

# Network properties, species abundance and evolution in a model of evolutionary ecology

Paul E. Anderson, Henrik Jeldtoft Jensen\*

*Department of Mathematics, Imperial College, 180 Queen's Gate, London, SW7 2AZ, UK*

Received 29 October 2003; received in revised form 20 February 2004; accepted 18 March 2004

Available online 19 October 2004

## Abstract

We study the evolution of the network properties of a populated network embedded in a genotype space characterized by either a low or a high number of potential links, with particular emphasis on the connectivity and clustering. Evolution produces two distinct types of network. When a specific genotype is only able to influence a few other genotypes, the ecosystem consists of separate non-interacting clusters (i.e. isolated compartments) in genotype space. When different types may influence a large number of other sites, the network becomes one large interconnected cluster. The distribution of interaction strengths—but not the number of connections—changes significantly with time. We find that the species abundance is only realistic for a high level of species connectivity. This suggests that real ecosystems form one interconnected whole in which selection leads to stronger interactions between the different types. Analogies with niche and neutral theory and assembly models are also considered.

© 2004 Elsevier Ltd. All rights reserved.

*Keywords:* Ecosystems; Networks; Species abundance distribution; Evolution and self-organization; Neutral and niche theory

## 1. Introduction

An important characteristic of an ecosystem is the total set of interactions between the various individuals. Organisms may influence each other in many ways and it is difficult to monitor and quantify all possible interactions except the most direct, such as simple trophic relationships. Despite this, some recent work has made good progress towards measuring interaction strengths in real ecosystems (Ives et al., 2003; Krause et al., 2003). The development of the set of interactions over evolutionary time scales is even more difficult to measure because of random mutations and the resulting adaptations. Gaining an understanding from observations is also problematic since *laws* may very well only be recognizable at the level of *averages*, see (Loreau and

Hector, 2001; Yedid and Bell, 2002). Here, we approach these issues within the framework of a simple individual-based model of ecosystem assembly and evolution (Christensen et al., 2002; Hall et al., 2002; diCollobiano et al., 2003).

We compare the early and late time connectivity and cluster properties of ecosystems evolving in two differently connected spaces such that genotypes interact with *either* a small or a large number of other genotypes. Note that the connectivity simply refers to whether any two organisms can influence each other or not. Clusters of genotypes are defined as being completely isolated from each other. This is similar to the ecological concept of compartmentalization usually used in the context of food webs (Begon et al., 1996; Pimm and Lawton, 1980; Krause et al., 2003). However, whilst interactions within compartments are strong, the interactions between compartments are weak but not necessarily non-existent. In this sense, clusters can be seen as isolated compartments.

Clearly, the number of interactions experienced by a site depends on which of all the possible mutations and

\*Corresponding author. Tel.: +44 20 7594 8541; fax: +44 20 7594 8517.

*E-mail addresses:* [h.jensen@ic.ac.uk](mailto:h.jensen@ic.ac.uk), [h.jensen@imperial.ac.uk](mailto:h.jensen@imperial.ac.uk) (H.J. Jensen).

*URL:* <http://www.ma.imperial.ac.uk/~hjens>.

adaptations have actually occurred, i.e. the network is dependent on its history. It turns out that the strengths of the interactions change significantly with time, whilst the degree distribution (number of active interactions) remains close to what would be expected if genotypes were occupied at random. The species abundance curve takes a log-normal form only for spaces where the genotypes are linked to many others, suggesting that a high connectedness leads to a more realistic ecology. As discussed later, our model is neutral in some cases, but is unrelated to niche theory in its current form.

How organisms interact and the consequences of connectivity and complexity in real ecologies has long been an important question for mathematical modelling (May, 1974). It is still a very active area of research: Haydon recently showed that the theoretically ‘stablest possible systems are characterized by heightened connectance’ (Haydon, 2000), provided that the species with strong self-interactions are connected to those with weak self-interactions.

In this paper, we are not directly concerned with the question of the stability of an ecosystem, but more in the assembly and structure of a realistic set of interactions. The idea of stability and resistance to invasion by mutants is implicit in this discussion, but the ultimate goal is to examine the sort of structures that appear in the network over evolutionary timescales. This work is related conceptually to models of community assembly (Drake, 1990; Law and Morton, 1996; Morton et al., 1996). The analogies are drawn in the next section when the nature of mutation and the interactions have been introduced. The main aim of this paper is to study directly the effect of altering the level of potential connectivity on an ecosystem model and its subsequent evolution.

