entropy

MDPI

Article

Entropy in the Tangled Nature Model of Evolution

Ty N. F. Roach [1], James Nulton [2], Paolo Sibani [3], Forest Rohwer [1] and Peter Salamon [2,*]

1 Department of Biology, San Diego State University, San Diego, CA 92182, USA; smokinroachjr@gmail.com (T.N.F.R.); frohwer@gmail.com (F.R.)
2 Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA; jnulton@mail.sdsu.edu
3 Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, Odense M 5230, Denmark; paolo.sibani@gmail.com
* Correspondence: salamon@sdsu.edu; Tel.: +1-619-594-7204

Academic Editor: Samuel A. Cushman
Received: 25 February 2017; Accepted: 19 April 2017; Published: 27 April 2017

Abstract: Applications of entropy principles to evolution and ecology are of tantamount importance given the central role spatiotemporal structuring plays in both evolution and ecological succession. We obtain here a qualitative interpretation of the role of entropy in evolving ecological systems. Our interpretation is supported by mathematical arguments using simulation data generated by the Tangled Nature Model (TNM), a stochastic model of evolving ecologies. We define two types of configurational entropy and study their empirical time dependence obtained from the data. Both entropy measures increase logarithmically with time, while the entropy per individual decreases in time, in parallel with the growth of emergent structures visible from other aspects of the simulation. We discuss the biological relevance of these entropies to describe niche space and functional space of ecosystems, as well as their use in characterizing the number of taxonomic configurations compatible with different niche partitioning and functionality. The TNM serves as an illustrative example of how to calculate and interpret these entropies, which are, however, also relevant to real ecosystems, where they can be used to calculate the number of functional and taxonomic configurations that an ecosystem can realize.

Keywords: configurational entropy; ecology; evolution

1. Introduction

The notions of entropy which have been put forward for use in ecology [1–3] are based on configurational, topological, and temporal heterogeneity of an ecosystem and/or on the uncertainty linked to its description. Thermodynamic entropy has been used in biology to quantify the efficiency of biological processes [4]. Additionally, statistical entropies have been used in biology to characterize the spatial distribution and structuring of ecologies [5,6], as well as to quantify the amount of complexity, organization, and order that emerge from evolutionary processes [7–10]. That more ordered structures gradually emerge in an evolving ecosystem is clear, but whether its entropy should increase or decrease in time remains an open question. As argued by Landsberg [11] long ago, disorder and entropy are not necessarily coupled, and increasing order can emerge along with increasing entropy. Herein we define two types of configurational entropy and investigate their behavior as well as the emergence of structure over time in an evolving in silico ecosystem. Computational models offer simplified descriptions of biological phenomena that can reveal generic properties of evolving systems through simulations and analyses. A case in point is Kauffman's NK and NKC models [12,13], where agents evolve in a fixed fitness landscape, and where uphill climbs are generated by mutations and by the competitive advantage of fitter individuals. In practice, at any given time, agents mainly inhabit

one local maximum of the landscape and nearby genome space is searched via mutation and drift. When a better maximum is found, the population rapidly shifts to that maximum, leading to an intermittent coarse-grained dynamics similar to punctuated equilibrium [14]. Furthermore, the models capture the decelerating rate of evolution and the red queen dynamics seen when interacting species are present, but do not capture that evolving species have a changing environment, a fact which drives much of natural selection. The latter aspect is central in a more recent computational model of evolution, the Tangled Nature Model (TNM). Introduced more than a decade ago by Christensen, Jensen and collaborators [15,16], the TNM has attracted attention ever since, see, e.g., [17–23] and references therein. A key feature of the TNM ecology is a growing network of extant species with increasingly mutualistic, i.e., positive interactions. Since these interactions are dynamically selected from a completely unbiased distribution, the degree of order appears to increase over time in our model ecology, precisely as it does in real ecosystems. This property makes the TNM highly relevant for a discussion of how order and entropy can evolve together in an ecology.

Utilizing data from TNM simulations, we study two different measures of configurational entropy, which provide different levels of description of the ecosystem. We find that these entropies both grow logarithmically in time. The per capita entropies, on the other hand, both decrease in time reflecting the structuring of the community. (See Figure 6 below.) The configurational entropies are interpreted in terms of the available niche space in the ecology and the possible taxonomic configurations available to fill these functional niches.

2. The TNM

Similar to the NK and NKC models, the Tangled Nature Model [15,16,19,24] (TNM) of biological evolution features individual agents, which reproduce asexually with the possibility for mutation. Agents are represented by binary strings, for convenience called *genomes*, and groups of agents with identical genomes are called *species*. The salient features of the model are summarized in Figure 1.

