with uncorrelated interactions. We note that their approach requires the storage and manipulation of huge sparse matrices. Laird and Jensen introduce correlated interactions by representing individuals as 16 dimensional vectors in phenotype space. Each element is an integer in the set \{0, 1, ..., 99999\} and contributes additively to the interaction of two species. This leads to a Gaussian interaction distribution where a change of a single element has a minor effect on the total, as desired. The Gaussian distribution of interactions used in both descriptions quickly makes the appearance of 'destabilizers', e.g. mutants receiving very strong positive interactions from extant species, extremely rare, whence the ecology evolves at a pace considerably slower than in Refs.

In the model versions we consider, the couplings between individuals arise here as products of two independent Gaussian variables with zero average and have a probability density function (PDF) described by a modified Bessel function of zeroth order. Apart from a small region near the origin, the latter resembles a Laplace distribution, i.e. \( e^{-|x|/a} \) for some positive constant \( a \). Specifically, we introduce and analyze in the following a one-parameter family of TNM models where an increasing degree of correlation between the interactions of an agent and those of its mutated offspring is obtained by increasing the value of a positive integer parameter, \( K \), where \( K = 1 \) corresponds the model version used in Refs. Choosing one type of interaction distribution over another has structural and dynamical effects which seem to have been overshadowed by issues of numerical convenience and, consequently, have not received sufficient attention in the literature.

The rest of the paper is organized as follows: after summarizing the background and notation, we first explain how the interactions are generated, and estimate, for different values of \( K \), their distribution and the correlation between the interactions of an agent and those of its mutants. We then describe how trait inheritance affects the dynamics, first qualitatively at the level of the emergent core species structure and then, more quantitatively, in terms of the species abundance distribution and the time decay of a cohort of species picked at different stages of the system evolution.
II. BACKGROUND AND NOTATION

TNM agents are binary strings which can either be interpreted as genomes or as cultural features\[13\], i.e. blueprints or strategies for action. Reproduction is asexual and error prone, and its rate depends on the ‘tangle’ of interactions connecting the agents, with positive, or mutualistic, interactions leading to a higher reproduction rate. Removals happen at a constant rate and independently of the interactions. Since extant agents draw resources from a shared and finite pool, they all have an indirect, global and negative effect on each other’s reproductive success. For sufficiently large values of the coupling probability $\theta$, and irrespective of the degree of correlation, a typical TNM ecology comprises a single group of interacting species.

As described in Ref. \[12\], the TNM ecology can be sub-divided into a small group of populous core species and a majority of intermittently populated cloud species. The core is an ordered structure, because core species are inevitably linked by mutualistic interactions dynamically selected from a symmetric distribution. The structure spontaneously emerges in a process driven by an overall increase of configurational entropy. Entropy and order generation are not always antithetic, as usually surmised. In the TNM they grow simultaneously in different parts of the system: core species carry the order and cloud species the entropy. We finally note that the mutualistic interactions between TNM core species do not have a direct interpretation in terms of trophic chains.

For completeness, we now briefly summarize the notation used e.g. in Refs. \[7, 12, 14\]. Time $t$ is measured in generations, each generation comprising the number of Monte Carlo updates needed to remove the whole population: If the population is $N(t)$ at time $t$, $N(t+1)$ is the population after $N(t)/p_{kill}$ updates, where $p_{kill}$ is the removal probability. The length of a generation is thus calculated iteratively based on the population of the previous generation. A species is a group of agents with identical genomes, and agent ‘$a$’ is queried with probability equal to the relative size of its species. When queried, the agent reproduces with probability

$$p_{off}(a) = \frac{1}{1 + e^{-H_a}}, \quad \text{where} \quad H_a = \frac{C}{N(t)} \left( \sum_i J_{ai} N_i \right) - \mu N_i. \quad (1)$$

In the rightmost expression, from left to right, $C$ is a scaling constant, the coupling $J_{ai}$ represents the influence of agent $i$ on agent $a$, $N_i$ is the current population of species $i$ and $\mu$ is a constant expressing the carrying capacity of the environment. Note that $J_{ai} \neq J_{ia}$ and that $J_{ii} = 0$, i.e. self-interactions are excluded. Each bit (gene) in the genome of a newly created off-spring differs from the parental gene with a constant probability $p_{mut}$. A last parameter, $\theta$, determines the probability that two species are connected by non-zero interactions. The parameters used in this work are $\theta = 0.25$, $\mu = 0.05$, $C = 50$, $p_{mut} = 0.01$ and $p_{kill} = 0.25$. Initially, the ecology consists of single species of 500 identical individuals. The system is then given a couple of generations to find a metastable configuration before data collection begins.