## 2. Methods

### 2.1. Definition of the model

Here, we briefly describe the structure and dynamics of the Tangled Nature model (Christensen et al., 2002; Hall et al., 2002). An individual is represented by a vector  $\mathbf{S}^\alpha = (S_1^\alpha, S_2^\alpha, \dots, S_L^\alpha)$  in the genotype space  $\mathcal{S}$ , where the ‘‘genes’’  $S_i^\alpha$  may take the values  $\pm 1$ , i.e.  $\mathbf{S}^\alpha$  denotes a corner of the  $L$ -dimensional hypercube. In the present paper we take  $L = 20$  as this gives a generously sized space to explore (over a million genotypes) whilst not being computationally prohibitive. We think of the genotype space  $\mathcal{S}$  as containing all possible ways of combining the genes into genome sequences. Many sequences may not correspond to viable organisms. The viability of a genotype is determined by the evolutionary dynamics. All possible sequences are made available for evolution to select from. The number of occupied sites is

referred to as the diversity, here analogous to the number of species or species richness (Krebs, 1999). As explained later, genotype, species, site and node are synonymous throughout.

For simplicity, an individual is removed from the system with a constant probability  $p_{kill}$  per time step. A time step consists of one annihilation attempt followed by one reproduction attempt. One generation consists of  $N(t)/p_{kill}$  time steps, which is the average time taken to kill all currently living individuals. All references to time will be in units of generational time.

The ability of an individual to reproduce is controlled by a weight function  $H(\mathbf{S}^\alpha, t)$ :

$$H(\mathbf{S}^\alpha, t) = \frac{1}{cN(t)} \left( \sum_{\mathbf{S} \in \mathcal{S}} J(\mathbf{S}^\alpha, \mathbf{S})n(\mathbf{S}, t) \right) - \mu N(t), \quad (1)$$

where  $c$  is a control parameter,  $N(t)$  is the total number of individuals at time  $t$ , the sum is over the  $2^L$  locations in  $\mathcal{S}$ , and  $n(\mathbf{S}, t)$  is the number of individuals (or occupancy) at position  $\mathbf{S}$ . Two positions  $\mathbf{S}^\alpha$  and  $\mathbf{S}^\beta$  in genotype space are coupled with the fixed random strength  $J^{\alpha\beta} = J(\mathbf{S}^\alpha, \mathbf{S}^\beta)$  which can be either positive, negative or zero. This link is non-zero with probability  $\theta$ , i.e.  $\theta$  is simply the probability that any two sites are interacting. In order to study the effects of interactions between species, there is no self-interaction, so that  $J^{\alpha\alpha} = 0$ . The present paper compares the three cases  $\theta = \frac{1}{1000}$ ,  $\theta = \frac{1}{200}$  and  $\theta = \frac{1}{4}$  (see Fig. 1). The non-zero values of  $J^{\alpha\beta} \neq J^{\beta\alpha}$  are determined by a deterministic but rapidly varying function of the two positions  $\mathbf{S}^\alpha$  and  $\mathbf{S}^\beta$ .

The method for computing  $J^{\alpha\beta}$  at  $t = 0$  is similar to multiplying two random numbers, each drawn from a normal distribution with mean zero. This would give a modified Bessel function of the second kind. The actual shape of the  $J$  distribution can be seen in the inset of Fig. 3 and it is indeed very close, but not identical, to a

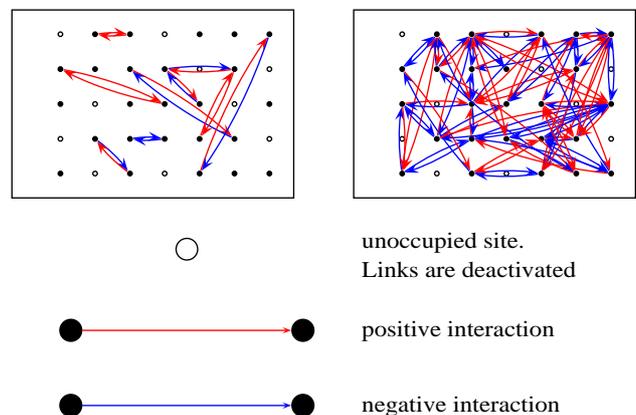


Fig. 1. Two types of network. Left: low  $\theta$ . Most sites are in two-clusters. There is no tendency to form larger clusters at later times. Right: high  $\theta$ . All species are connected in one giant cluster. In both cases, not all sites are occupied.

Bessel function. The exact procedure used (Hall et al., 2002) was slightly different and chosen for computational convenience.

The conditions of the physical environment are simplistically described by the term  $\mu N(t)$  in Eq. (1), where  $\mu$  determines the average sustainable total population size. An increase in  $\mu$  corresponds to harsher physical conditions. Note that we cannot identify  $\mu$  with the carrying capacity, since the maximum population size that can be supported for a given  $\mu$  is not a fixed limit and can indeed be increased by different configurations of the  $J$  couplings. Thus,  $\mu$  can be tuned to obtain a particular average population size but cannot set a fixed cap on  $N(t)$ . Genotypes adapt to each other and the physical environment represented by  $\mu$ .