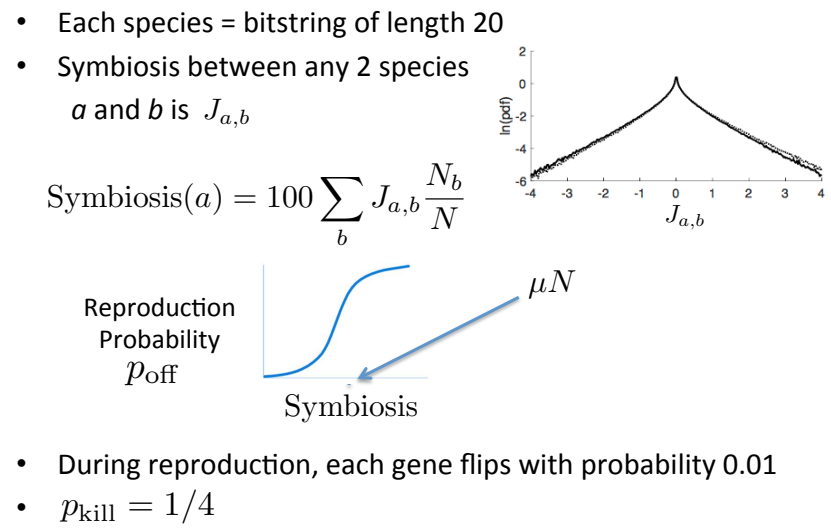

Figure 1. Summary of the Tangled Nature Model (TNM) showing our parameter choices. The top graph shows the probability density function (pdf) of the symbiosis between two random species. The formula shows the overall symbiosis of a species as the average of its pair symbioses. This average determines the likelihood that an individual in species *a* reproduces.

In all of the above models, each dynamical trajectory is characterized by periods of metastability of increasing duration, referred to as quasi evolutionary stable states (QESSs), which mimic the punctuated equilibria of biological macroevolution [25]. During a QESS, macroscopic quantities such as the total population fluctuate slightly around fixed plateau values. In the NK model, the reproductive

success of an individual is uniquely determined by its genome, and the vast majority of the extant population consists of identical and equally fit individuals perched on a local fitness maximum of the configuration space, aptly called the *fitness landscape*. Sudden changes are induced by mutants with higher fitness whose progeny find a higher peak in the fitness landscape and quickly dominate the population. In the TNM, reproductive success of an individual is not uniquely determined by its genome but depends on the "tangle" of symbiotic interactions connecting it to others, with positive interactions leading to a higher reproduction rate. The transitions between consecutive QESSs are rapid and turbulent events, called *quakes*, which entail considerable rearrangements of the network structure [22,23].

Specifically, extant species in the TNM can be grouped into two groups: a few highly populated *core* species and numerous sparsely populated *cloud* species [22]. By definition, core species have a population exceeding 5% of the most populous species, and all species not belonging to the core belong to the cloud. The metastable group of core species owes its reproductive success to mutualistic symbiotic interactions. Cloud species are sparsely and intermittently populated, mainly by an influx of mutants from the core. The interactions between cloud species are distributed randomly according to the distribution for all species. Each QESS in a TNM trajectory is characterized by a core, which is in most cases completely renewed during a quake. At a coarse grained level, a TNM evolution consists of a series of QESSs, each with a different core. As the system ages, cores, on average, gradually have stronger (i.e., more positive) mutualistic interactions, and both the number of core species and the total population grow in the process. Positive interactions are responsible not only for the stability of a core, but also for its eventual demise; occasionally, a mutant in the cloud is linked to the core by strong positive interactions, and, as a consequence, its offspring grow rapidly. This destabilizes the core, and induces a quake [22]. "Old" cores are those able to coexist with large clouds which do not threaten their survival. To summarize, an increasingly structured network of core species can be identified in the evolution of the TNM. This evolving structure is exploited below to count the number of microscopic configurations that underpins them.

The TNM implementation used in our simulations is identical to that described in Reference [22], which should be consulted for more information on the model's phenomenology. The key details are, for convenience, summarized in Appendix A. The nomenclature and notation are given below:

- The TNM's elementary variables are binary strings of length K, i.e., points of the K-dimensional hypercube. Each point represents a genome and is populated by a varying number of individuals or agents. This number is the abundance of the corresponding species.
- The probability p_{kill} that a queried agent is removed is the same for all agents.
- Simulation time is given in *generations*, each comprising the number of updates needed to remove all extant individuals. Specifically, when the initial population is N, the first generation comprises N/p_{kill} updates. The length of subsequent generations is computed similarly, but with N denoting in each case the population present at the end of the preceding generation.
- The reproduction probability p_{off} of an agent is the same within each species, and depends for species a on the average symbiosis species a enjoys with other extant species.

3. Structure and Entropy in the TNM

3.1. The Two Entropies

In the following, we deal with two ostensibly distinct notions of entropy, which nevertheless exhibit an interesting empirical relation, the basis for which is speculated upon in the discussion section.