The basic procedure to generate the interactions between species ‘$a$’ and ‘$b$’ from the corresponding genomes $G(a)$ and $G(b)$ is as follows: Assuming a binary genome of length $L$, species occupy a $2^L$ dimensional hypercube, and two static arrays of length $2^L$, $A$ and $B$, are initially constructed. The first contains standard independent Gaussian deviates of zero average, and the second contains zeros with probability $1 - \theta$ and ones with probability $\theta$. The two genomes can be read as integer indexes to $A$ and $B$, and so can their logical exclusive or, $c =$XOR$(a, b)$. The procedure to generate $J_{ab}$ is as follows:

1. If $B(c) = 0$ or $a = b$ then $J_{ab} = 0$, else
2. Read out the three Gaussian deviates
\[ z_1 = A(G(a)), \quad z_2 = A(G(b)), \quad \text{and} \quad z_3 = A(G(c)). \]
3. Set \( J_{ab} = z_1 z_3 \) and \( J_{ba} = z_2 z_3 \).

With this algorithm, all interactions are known in potential before the algorithm starts and do not need to be stored in their totality. Computationally this is advantageous, since dynamically generated interactions between extant species only comprise a tiny fraction of the \( 2^{2L} \) possible values of the interaction matrix. The evolution of the ecology in configuration spaces gradually shifts the interaction distribution of extant species toward more positive values \[8, 12, 14\], which is the very feature that underlies the aging dynamics of the model.

As mentioned, when following the above scheme, a single bit change in a genome produces a completely different set of interactions. In the next section we illustrate how the scheme can be improved to allow a mutant to inherit some of the traits of its parent.

III. TRAIT INHERITANCE AND CORRELATED INTERACTIONS

In order to introduce trait inheritance, a gene of length \( L \) is first subdivided into \( K \leq L \) contiguous parts of integer size \( L/K \), each part indexed by an integer \( s \), \( 1 \leq s \leq K \). Secondly, we let \( G_s(a, 0) \) denote the binary string whose \( s \)’th part is identical to the corresponding part of \( G(a) \), and whose other bits are set to zero. To calculate the interactions between \( a \) and \( b \) we repeat the procedure described in the previous section, except that we now read out \( 3K \) standard Gaussian deviates, \( z_{1,s} = A(G_s(a, 0)) \), \( z_{2,s} = A(G_s(b, 0)) \) and \( z_{3,s} = A(G_s(c, 0)) \), and define

\[ J_{ab} = z_1 z_2; \quad J_{ba} = z_1 z_3, \quad \text{where} \quad z_l = \frac{1}{\sqrt{K}} \sum_{s=1}^{K} z_{l,s} ; \quad 1 \leq l \leq 3. \] (2)

We note that all three \( z_l \)'s are standard Gaussian deviates of zero average and that changing one bit in, say, gene \( a \) only affects one of the \( K \) contributions to \( z_1 \). This produces the desired correlations between the interactions of similar genes. We did not implement an analogous procedure to correlate whether an agent and its one-point mutant have similar sets of zero interactions. Finally, since each part of the genome corresponds to \( 2^{L/K} \) integers, only \( K 2^{L/K} \) elements of the array \( A \) are utilized out of the \( 2^L \) available. For \( K \) near \( L \) this introduces undesired statistical correlations between the \( z_{l,s} \) values generated by the algorithm.

The statistical properties of the coupling distribution are summarised in figure 1, which displays an estimate of the PDF of the interaction strengths generated by the rule just described for \( L = 20 \) and two values of \( K \): \( K = 1 \), corresponding to the uncorrelated model (full line) and \( K = 5 \). The data are obtained by generating a million non-zero interactions between random species. Hence they do not include any dynamical selection effects. The theoretical PDF of the product of two standard Gaussian deviates with zero average is also plotted, but is basically hidden behind the \( K = 1 \) curve. As anticipated, the three PDFs have very similar shapes. The \( K = 5 \) data PDF has a slight positive bias, at very high values of its argument. This is an effect caused, as mentioned, by only using \( K 2^{L/K} \) elements of the normally distributed array \( A \): the probability of picking two random numbers with the same sign increases, and when these are multiplied together, the result is always positive.

To describe how the interactions between two agents \( a \) and \( b \) change as \( b \) undergoes mutations, we let \( J_{ab}(m) \), \( m = 0, 1, 2 \ldots \) be the interaction between \( a \) and an \( m \) times mutated \( b \). We then define the correlation function \( C(m) = \langle J_{ab}(0) J_{ab}(m) \rangle_{\{a,b\}} \), where the brackets indicate an average
FIG. 1. The PDF of the interaction strengths for two different values of $K$. The theoretical PDF of the product of two standard Gaussian deviates with zero mean is also plotted but is almost indistinguishable behind the $K = 1$ curve.