We use asexual reproduction consisting of one individual being replaced by two copies mimicking the process of binary fission seen in bacteria. Successful reproduction occurs with a probability per unit time given by

$$p_{off}(\mathbf{S}^z, t) = \frac{\exp[H(\mathbf{S}^z, t)]}{1 + \exp[H(\mathbf{S}^z, t)]} \in [0, 1]. \quad (2)$$

This function is chosen for convenience. We simply need a smoothly varying function that maps  $H(\mathbf{S}^z, t)$  to the interval  $[0, 1]$  and it is otherwise arbitrary. We allow for mutations in the following way: with probability  $p_{mut}$  per gene we perform a change of sign  $S_i^z \rightarrow -S_i^z$  during reproduction.

Since the interaction between a new species, its parent and the other extant species is random, Tangled Nature is very similar to assembly models where an immigrant species is drawn at random from a pool of species (Drake, 1990; Law and Morton, 1996; Morton et al., 1996). In both cases, a species needs to be able to succeed when rare in order to become established within the community. However, in Tangled Nature a new mutant can be produced, potentially, in every time step. This means that although the species are introduced one at a time, there is no guarantee that the system will settle into an equilibrium before the next new species is inserted. This is in contrast to assembly models where the resident community has a chance to respond to the newcomer and in some sense settle down before another immigrant is introduced (Drake, 1990; Law and Morton, 1996). Morton et al. highlight this neglect of ‘simultaneous invasions on the transient trajectories of resident communities’ (Morton et al., 1996).

Furthermore, our pool has no explicit trophic structure unlike the model in Drake (1990) where the source pool is divided into omnivores, predators, herbivores and producers. Despite this, it is very useful to consider new species in Tangled Nature as randomly drawn from a pool. After all, the one-step mutant genotypes of highly populated sites are likely to be born (and possibly go extinct) regularly. This is analogous to

members of the pool having several chances to invade (Drake, 1990), and emphasizes that the order in which species are drawn from the pool matters. The Tangled Nature model relates this idea of community assembly to evolution.

Initially, we place  $N(0) = 500$  individuals at randomly chosen positions. Their initial location in genotype space does not affect the dynamics. A two-phase switching dynamic is seen consisting of long periods of relatively stable configurations (quasi-Evolutionary Stable Strategies or q-ESSs) interrupted by brief spells of reorganization of occupancy which are terminated when a new q-ESS is found, as discussed in Christensen et al. (2002).

## 2.2. Parameters

As mentioned previously, the only parameter that is changed in this paper is the connectivity,  $\theta$ . The values used throughout are  $L = 20$ ,  $c = 0.005$ ,  $\mu = 0.01$ ,  $p_{kill} = 0.2$  and  $p_{mut} = 0.015$ . This selection provides several transitions between different q-ESSs for each run, see (Hall et al., 2002; diCollobiano et al., 2003). We consider three  $\theta$  values: 0.001, 0.005 and 0.25, which we will refer to as very low, low and high  $\theta$ , respectively. These correspond to below, near and above the percolation threshold. That is, the point where there is a *non-zero* probability that all living sites are connected in one dominant cluster (see Albert and Barabási (2002) for a review of network models and terminology). A realistic species abundance curve was only obtained above the threshold. We will be contrasting results at  $t = 500$  (primal time),  $t = 5000$  (early time) and  $t = 500,000$  (late time). Early time is well outside the system’s initial transient search for a quasi-stable configuration in genotype space. The low and high  $\theta$  ensembles consist of 500 realizations. Each run uses a different random number seed but, for any given run, only  $\theta$  is changed between the two ensembles. Such large ensembles were obtained using a Super Computer At Night (SCAN) cluster of undergraduate machines running in parallel (with the FreeBSD operating system) overnight and during weekends and holidays.

Figs. 2–4 show two sets of data: one labelled *simulation* (which are the results generated by the dynamics of the model) and the other *random*. In the random case, rather than evolving the network, for any specified time we read in the diversity and number of individuals alive from the simulated run. The individuals are then thrown on to the network of  $2^{20}$  genotypes at random with the constraint that the diversity is the same as the simulation. That is, the species are simply selected at random and then the various properties are measured. There is no dynamics. Thus random data is *not* dependent on the history of the network, but has the same global properties (diversity and population size) as the simulations. This provides a very useful null model.