Entropy 1:

Our first configurational entropy is to use the number of configurations in the most likely realization of the species abundance distribution (SAD). This gives simply the number of species times the Shannon information of the SAD, familiar to ecologists [3,26]. Its selling feature is its experimental

accessibility for real ecosystems. It is included here because it mirrors closely the behavior of our second configurational entropy.

Note that this is not the Shannon–Wiener entropy of the rank abundance distribution (RAD), which is often used as a measure of diversity.

Entropy 2:

The second configurational entropy arises in an analysis of the structure of a QESS in the TNM model. It is a configurational entropy derived from a count of the number of ways the progeny of the core can be produced within the space of potential species in a way consistent with the reproduction and mutation characteristics of a QESS in the TNM.

3.2. Simplified QESS Structure

We assume a slightly simplified picture, taking the structure to be

- n core species
- each core species is surrounded by a cloud of its mutants
- only core species reproduce and mutate; cloud species never reproduce
- each cloud-core distribution of individuals is spread out to k point mutations according to a binomial distribution (described below).

These assumptions, although simple, give a reasonable approximation to the real structure of a QESS. They suggest a simple probabilistic model (see Figure 2) for reproduction and mutation during a QESS. Our purpose here is to clarify the structure of a QESS enough to make possible the definition of Entropy 2.

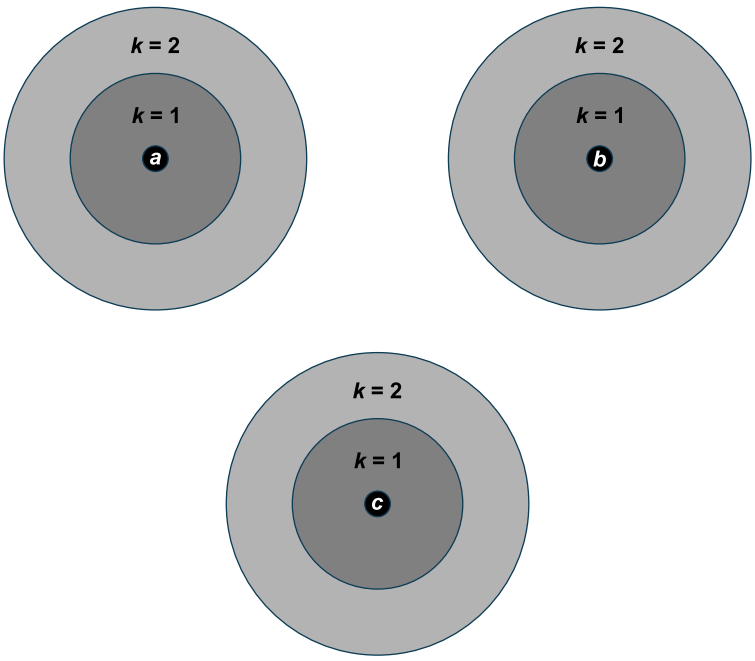

Figure 2. The simple probabilistic model showing three core species a, b, and c along with their associated clouds having $k = 1$ and $k = 2$ mutations.

It is clear that Entropy 2 will reflect only those features put explicitly into the simple model. Therefore, as an indirect check for loss of generality, we carry out the following comparison of theory (probabilistic model) with experiment (actual runs of the TNM algorithm). Based on the probabilistic

model, we calculate an expected SAD, which is compared to the empirical SAD in the Experimental Findings section below. The figure presented there compares the two distributions at two vertical scales appropriate to classes of low abundance (cloud) and high abundance (core). In addition, Figure A1 in Appendix B compares the predicted and empirical values of Entropy 1 for all seven of the times. The closeness of the match, we believe, justifies using the probabilistic model as the basis for defining Entropy 2 without loss of generality or unnecessary narrowness.

We begin by identifying the important quantities expressing the structural features above that serve to specify our simplified view of the QESS produced in a single run. Let n denote the number of core species. The space of potential species or phenotypes is represented as a binary hypercube of dimension K consisting of 2^K sites ($K = 20$ in our implementation). The n core species are characteristic of the QESS and therefore correspond to n fixed sites in the hypercube. Let N denote the total number of individuals. In reality, N fluctuates slightly during a QESS. Nevertheless, we take N as a fixed value characteristic of a run. We also assume the following:

- the separate core-clouds do not overlap;
- all core species are equally likely to reproduce.

The process must be far enough along for the first to be true. The second is not strictly true but is made for convenience. The resulting predictions are in good agreement with the observed experimental values.

Consider a core species x and one of its progeny y. If k distinct positions in the genome of y have mutated, then y will be found in the Hamming sphere $s_x(k)$ of radius k about the site x. We note that the number of such sites is

$$c(k) = \binom{K}{k}. \tag{1}$$

Since each position along the genome of a new offspring has probability p_{mut} to mutate, the fraction of any parent's progeny that lie on the sphere $s_x(k)$ is given by the binomial distribution

$$f(k) = \binom{K}{k} p_{\text{mut}}^k (1 - p_{\text{mut}})^{K-k} \tag{2}$$

($p_{\text{mut}} = 0.01$ in our implementation).

