

Emergent Ecology: Neutral Theory of Ecosystems

Zihan Wang

May 2021

Abstract

This essay discusses the neutral theory (NT) of ecosystems, from the aspects of analytical model, simulation results and comparison to empirical data. Certain macroecological properties are shown to be emergent from NT. It also discusses how the neutrality itself could be viewed as an emergent property from evolutionary dynamics, and how can the NT be possibly be reconciled with the classical niche theory.

1 Introduction

Since the time of Charles Darwin, people have been amazed by the idea, that a system as complex as the ecosystem could exhibit deep universalities over wide space and time scales. On one hand, such systems consist a large population of individual units, that interact with one another, and also with the environment, through various complicated mechanisms; on the other hand, data collected from the field has revealed general and relatively simple patterns, that can characterize very different ecological communities around the world. One classical example of such patterns is that the rank-abundance curve of different closed canopy forests seem to represent a single family of mathematical functions (Fig. 1). The existence of such patterns suggests that the essential properties of certain communities might be well captured by a simple model with few parameters. People thus have been seeking theories that relate key biological features at the individual scale to the macroecological features at the community scale.

From the classical viewpoint of niches, the community is described by deterministic models that contains species-specific parameters. These models have been relatively well studied on smaller systems, but as the number of species increase, they quickly become analytically difficult and computationally expensive. Therefore, for such models it remains challenging to study the macroecology, or to understand the universal patterns.

On the contrary, the neutral theory (NT) assumes all species and individuals in a community are ecologically equivalent, meaning that they have identical per-capita probabilities of birth, death, migration, speciation, etc. The validity of neutrality in ecological communities has been controversial since NT was first proposed, but the idea has spawned an active field of study and a fair amount of work, especially since Hubbell's monograph came out in 2001[1][2]. It is hard to summarize these work into a specific framework of theory. They are more like different theories that share the common ideology of ecological patterns can be described as emergence from neutrality of individuals. And from a physicist's view, NT is appealing because it is a basic theory that provides the essential ingredients of an ecosystem, where the emergent properties could be further explored.

In this essay, I will discuss two macroecological patterns that were observed in the experiments, and show how NT can derive these patterns as emergence from

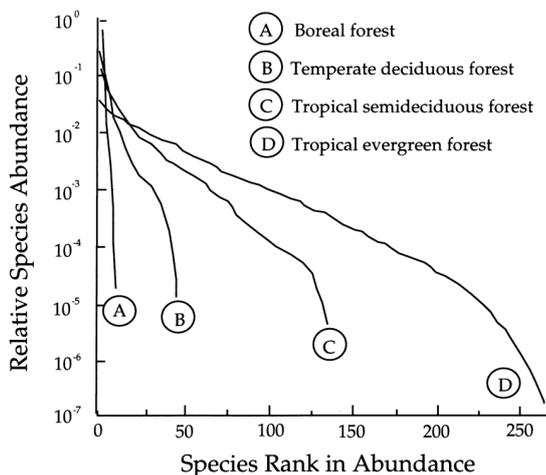


Figure 1: Dominance-diversity curves for tree species in four closed canopy forests, spanning a large latitudinal gradient. [1].

neutrality, along with how NT’s predictions are compatible with the data. I will then discuss how the controversial concept of neutrality can be viewed as emergent from non-neutral evolutionary dynamics, and how that could possibly help to reconcile the niche and the NT perspectives.

2 Emergence of Relative Species Abundance Pattern

The relative species abundance (RSA) pattern is defined as the probability that a species has n individuals in a given region. When multiplied by the total number of species in the region, this gives the number of species with n individuals. The RSA pattern is one of the most commonly used descriptor of static biodiversity, and has been measured for various empirical ecological communities, such as the ones reflected in Fig. 1.

2.1 Experimental Results

The first insights of the RSA pattern being universal came out in the 1940s[3][4]. The observed RSA patterns were fitted with a mode-less distribution “log-series” or a lognormal distribution. Since then there has been many experimental studies on the RSA, and most of them could also be fitted with these shapes. Successfully predicting these shapes has been one of the major successes of the NT.