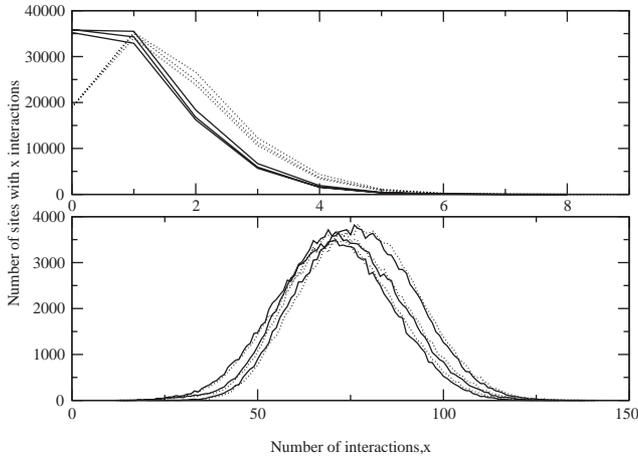


Fig. 2. Degree histograms. Top: Degree histogram for  $\theta = 0.005$ . Bottom:  $\theta = 0.25$ . Solid lines, random; dotted lines, simulation. From the left, the pairs of curves are for  $t = 500, 5000$  and  $500,000$ . At later times, the number of active links increases for both the simulation and random data.

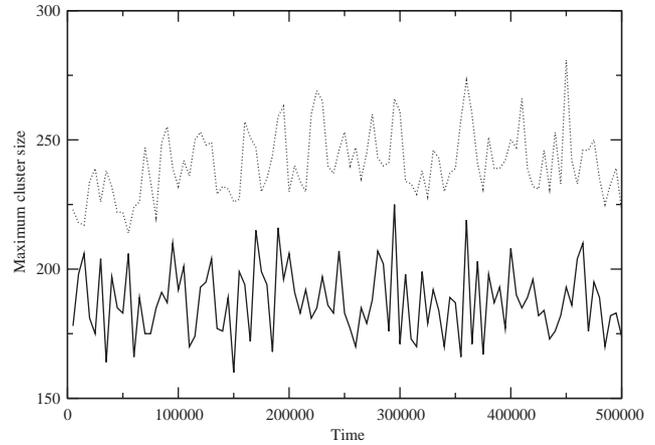


Fig. 4. Maximum cluster size. Maximum cluster size across all realisations for  $\theta = 0.005$ . Solid line, random; dotted line, simulation. Clusters produced by the simulation are larger than those produced in a history-independent network.

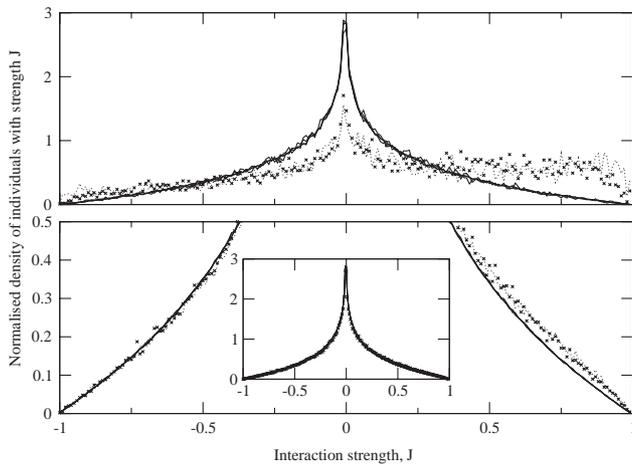


Fig. 3. Interaction distributions. Top: Distribution of interaction strengths between individuals for  $\theta = 0.005$ . Bottom:  $\theta = 0.25$ . Inset: Entire distribution. Solid lines, random; crosses, simulation at  $t = 500$ ; dotted lines, simulation at  $t = 500,000$ . All plots are normalized so that their area is one. For high  $\theta$ , a significant increase in positive interactions is seen. For low  $\theta$ , a change is seen but for trivial reasons.

Comparisons with this will reveal whether the network is really evolving, or the results are just by-products of increasing diversity. Simulated data is always shown as a dotted line and random data as a continuous line.

### 3. Results

#### 3.1. Connectivity

We study the temporal evolution of the network connectivity in the space of occupied positions for different  $\theta$  values. Note that the hard-wired configura-

tion of couplings  $J(\mathbf{S}^a, \mathbf{S}^b)$  between all  $2^{20}$  positions in genotype space is determined at  $t = 0$  and remains constant. The network of *occupied* sites will nevertheless change with time and so the network properties at any given time depend on which genotypes are inhabited. Interactions between other sites can be explored by mutations away from the occupied sites. The degree distributions in Fig. 2 show the number of genotypes having  $x$  active interactions.