Since all spheres $s_x(k)$ are supplied with new individuals at a rate proportional to $f(k)$, but all individuals disappear at a fixed rate p_{kill}, it is clear that, in the stationary state, the individuals occupying the core-cloud for x will be distributed according to $f(k), k = 0, \ldots, K$.

This description of a QESS suggests a probabilistic experiment that captures most features that are relevant to formulating any entropy for the QESS. There are three steps to carrying out a reproduction event: choose (1) the core species that is reproducing, (2) the Hamming sphere about the core species, i.e., the value of k, the number of bits mutated, and (3) the site within the Hamming sphere. We treat all core species as having equal fecundity, so the first step has probability $1/n$. The second step has probability $f(k)$, where k is the radius of the Hamming sphere. The third step has probability $1/c(k)$, since all sites within the sphere are equally likely. Thus, p_i is given by

$$p_i = \frac{f(k)}{nc(k)}. \tag{3}$$

Equation (3) holds for any i lying within any Hamming sphere of radius k. The set of all I with this exact value of p_i is the union $U(k)$ of all Hamming spheres of radius k. For purposes of the probabilistic model described below, it is useful to note that its cardinality is $|U(k)| = nc(k)$.

We propose the following toy model for the generation of a run producing N individuals with n reproducing core species.

Probabilistic Model:

N independent trials are performed in which the outcome of a single trial (mimicking a reproduction event) is chosen from the space of potential species according to the probability p_i, in Equation (3).

3.3. Predictions from the Simple Model

Using this model, we can set up the entropies described in the last section precisely. Each is formulated as a configurational entropy in which the configurations that are counted are defined by constraining the expected values of certain families of random variables. Therefore, strictly speaking, they should be regarded as expected entropies within the framework of this model.

Entropy 1 (expected):

Define the random variables Y_a = the number of distinct species seen exactly a times among the outcomes of all N trials. The expected values $L_a = E[Y_a]$ for $a = 1, \ldots, N$ constitute what is usually called the species abundance distribution (SAD). If we define $L = \sum_a L_a$ and $\pi_a = L_a/L$, we can write a formal expression for Entropy 1:

$$S_1 = -L \sum_a \pi_a \ln \pi_a. \tag{4}$$

A derivation of L_a as a function of a and the parameters N, n of the process is given in Appendix B. This permits a calculation of an expected value of S_1 characteristic of each evolutionary time studied empirically. These values can be compared with the values for S_1 obtained directly from the observed values for Y_a. The results show a close match, as exhibited in Figure A1.

The close agreement between the toy model and the in silico experiment justifies the assumptions in the toy model as reasonable for describing the manner in which observed species are distributed over the space of potential species during a QESS. The import of this bears critically on our discussion of Entropy 2 below, as the features of the toy model supply the basis for its definition and calculation.

We remark that Stirling's approximation allows us to interpret S_1 as a configurational entropy in which the multinomial can be interpreted as a count of the number of process outcomes consistent with the expected values L_a:

$$-L \sum_a \pi_a \ln \pi_a \approx \ln\left(\frac{L!}{\prod_a L_a!}\right). \tag{5}$$

Entropy 2 (expected):

Define the random variables Z_k = the number of trial outcomes that end up in the set $U(k)$, as defined above. Since p_i is given by Equation (3) uniformly for all $i \in U(k)$, the probability of any outcome winding up in $U(k)$ is just $nc(k)p_i = f(k)$. Therefore, we have the expected values $E[Z_k] = Nf(k)$. Suppose we view a typical outcome to be one that is consistent with these expected values. Then, for each $k = 0, \ldots, K$, a typical outcome must deliver the expected $Nf(k)$ outcomes to the distinct $nc(k)$ sites of the set $U(k)$. These allocations to the sets $U(k)$ are all independent. Consequently, the total count of typical outcomes is a product and the corresponding Boltzmann entropy is given by

$$S_2 = \ln \prod_{k=0}^{K} \binom{Nf(k) + nc(k) - 1}{Nf(k)}. \tag{6}$$

A value of S_2 can be calculated for each of the 800 experimental runs and the average computed. This can be done for each of the seven evolutionary periods studied.

We remark that in Equation (6) the entropy cannot be related to Shannon information in any obvious way. Nevertheless, as pointed out in the discussion section below, the ratio S_2/S_1 seems to be relatively constant over evolutionary time. Thus, the possibility of putting these two entropies on a similar conceptual footing remains.

4. Experimental Findings

An ensemble of 800 independent copies of the TNM was run with parameter values $p_{\text{kill}} = 0.25$ and $p_{\text{mut}} = 0.01$, always starting with a single species of 500 individuals. Snapshots of the extant species were collected at times $t = 100, 500, 1000, 5000, ...100,000$.