FIG. 2. The normalised correlation function for species pair interactions is plotted as a function of the number $m$ of point mutations which one of the two species undergoes.

over different $a$’s and, for each $a$, over the $\binom{L}{m}$ possible ways to introduce $m$ point mutations in $a$ to generate $b$. To estimate the correlation function $C(m)$, $10^4$ starting pairs were used. The result, normalized to $C(0) = 1$, is plotted as a function of $m$ in fig. 2 for different values of $K$. As expected, for $K = 1$ we obtain a Kronecker delta, $C(m) = \delta_{m0}$, while for increasing values of $K$ the correlation decays more and more slowly as a function of $m$. In conclusion, our algorithm ensures that a mutant inherits the interactions of its parent with the typical amount of change tunable using the parameter $K$. 
FIG. 3. A 3-dimensional rendering of the ecology after 10,000 generations. Left, $K = 1$ and right $K = 5$. Species are represented by spheres of volume proportional to their population. The mutualistic interactions connecting core species are represented by lines with the same color as the species they affect and with a thickness proportional to their strength. Interactions linking cloud species are omitted for clarity.

IV. DYNAMICAL EFFECTS

Earlier studies \[9, 16\] suggest that correlated interactions do not change the dynamical behavior of the TNM qualitatively. In this section we reexamine the question for the model version considered, focusing on structural features and aging dynamics. Our first analysis concerns the structure of the core species \[12\] and the log-normal distribution of species abundance. We then proceed to analyze, for different values of $K$, the time dependence of the survival probability of species which are extant at different ages.

A. Core structure

Core species make up the bulk of the population since the mutualistic interactions which link them together endow them with high reproduction rates. Most species belong however to the cloud and are intermittently populated, mainly through an influx of mutants from nearby core species. All together, cloud species only comprise a minor fraction of the population. Following \[12\], a practical criterion which can be used to distinguish cloud from core species on the fly is that a core species is larger than 5% of the most populous species.

Figure 3 shows a three dimensional rendering of a TNM ecology evolved for $10^4$ generations starting from a single species with 500 individuals. The mapping is obtained using Principal Component Analysis, see e.g. \[13\] for further details, and represents species by spheres of volume proportional to their population. Distances reflect the Hamming distances between the corresponding species of the ecology. For $K = 5$ the ecology appears more diverse than in the $K = 1$ case where interactions are uncorrelated. This seems a natural consequence of the fact that, in the former case, a mutant species inherits to some degree the good connections of its parent core species and can establish itself as a new core species. This seems to be the case for the two nearby core species drawn near the
FIG. 4. Each line represents the population of a different species in the frozen ecology obtained by setting the mutation rate to zero after $10^4$ generations. The left and right panels show uncorrelated and correlated interactions, respectively.

FIG. 5. The population abundance distributions vs. the logarithm of the species population for $K = 1$ and $K = 5$, respectively. Lines are gaussian fits. Notice the clear separation of the maxima of the curves on the horizontal scale. Insets zoom in on core species data at large populations.

Lower edge of the right panel. Besides one extra core species, the $K = 5$ version features many more cloud species and a larger total population. To show in more detail the differences in population and diversity induced by changing $K$, simulations were continued for a few generations after forcing the ecology to 'freeze' at $t = 10^4$, by setting the mutation rate $\mu_{mut}$ to zero. The immediate effect of removing mutations, seen in both panels of Fig. 4, is that core species grow in size, due to the fact that all cloud species, no longer replenished by mutants, die out. The long term effect is to force the dynamics into a stationary state by preventing the generation of destabilising mutants. Figure 4 depicts core species populations in such state for $K = 1$ and $K = 5$ as a function of time and shows that turning on the correlations increases both the diversity and the total population of the ecology.
B. Population abundance distribution

Figure 5 illustrates how the population is distributed across different species. The abscissa is the logarithm of the species population size, and the ordinate is our estimate of the corresponding probability distribution function, obtained as the frequency with which a species of given size appears in the population. The data were obtained from an ensemble of 100 independent runs each lasting for 10^4 generations. Comparing the left and right panels, we note that, irrespective of K, the abundance distribution is bimodal, with two widely separated maxima corresponding to values of the abscissa near 2 and 7. The separation supports the distinction between cloud and core species. The insets show that, for K = 5, core species are distributed in a nearly log-normal fashion while, for K = 1, a log-normal distribution still provides a reasonable, but far less accurate, description. We also see that the distribution of the most populous half of the cloud species is also approximately log-normal, while very sparsely populated species fall outside the description. The nearly log-normal distribution of cloud species is expected, as these species are populated by an influx of mutants from 'parent' core species with the size of the influx mirroring the size of the parent species.

Anderson and Jensen [8] already pointed out the relevance for the TNM of log-normal distributions, which are known to describe many natural systems, including abundance distributions in several real ecologies [17]. Our results concur with theirs and furthermore show i) that the distinction between core and cloud species emerges naturally from a statistical analysis and ii) that introducing correlations improves the quality of the log-normal description of the species abundance distribution.