The leftmost pair of curves represents primal time, the next, early time and the rightmost late time. Considering only the simulation data for now, a clear shift to a greater number of active links is seen in the high  $\theta$  case, whilst a slight change occurs for low  $\theta$ . The difference between early and late time is bigger than that between early and primal time. The degree of a site is equal to the number of direct interactions it has with all other occupied sites. This explains why any particular site in the low  $\theta$  runs only has at most nine and usually only one or two direct interactions. The data is summed up over the entire ensemble.

How much of this shift is due to a genuine change in network connectivity? For high  $\theta$ , the null model data shows that there is very little difference between evolving the network and throwing individuals down randomly. Low  $\theta$  appears to show a change. However, any site that does not interact with any others will die very quickly in a simulation. If for any instant in time genotype positions are chosen by chance, such a low connectivity will give a disproportionate number of isolated genomes that would be forbidden by the dynamics. There is no fair way to simulate this effect, but it can be seen that the differences between the time curves in the random and simulated runs is similar and thus the network connectivity does not evolve for either value of  $\theta$ .

### 3.2. Interaction strength

For both  $\theta$  values, the diversity gradually increases with time. What is causing this? It turns out that the strength of the interaction between sites is crucial to the ability of the network to support larger numbers of individuals. Fig. 3 shows the distribution of interaction strengths between each *living site* and all other living sites at a given time. Interaction strengths are assigned at random and are not necessarily symmetric (Hall et al., 2002). For example,  $J(S^{18}, S^{59}) = 0.3$ , but  $J(S^{59}, S^{18}) = -0.2$ . For all times and both values of  $\theta$ , the distribution for the random data approximately follows a modified Bessel function of the second kind as described earlier. It is a symmetrical curve peaking at  $J = 0$ . This makes sense because there is no bias in the ratio of positive to negative links when the links are assigned to the “bare” network at  $t = 0$ .

For reasons of clarity, the simulation results are only shown for primal and late time. Clearly, a significant change takes place for high  $\theta$  between  $t = 0$  and primal time. Some weight is taken in the fall of the peak at  $J = 0$  and the drop in negative  $J$  values, and redistributed into positive strengths, that is the curve moves right. This comprises a significant shift in the probability density. The change from primal to late time is smaller—but still noticeable—since the large number of reorganizations of genotype space in the early generations drops to occasional punctuation of q-ESSs later in the run. (Typically, there are only one or two transitions from early time onwards.) Despite this, the curve continues to drift to the right. On first inspection, the low  $\theta$  runs seem to have changed dramatically from the initial configuration. However, nothing particularly interesting is happening here: it is simply an effect of the structure of clusters in the low  $\theta$  space and is explained below.

### 3.3. Clustering

The clustering (how sites are linked to each other through other sites) is another useful network measure. For high  $\theta$  we find that at any given time, all occupied sites in genotype space belong to one and the same cluster. Thus the cluster size for high  $\theta$  simply follows the diversity. In contrast, we observe the formation of distinct clusters (or compartments) for low  $\theta$ . The rest of this section will deal solely with low  $\theta$ .

The overall structure of the clusters does not seem to change much with time. This is the first indication that the clustering is not an evolving property of the network. As would be expected, one-clusters are transient. They are born on a new site as mutants from a parent but are isolated from other sites and so are extremely unlikely to reproduce. (Since  $\mu = 0.01$  and the average total population is about 2700,  $p_{off} \approx$

$10^{-12} \ll p_{kill} = 0.2$  so an isolated site is much more likely to be killed than multiply when chosen.) These sites are simply flashing in and out of existence.

Simulations indicate that the building blocks of larger groups are two-clusters. These tend to be two very old sites that have mutually positive links. Large clusters are formed mainly from very old two-clusters joined together by a mutant. The continual background of mutants flitting in and out of the network plugs these building blocks together. However, the entire cluster is rarely long lived whereas the two-clusters are formed early in any particular run, quickly build up their population and are very persistent since their occupation is high.

Clusters do indeed generally increase in size with time. There are, however, large fluctuations in the record size which gives an indication of how unstable these large clusters are, see Fig. 4. The largest recorded cluster in any run at any time contained 281 sites. It is revealing to compare the results obtained from the null model, where the maximum cluster size is much smaller but just as variable. When individuals are thrown down at random, two-clusters are no longer the building blocks of the large clusters and any long string of connected sites is determined purely by chance. Hence, the biggest cluster will always be smaller than that produced by the dynamics. The temporary nature of the large clusters is further borne out by time and ensemble averaging the cluster sizes. The number of clusters of a particular size  $S$  is stored at intervals of 5000 generations from early to late time for each of the 500 runs and the time and ensemble average is then calculated. As expected, there are many one-clusters and fewer large clusters.