Figure 3 illustrates the time evolution of the ensemble: a logarithmically growing average population $\langle N \rangle$ and average number of core species $\langle n \rangle$.

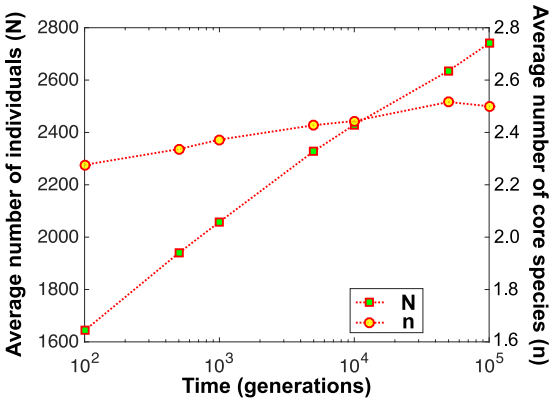

Figure 3. The average number of individuals (green squares, left axis) and average number of core species (yellow circles, right axis) plotted vs. the logarithm of time.

The species abundance distribution obtained from these data at $t = 10,000$ is shown in Figure 4 along with the predictions (blue Xs) from the simple probabilistic model. Most species are in the cloud, and most individuals are in the core. Hence, the abundance of the relatively few core species that are not visible in the vertical scale of the plot is replotted separately in the inset. SAD distributions obtained at different times all have similar shapes. These are systematically translated to the right or left, in order to reflect the larger or smaller populations of the ecosystem at different times.

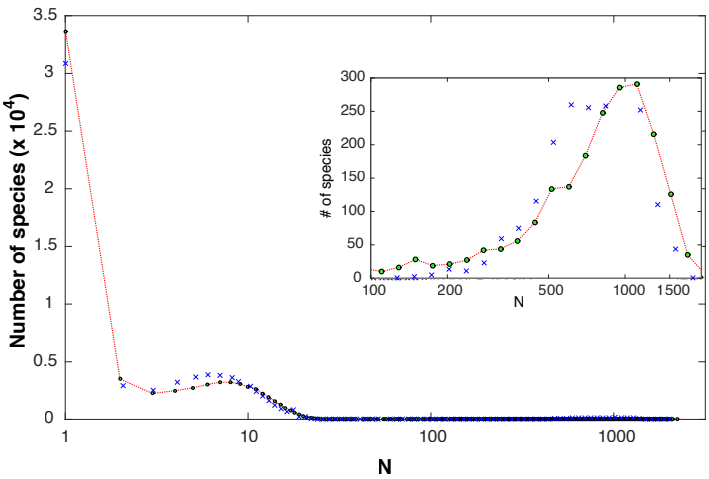

Figure 4. The empirical un-normalized species abundance distribution at $t = 10,000$, aggregated from 800 independent runs, is plotted vs. the logarithm of the abundance. The abundance of core species, which is too small to be visible in the main plot, is re-plotted separately in the vertically zoomed-in insert. Green circles represent empirical data from the Tangled Nature Model (TNM) simulation and blue Xs represent the predictions from the simple model.

The approximately lognormal form of the SAD has been used to highlight the relevance of the TNM for ecosystem dynamics [17]. We note, however, that our SAD has two approximately lognormal peaks [23] in addition to a very large maximum at $N = 1$ describing species with a single individual. These species die out quickly and get either replenished or replaced by other species. The peak located at relatively small values of the abscissa describes the many sparsely populated species of the cloud. The much lower peak centered at larger values of the abscissa describes the relatively few but highly populated core species. Interestingly, the split between cloud and core emerges naturally and in a robust fashion from the empirical shape of the SAD.

5. Discussion

Overall, we define two distinct configurational entropies (see Figure 5): one based on the SAD and one based on the specific structure displayed during a QESS. The two entropy measures are nearly proportional to each other and grow logarithmically with time in our simulations (Figure 6 left vertical axis). The number of individuals in the ecosystems also grows in a logarithmic fashion, while either entropy per individual decreases in time, reflecting the emergent macroscopic structuring (Figure 6 right vertical axis).

Entropy Type	Qualitative description	Equation	Biological interpretation
Entropy 1	Configurational count based on the Shannon entropy of the species abundance distribution (SAD)	$S_1 = -L \sum_a \pi_a ln\, \pi_a$ (Eq. 4)	Quantifies the niche partitioning of an ecosystem.
Entropy 2	Configurational count based on the Boltzmann entropy of a quasi evolutionary stable state (QESS)	$S_2 = \prod_{k=0}^{K} \binom{Nf(k) + nc(k) - 1}{Nf(k)}$ (Eq. 6)	Quantifies the number of taxonomic configurations compatible with a given niche partitioning in an ecosystem.