V. SPECIES PERSISTENCE CURVES

Since, by definition, TNM agents die at a constant rate, they have a finite expected life-time. In contrast, and reflecting the complexity of the dynamics, species die at a slow and decelerating rate, and their life-times do not possess finite averages. Studying species persistence curves throws light on important aspect of the TNM dynamics: first, these curves concisely describe the aging dynamics of the model and, second, they change in a measurable and systematic way when inherited traits are introduced.

To explore these issues we define a cohort as the set of species extant at time t_w. We then measure the persistence P(t_w, t) as the fraction of the cohort which is still extant at time t > t_w. The persistence provides an estimate of the probability that a species extant at time t_w still is extant at later times and the distinction between the two quantities is glossed over in the following. The life-time probability density function of a species extant at t_w is then

\[ S(t_w, \tau) = -\frac{d}{d\tau} P(t_w, t_w + \tau) \quad 0 \leq \tau < \infty. \]  

(3)

To measure persistence of species we ran 51 independent simulations in the K = 1 case and 68 in the K = 5 case, all lasting 2^{24} > 10^7 generations, which is a very long time by most criteria. Extant species are registered at each t = 2^k, k = 4, 5, ..., 24, with the delay on the first cohort introduced to ensure that the ecology has properly stabilised. We thus end up with 21 cohorts, one for each k, of which the last one is discarded as it only has one data point. Each data set in figure 6 shows for t > t_w the persistence of the cohort of species extant at time t_w. The dip for t ≈ t_w stems from the fast disappearance of cloud species. At later times the curves all tend to approach straight lines on our log-log plot, which is indicative of a power-law dependence. Since pure or t/t_w scaling is known
FIG. 6. Species persistence versus time. A new cohort of extant species is registered for each $t = 2^k$, $k = 1, 2, 3, \ldots$, and the corresponding fraction surviving at later times is plotted on a logarithmic scale. Not all cohorts are shown for graphical reasons. $K = 1$ data are shown by crosses, and $K = 5$ data by plusses. Lines are only guides to the eye.

FIG. 7. Same data as in figure 6, now plotted as a function of $t/t_w$. $K = 1$ data are shown by crosses, and $K = 5$ data by plusses. The lines are least squares fits to power-laws $y = a(t/t_w)^b$, where $b = -0.283(14)$ and $b = -0.117(6)$ for $K = 1$ and $K = 5$, respectively. All available cohorts have been used in the fits.
to hold approximately in many aging systems of physical nature, see Ref. [18] and references therein, it is interesting to investigate its applicability to a model of biological evolution such as the TNM. Figure 7 shows that plotting our persistence data as a function of $t/t_w$ produces a good data collapse. The lines are fits to a power-law $y(t/t_w) = a(t/t_w)^b$. For $K = 1$ the exponent is $b = -0.283(14)$ and for $K = 5$ it is $b = -0.117(6)$. Three comments are in order: first, independently of the degree of inheritance, Eq. (3) shows that the life-time distribution lacks a finite average. Second, we see that species created at a late stage of the evolution process (large $t_w$) are more resilient than those created early on, implying that the rate of quakes decreases in time. Third, for $K = 5$ inheritance more than halves the $K = 1$ exponent of the persistence decay, clearly showing that inheritance produces a more robust ecology where species live longer.

VI. CONCLUSION AND OUTLOOK

We introduced a version of the TNM that allows a mutant to inherit part of the interactions of its parent, with the amount of modification controlled by a parameter $K$ but where, independently of $K$, the genome remains a point in an $L$ dimensional hypercube. Our approach does not require the storage and manipulation of huge sparse matrices and only leads to insignificant changes of the distribution of the non-zero couplings linking different species. We concur with Refs. [9, 16] that introducing trait inheritance does not radically change the basic properties of the TNM: irrespective of the value of $K$, we see a decelerating aging dynamics where sudden quakes lead to considerable rearrangements of the network structure of the TNM ecology. More specifically, we have shown that independently of $K$ the aging dynamics of the TNM is characterized by $t/t_w$ scaling of species persistence curves which is the a scaling form common in aging systems of physical origin [15]. Increasing the degree of inheritance has nevertheless both structural and dynamical effects: The subnet of core species becomes larger and more highly populated, the species abundance distribution is better approximated by a log-normal distribution, and the decay of the species survival probability becomes noticeably slower, together with the rate of quakes which cause the demise of core species.

In conclusion, adding trait inheritance renders the emergent structural and dynamical properties of the TNM model closer to reality and improves the model’s applicability to the evolution of ecologies, be they of biological or socio-economic nature.

ACKNOWLEDGMENTS

PS would like to thank Peter Salamon for motivating part of this work and Rudy Arthur for several stimulating discussions on TNM dynamics.