The distribution follows the functional form  $n_s \sim s^{-5/2} e^{-(p-p_c)^2 s}$  anticipated for the cluster size distribution on a random graph of  $D$  ( $\approx 195$ ) nodes. (See Eq. (36) in Albert and Barabási (2002).) The percolation threshold  $\frac{1}{D}$  is very close to the considered connectivity  $\theta = 0.005$ . A comparison with the random data shows that this scale-independent distribution is not due to the dynamics of the system, the only difference being the appearance of larger record clusters in the simulation, as shown in Fig. 4. This is perhaps the most compelling piece of evidence that the low  $\theta$  regime does not show any emergent structure. We also ran simulations for very low  $\theta$  (0.001) and found that the cluster sizes were exponentially distributed as would be expected below the percolation threshold.

Fig. 3 (low  $\theta$ ) can now be easily explained. Unconnected sites die extremely quickly in the first few generations leaving behind two-clusters and other sites with positive interactions. The slight increase in sites with  $J > 0$  for late time is caused by well established positive–positive two-clusters. So what looked like an interesting result initially proves to be due to the fairly constant microscopic structure of the network.

Until recently (Krause et al., 2003), there has been little evidence of this clustering or compartmentalization in nature (Pimm and Lawton, 1980). Krause et al. indicate that compartments may have been overlooked in several well-known food webs, and may well play an important role. In the Tangled Nature model, whilst clustering is inevitable at  $t = 0$  when individuals are randomly thrown down due to the low underlying connectivity, it could have been the case that the dynamics found a large, interconnected cluster that existed in the bare network. Clearly this does not happen and compartments remain. This further adds to the evidence that the low  $\theta$  case is ecologically unrealistic. Even the Krause et al. study finds only two compartments in a 45 taxa ecosystem, as opposed to the mass of isolated, tiny compartments seen in Tangled Nature.

### 3.4. Species abundance

The Species Abundance Distribution or SAD is important in characterizing ecosystems. It is the proportion of species that contain  $\rho$  individuals. We define a species as one site in genotype space. Ideally, we would like to use a coarse-grained definition more likely to reflect real ecologies, where species are defined as groups of points in genotype space echoing the genotypic cluster species definition introduced by Mallet (1995). Since the maximum number of genotypes in our model is only around  $10^6$  anyway, the single site species approach is more appropriate. This is perhaps not so unreasonable since any two gene sequences differ by at least 5%, since  $L = 20$ . We have been able to extend the initial results obtained in Hall et al. (2002) and can consider the evolution of the SAD for high and low  $\theta$  integrated across all 500 runs, as seen in Fig. 5. The

larger ensembles allow enough statistics for illuminating conclusions to be drawn. Note that the null model is absent since when individuals are sprinkled randomly across the living sites, there is no tendency for accumulation on any particular site, so the individuals follow a multinomial distribution.

The key result of this paper is that only high  $\theta$  leads to a SAD similar to those observed in nature. Low  $\theta$  is skewed by its heavily populated two-clusters. The plots for high  $\theta$  show the log-normal form observed in many real ecosystems and in other ecological models (McKane et al., 2000; Hubbell, 2001). They appear to become more log-normal as time increases with the dip between four and eight individuals falling, even though the diversity is rising. Hence, the SAD is evolving. From this, it seems that the high  $\theta$  case structures itself more like a real ecosystem than low  $\theta$ , whose SAD develops a sharper peak as the two-clusters become densely occupied. The single cluster of highly interdependent genotypes produces a reasonable SAD that cannot be formed by patches of isolated clusters.

Thus the abstract parameter  $\theta$ , which cannot be measured in a simple way in real systems, is directly linked to the easily observed SAD. We recall that low values of  $\theta$  correspond to a world in which different species, or types, are able to influence only a small number of other species. High values of  $\theta$  correspond to the situation where different types may have an impact on the vitality of a large number of other species.

The initial descent in both curves from the global peak at  $\rho = 1$  is due to the large number of sites with only one occupant. In nature, sampling difficulties would mean that these sites would not be detected so this first aspect is not seen or is at least much smaller in observed SADs. (It is particularly marked for our model since we use each site as one species and do not coarse-grain.) But the second peak does correspond well to results from the field, though it should be pointed out that the proportion of all sites with more than two individuals is only about 30% in each case. However, this is sufficient to detect the evolution of the SAD. It should also be remembered that whilst stochasticity is an important effect when the number of individuals is small (as it is in any one run), we integrate across 500 runs and so the second peak is a real effect and not just due to fluctuations.