Figure 5. Table of qualitative, quantitative, and biological descriptions of the two entropies. Entropy 1 is a Shannon entropy derived from the species abundance distribution, which describes the available niche space in an ecosystem. Entropy 2 is a Boltzmann entropy, which quantifies the number of taxonomic configurations compatible with a given number of functional niches. List of variables in Equations (4) and (6): S_1 = Entropy 1, S_2 = Entropy 2, L = number of species, π_a = fraction of species with abundance a, K = length of genome, N = total number of individuals, n = number of core species, k = number of distinct mutated positions in the genome, $f(k)$ = the fraction of progeny with k mutations, $c(k)$ = total number of k mutated sites around a core species.

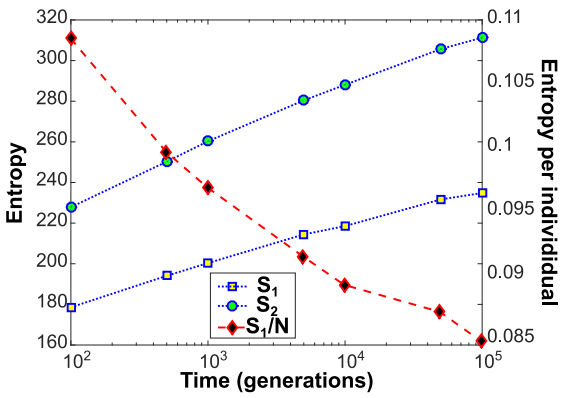

Figure 6. Left vertical axis: the squares show Entropy 1 (S_1) calculated from the empirical species abundance distribution shown in Figure 4. The circles show Entropy 2 (S_2) from the analytical configuration count (see Equation (6)) using empirical N and n values. In the range of observation, $S_2 \approx 1.27 S_1$. Right vertical axis: the diamonds show Entropy 1 per individual in the community. Since the two entropies are approximately proportional, Entropy 2 per individual looks the same and was omitted for readability. All quantities are plotted vs. the logarithm of time.

The configurational entropy of our system depends upon the level of description utilized. In our approach above, this was handled by the random variables used to describe a configuration. The entropy obtained from the SAD (i.e., Entropy 1, the lower of the two) describes the number of ways of distributing species into abundance classes. Thus, Entropy 1 can be thought of as describing the niche space available in an ecosystem, which measures the functionality of the ecosystem. That is, how many different functional niches are there in an ecosystem, and how populated are they?

Entropy 2 on the other hand is based on the structure of a QESS and describes the number of configurations that can be obtained by placing individuals into well defined classes of species. While in the TNM, these classes are the k-bit mutations from a core species, and we can envision these classes consisting of species that perform a specific function within the ecosystem. For our count, it just needs to be a well-defined class of species with a specified fraction $f(k)$, where k now indexes the functionality features that need to be present in the ecosystem. If $U(k)$ represents those species able to perform function k, we again end up with a similar count

$$S_2 = \ln \prod_{k=0}^{K} \binom{Nf(k) + |U(k)| - 1}{Nf(k)}. \tag{7}$$

Thus, Entropy 2 appears to describe the number of taxonomic configurations compatible with a given functional niche partitioning of an ecosystem. This is of particular biological relevance as this is, to our knowledge, the first time that an entropy measure has been proposed to describe the niche partitioning of an ecology. Furthermore, our findings highlight the fact that multiple taxonomic configurations can provide the needed functions. Our simulation data show an approximate proportionality between the log of the number of functional niches in an ecosystem and the log of the number of ways to fill those niches with a given number of species, $S_1 \propto S_2$. We propose that, in general, there exists a relationship between these two entropies that is implicit from the dynamic structuring in any ecosystem. Such a relationship would represent a sort of "equation of state" for the ecosystem, relating the number of functional niches to the number of ways to fill these niches.

These notions of functional and taxonomic entropy may be of particular relevance to the understanding of community structure in ecosystems called holobionts. A holobiont is a host organism along with all of its associated microbiota (i.e., viruses, bacteria, protists, and fungi). Much like the TNM, holobionts have core members of the community, which are stable long-term associations and cloud members, which are transient taxa that only serve to provide particular functionalities [27]. Although in real ecosystems these different transient members are most likely replaced by migration rather than by mutation as in the TNM, we believe that the results shown herein are general and do not depend on the underlying microevolutionary mechanism. Our notions of functional and taxonomic entropies may provide a way to describe the relationship between core and cloud members of holobiont communities, as well as a way to describe large scale configurational patterns of macroscopic ecological systems.

6. Conclusions

In this paper, TNM simulation data are uilized to calculate and describe two configurational entropies that behave similarly over time, but capture two different notions of structure in the same system. The TNM and other ecological models offer a convenient way to explore these entropies. However, the real value of these entropies is that they can serve to calculate the numbers of functional and taxonomic configurations for real ecosystems at scales ranging from holobionts to large-scale ecosystem patterning.

Acknowledgments: The authors would like to thank Ben Felts and Bjarne Adresen for their thoughtful discussion and insight during the writing process. We would also like to thank the National Science Foundation for funding provided to T.N.F.R. (NSF Division of Graduate Education G00009988).