We note that recent studies of a simplified version of the Tangled Nature model described in Rikvold and Zia (2003a,b); Zia and Rikvold (2004) found no temporal evolution of the statistics of the model. The reason for this may be that they use a relatively short genome length  $L = 13$  together with a very substantial simulation time of order  $10^7$  generations. We have observed previously that the time to reach a stationary state explodes with genome length (Christensen et al., 2002).

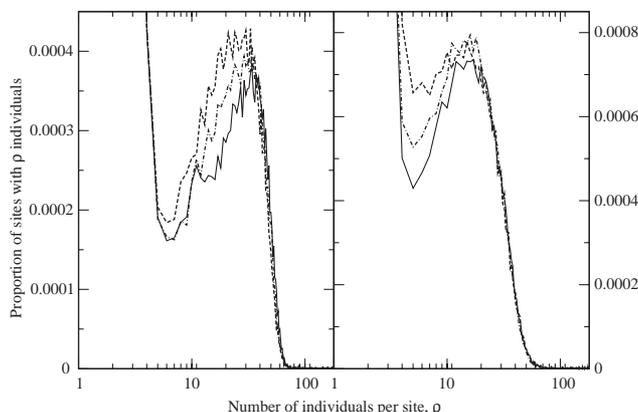


Fig. 5. Species abundance distributions. Species abundance distributions for the simulations only. Dashed line,  $t = 500$ ; dashed-dotted line,  $t = 5000$ ; solid line,  $t = 500,000$ . Low  $\theta$  on the left, high  $\theta$  on the right. The ecologically realistic log-normal form is only seen for high  $\theta$ .

#### 4. Discussion

Before discussing how these results relate to ecology in practice, we shall pause to consider the relationship between the Tangled Nature model and the neutral/niche theory debate. There has lately been much interest in the neutral theory of biodiversity (Hubbell, 2001; Bell, 2001). Despite making assumptions that are anathema to traditional niche models (all individuals are the same and adaptations to specific environmental niches are essentially unimportant) it has been successful in making predictions about real world ecology—although its effectiveness in modelling the species abundance has recently been called into question (McGill, 2003).

If all individuals are on the same genotype at  $t = 0$ , the Tangled Nature model is neutral since all individuals are identical. However, the dynamics immediately breaks this neutrality as configurations are spontaneously generated. Once individuals become differentiated, interactions matter and the model is no longer neutral—although  $\mu$  and  $p_{kill}$  remain independent of the individual and thus neutral. Technically, Tangled Nature is also neutral if  $J = 0$  for all time, but this is obviously not the case in this paper. Further, unlike the models in Hubbell (2001), we have no spatial aspect: we deal with only one large metacommunity as opposed to aggregating many local communities to form the metacommunity.

Niche models use a small number of limiting factors to model phenotypic evolution. Typical examples include food size (May, 1974), plant biomass (Jansen and Mulder, 1999) and body size (Kisdi, 1999). In Tangled Nature, no trade offs are specified so the number of limiting factors is infinite. As it stands, the model is not related to niche theory but it could be developed into a niche model by changing the nature of interactions and adding a resource that the species compete for. This line of work is currently being pursued. A spatial aspect would also bring it in to line with the existing models of migration between patches of species.

In summary, although the Tangled Nature model is not directly related to either class of model, perhaps in the future, simulation and individual-based approaches could be used to investigate the relative importance of niche and neutral effects.

Our most important results are that temporal evolution of the network properties of an ecosystem and a realistic form for the species abundance are only seen if the genotype space is well connected. This is interpreted here as meaning that an occupied genotype is likely to interact with many other (potentially occupied) genotypes. No evolution at the level of ecosystems can occur in a world where most possible genotypes are *inert*, i.e. whether they are present or not will have very little

influence on other organisms. It is easy to overlook the importance of the entire network of interactions when dealing with small communities of organisms on a macroscopic scale, but easier to visualize with colonies of billions of bacteria.

We suggest that this observation can be used to gain insight into the potential underlying connectivity between biota. Imagine two microbial evolution experiments. In one case, the microbial ecosystem evolves towards an interwoven or entangled ecology. In the other, little evolution is observed in the structure of the ecological properties of the microbial community. One might, according to the result from our model, anticipate that the first system consists of microbes from a part of genotype space in which types influence each other, whereas the second system consists of genotypes from a region of space consisting of mainly *inert* organisms.

The immediate ecological implication of this work, is that macroscopic properties that can be easily measured can give indications of the nature of microscopic properties that cannot be directly obtained. More specifically, the functional form of the SAD—a large scale, comparatively simple quantity—is important in giving an insight into the connectivity of the ecosystem—a notoriously slippery and complicated unknown. More generally, this work agrees with the idea that species are highly dependent on each other in ways that are not necessarily immediately obvious, so ecosystems should be treated as an interconnected whole rather than a set of isolated species dealt with on a case-by-case basis.