Author Contributions: T.N.F.R., P.Sa., P.Si. and J.N. conceived and designed the experiments; P.Si. performed the experiments; T.N.F.R., P.Sa., P.Si., J.N., and F.R. analyzed the data and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. TNM Implementation

In this appendix, our TNM implementation is reviewed for convenience. The couplings J_{ab} between two genomes a and b could in principle be stored in a huge sparse matrix. This is computationally inefficient, since species with zero population, by far the most common at any given stage, do not influence the dynamics. The standard procedure recomputes the interactions each time they are needed, as described below.

The first step is to create three random arrays of length 2^K, called I, 1 and F_2. The first array controls whether two species have non-zero interactions or not, and contains ones with probability $p_{\text{interaction}} = 1/4$ and zeros otherwise. The other two arrays fix the value of the interactions $J_{ab} \neq 0$ and contain random numbers independently drawn from a Gaussian distribution with zero mean and unit variance. These arrays never change dynamically during a simulation. To determine the couplings linking genomes a and b, the binary string $c = XOR(a, b)$ is calculated first. Now reading binary strings as integer labels whenever needed, the coupling J_{ab} is set to zero if and only if $I(c) = 0$. Otherwise, $J_{ab} = GF_1(c)F_2(b)$ and $J_{ba} = GF_1(c)F_2(a)$, where G is a constant. Note that if $J_{ab} \neq 0$ then $J_{ba} \neq 0$ and $J_{ba} \neq J_{ab}$.

Reproduction probabilities are calculated as follows: at time t, let $N_b(t)$ denote the population size of species b, and $N(t) = \sum_b N_b(t)$ denote the total population size. An individual of type a is chosen as a candidate for reproduction with probability N_a/N and successfully reproduces with probability $p_{\text{off}}(a) = 1/(1 + e^{-H_a})$, where

$$H_a(t) = -\mu N(t) + \frac{C}{N} \sum_b J_{ab}(t) N_b, \tag{A1}$$

and where $C = 100$. In Equation (A1), μ is a positive constant which limits the size of the ecosystem by introducing an indirect negative coupling between all extant individuals. At each successful reproduction step, a point mutation in the offspring occurs with probability of mutation per bit by p_{mut}. Parent and offspring then differ by k bits with probability $\text{Bin}(k; K, p_{\text{mut}})$, the binomial distribution. Finally, an individual selected for reproduction, succesful or not, is killed with probability p_{kill}.

Appendix B. Expected Species Abundance

We calculate here $L_a = E[Y_a]$, as defined for entropy S_1 above.

Let Y_{ia} denote the random variable characteristic of the event "The site i is chosen a times". Then, $E[Y_{ia}]$ is just the probability of the event, as given by the binomial expession

$$E[Y_{ia}] = \binom{N}{a} p_i^a (1 - p_i)^{N-a}. \tag{A2}$$

Since $Y_a = \sum_i Y_{ia}$, we can sum Equation (A2) over i to obtain

$$L_a = E[Y_a] = \binom{N}{a} \sum_i p_i^a (1 - p_i)^{N-a}. \tag{A3}$$

Replacing p_i with Equation (3) and summing over k, instead of i, we have an explicit formula for the expectations L_a in terms of a, N, and n:

$$L_a = \binom{N}{a} \sum_{k=0}^{K} nc(k) (\frac{f(k)}{nc(k)})^a (1 - \frac{f(k)}{nc(k)})^{N-a}. \tag{A4}$$

This formula enables a comparison of the predictions of the probabilistic model with experimental findings, illustrated in Figure A1. The red circles are produced as follows. The distribution $\{L_a\}$ is computed for the pair of values N and n for each empirical run. These distributions are then averaged over the 800 runs for each evolutionary time to obtain a distribution characteristic of that time, say $\{L_a^*\}$. Equation (4) then produces an expected value for Entropy 1 at each time. The squares in Figure A1 repeat the empirical values for Entropy 1 from Figure 4 for comparison.

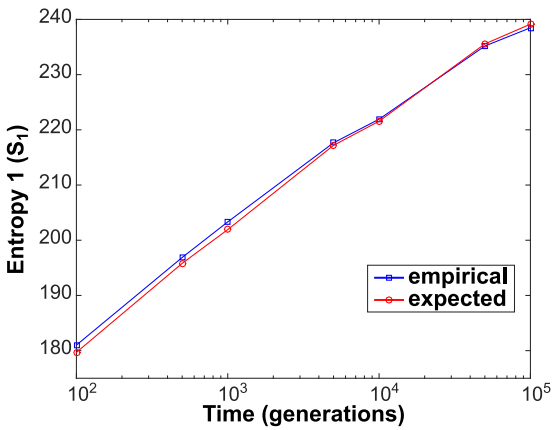

Figure A1. A comparison of observed with expected Entropy 1. The squares repeat the values shown in Figure 6. The calculation of the circles is based on the simple model of a quasi evolutionary stable states (QESS).