From our results, it is tempting to speculate that the observed degree of diversity, complexity and adaptation of living matter may be directly related to a high level of interdependence between organisms. Thus Darwin's entangled bank may be a useful image to keep in mind when studying the evolution of large collections of individuals.

#### Acknowledgements

We are grateful to Gunnar Pruessner for extensive help with the development of the parallelized code that enabled us to probe the model at much deeper timescales and for broader ensembles than had been possible before. We wish to thank Andy Thomas for his fantastic technical support. Without his help and dedication, this project would not have been possible. We are indebted to Dan Moore, Brendan Maguire and Phil Mayers for their continuous support. We thank Albert Diaz-Guilera for inspiring discussions and Ed Johnson for reading the manuscript. Paul Anderson thanks EPSRC for research funding.

## References

- Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.
- Begon, M., Harper, J.L., Townsend, C.R., 1996. *Ecology—Individuals, Populations and Communities*, third ed. Blackwell Science, Oxford.
- Bell, G., 2001. Neutral macroecology. *Science* 293, 2413–2418.
- Christensen, K., di Collobiano, S.A., Hall, M., Jensen, H.J., 2002. Tangled nature: a model of evolutionary ecology. *J. Theor. Biol.* 216, 73–84.
- di Collobiano, S.A., Christensen, K., Jensen, H.J., 2003. The tangled nature model as an evolving quasi-species model. *J. Phys. A: Math. Gen.* 36, 883–891.
- Drake, J.A., 1990. The Mechanics of Community Assembly and Succession. *J. Theor. Biol.* 147, 213–233.
- Hall, M., Christensen, K., diCollobiano, S.A., Jensen, H.J., 2002. Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E* 66 (011904), 1–10.
- Haydon, D.T., 2000. Maximally stable model ecosystems can be highly connected. *Ecology* 81 (9), 2631–2636.
- Hubbell, S.P., 2001. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, Princeton.
- Ives, A.R., Dennis, B., Cottingham, K.L., Carpenter, S.R., 2003. Estimating community stability and ecological interactions from time-series data. *Ecol. Monographs* 73 (2), 301–330.
- Jansen, V.A.A., Mulder, G.S.E.E., 1999. Evolving biodiversity. *Ecol. Lett.* 2, 379–386.
- Kisdi, É., 1999. Evolutionary branching under asymmetric competition. *J. Theor. Biol.* 197, 149–162.
- Krause, A.E., Frank, K.A., Mason, D.M., Ulanowicz, R.E., Taylor, W.W., 2003. Compartments revealed in food-web structure. *Nature* 426, 282–285.
- Krebs, C.J., 1999. *Ecological Methodology*, second ed. Benjamin/Cummings, Menlo Park, CA.
- Law, R., Morton, R.D., 1996. Permanence and the assembly of ecological communities. *Ecology* 77 (3), 762–775.
- Loreau, M., Hector, A., 2001. Partitioning selection and complementarity in biodiversity experiments. *Nature* 412, 72–76.
- Mallet, J., 1995. A species definition for the modern synthesis. *Trends Ecol. Evol.* 10 (7), 294–299.
- May, R.M., 1974. *Stability and Complexity in Model Ecosystems*, second ed. Princeton University Press, Princeton, NJ.
- McGill, B.J., 2003. A test of the unified theory of biodiversity. *Nature* 422, 881–885.
- McKane, A., Alonso, D., Solé, R.V., 2000. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E* 62 (6), 8466–8484.
- Morton, R.D., Law, R., Pimm, S.L., Drake, J.A., 1996. On models for assembling ecological communities. *OIKOS* 75 (3), 493–499.
- Pimm, S.L., Lawton, J.H., 1980. Are food webs divided into compartments? *J. Animal Ecol.* 49, 879–898.
- Rikvold, P., Zia, R.P., 2003a. Flicker noise in a model of coevolving biological populations. In: Landau, D., Lewis, S., Schüttler, H.B. (Eds.), *Computer Simulation Studies in Condensed-Matter Physics XVI*. Springer, Berlin. in press, preprint at arXiv:nlin.AO/0303010.
- Rikvold, P.A., Zia, R.K.P., 2003b. Punctuated equilibria and  $1/f$  noise in a biological coevolution model with individual-based dynamics. *Phys. Rev. E* 68 (031913).
- Yedid, G., Bell, G., 2002. Macroevolution simulated with autonomously replicating computer programs. *Nature* 420, 810–812.
- Zia, R.K.P., Rikvold, P.A., 2004. Fluctuations and correlations in an individual-based model of biological coevolution. *J. Phys. A: Math. Gen.* 37, 5135–5155.