References

1. Cushman, S.A. Calculating the configurational entropy of a landscape mosaic. *Landsc. Ecol.* **2016**, *31*, 481–489.
2. Vranken, I.; Baudry, J.; Aubinet, M.; Visser, M.; Bogaert, J. A review on the use of entropy in landscape ecology: Heterogeneity, unpredictability, scale dependence and their links with thermodynamics. *Landsc. Ecol.* **2015**, *30*, 51–65.
3. Harte, J. *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*; Oxford University Press: Oxford, UK, 2011.
4. Odum, H.T.; Pinkerton, R.C. Time's speed regulator: The optimum efficiency for maximum power output in physical and biological systems. *Am. Sci.* **1955**, *43*, 331–343.
5. Ulanowicz, R.E. *Ecology, the Ascendent Perspective: Robert E. Ulanowicz*; Columbia University Press: New York, NY, USA, 1997.
6. Ulanowicz, R.E. *Growth and Development: Ecosystems Phenomenology*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
7. Brooks, D.R.; Wiley, E.O. *Evolution as Entropy*; University of Chicago Press: Chicago, IL, USA, 1988.
8. Wicken, J.S. A thermodynamic theory of evolution. *J. Theor. Biol.* **1980**, *87*, 9–23.
9. Wicken, J.S. *Evolution, Thermodynamics, and Information: Extending the Darwinian Program*; Oxford University Press: Oxford, UK, 1987.
10. Depew, D.J.; Weber, B.H. *Entropy, Information, and Evolution: New Perspectives on Physical and Biological Evolution*; MIT Press: Cambridge, MA, USA, 1988.
11. Landsberg, P.T. Can entropy and "order" increase together? *Phys. Lett. A* **1984**, *102*, 171–173.
12. Kauffman, S.A.; Levine, S. Towards a general theory of adaptive walks in rugged landscapes. *J. Theor. Biol.* **1987**, *128*, 11–45.
13. Kauffman, S.A.; Johnsen, B. Coevolution to the Edge of Chaos: Coupled Fitness Landscapes, Poised States, and Coevolutionary Avalanches. *J. Theor. Biol.* **1991**, *149*, 467–505.
14. Eldredge, N.; Gould, S.J. Punctuated equilibria: An alternative to phyletic gradualism. In *Models in Paleobiology*; Freeman, Cooper and Company: San Francisco, CA, USA, 1972; pp. 82–115.

15. Christensen, K.; de Collobiano, S.A.; Hall, M.; Jensen, H.J. Tangled nature: A model of evolutionary ecology. *J. Theor. Biol.* **2002**, *216*, 73–84.
16. Hall, M.; Christensen, K.; di Collobiano, S.A.; Jensen, H.J. Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E* **2002**, *66*, 011904.
17. Anderson, P.E.; Jensen, H.J. Network properties, species abundance and evolution in a model of evolutionary ecology. *J. Theor. Biol.* **2005**, *232*, 551–558.
18. Volkan, S.; Rikvold, P.A. Effects of correlated interactions in a biological coevolution model with individual-based dynamics. *J. Phys. A* **2005**, *38*, 9475–9489.
19. Lawson, D.; Jensen, H.J. The species-area relationship and evolution. *J. Theor. Biol.* **2006**, *241*, 590–600.
20. Laird, S.; Jensen, H.J. The Tangled Nature model with inheritance and constraint: Evolutionary ecology restricted by a conserved resource. *Ecol. Complex.* **2006**, *3*, 253–262.
21. Yoshuke, M.; Shimada, T.; Ito, N.; Rikvold, P.A. Random walk in genome space: A key ingredient of intermittent dynamics of community assembly on evolutionary time scales. *J. Theor. Biol.* **2010**, *264*, 663–672.
22. Becker, N.; Sibani, P. Evolution and non-equilibrium physics: A study of the Tangled Nature Model. *EPL* **2014**, *105*, 18005.
23. Andersen, C.W.; Sibani, P. Tangled nature model of evolutionary dynamics reconsidered: Structural and dynamical effects of trait inheritance. *Phys. Rev. E* **2016**, *93*, 052410.
24. Anderson, P.E.; Jensen, H.J.; Oliveira, L.P.; Sibani, P. Evolution in complex systems. *Complexity* **2004**, *10*, 49–56.
25. Gould, S.J.; Eldredge, N. Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* **1977**, *3*, 115–151.
26. Baldridge, E.; Harris, D.J.; Xiao, X.; White, E.P. An extensive comparison of species-abundance distribution models. *PeerJ* **2016**, *4*, e2823.
27. Hester, E.R.; Barott, K.L.; Nulton, J.; Vermeij, M.J.; Rohwer, F.L. Stable and sporadic symbiotic communities of coral and algal holobionts. *ISME J.* **2015**, *10*, 1157–1169.

© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).