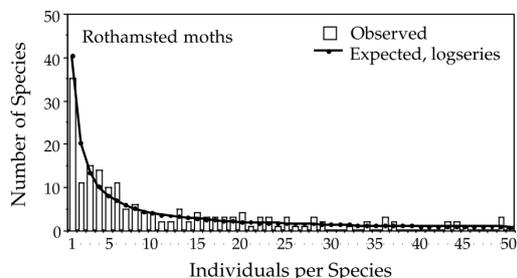


Figure 2: An example of the use of the log series distribution to fit data on species abundance in collections of months at light trap over a 4-year period at Rothamsted Field Station, U.K.[3]

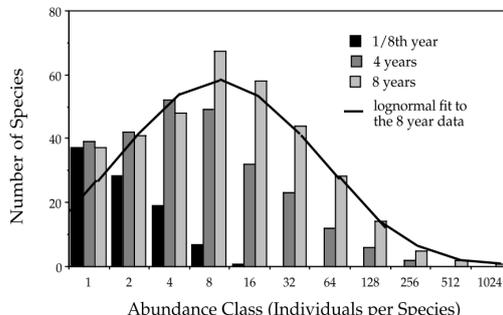


Figure 3: As the survey of months at light traps at Rothamsted Field Station was extended over more years, the distribution of individuals per species became lognormal, as Preston predicted[1]. Note the logarithmic x axis.

2.2 A Null Model

It is instructive to first have a null expectation of the RSA pattern. Assuming perfect neutrality, and the species being independent and randomly distributed, we shall re-

sult in a Poisson distribution of mean ρA , where ρ is the overall density of species within the concerned area A . This deviates from the empirical observations, suggesting that there must be non-trivial ecological mechanisms driving the community towards the log-series or lognormal distributions.

2.3 A Neutral Model

Information in this section comes from [5].

To account for these ecological mechanisms, we can consider four distinct processes: birth, death, migration, and speciation. They form the set of basic elements that is the simplest one to give the empirical distributions. We also assume that all the above mechanisms apply equally to all individuals regardless of which species they belong to.

In the continuous time limit, the community could be described by a simple master equation

$$dP_k/dt = A_{k-1}P_{k-1} + C_{k+1}P_{k+1} - (A_k + C_k)P_k \quad (1)$$

where $P_k(t)$ is the probability that k individuals belong to the same species at time t , and the coefficients A_k and C_k represent respectively the increase and decrease of a species with abundance k during the lifetime of an individual.

We will first consider a metacommunity with zero-sum dynamics, which means it has the summed abundance across all species strictly constant, and speciation present. Then, we will consider a local community, that is in contact with the metacommunity, and has the average total population fixed. The two scenarios roughly corresponds to the microcanonical and canonical ensemble in statistical physics.

2.3.1 Metacommunity

In the zero-sum dynamics of the metacommunity, the coefficients in master equation are as follows:

$$A_k = \frac{(J_M - k)k}{J_M(J_M - 1)}(1 - \nu), \quad (2)$$

$$C_k = \frac{k[J_M - k + (k - 1)\nu]}{J_M(J_M - 1)}, \quad (3)$$

where J_M is the total abundance, and ν is the speciation rate. The increase A_k is given by the probability at each timestep that the death happens in another species ($(J_M - k)/J_M$), times the probability that the birth happens in the investigated species ($k(J_M - 1)$), times the probability that it does not mutate ($1 - \nu$). C_k can also be derived in a similar manner.

The average number of species at abundance k and time t could be noted down as $\langle \phi_k(t) \rangle_M$. It could be calculated by summing up all species that were speciated at time $t - u$ and reached size k at time t . Therefore,

$$\langle \phi_k(t) \rangle_M = \int_0^t P_k(t-u)p(u)du = \nu \int_0^t P_k(u)du. \quad (4)$$

The equilibrium distribution of $\langle \phi_k(t) \rangle_M$ at a given overall abundance J_M and initial condition $P_k(0) = \delta_{k,1}$ could be analytically solved:

$$\langle \phi_k \rangle_M = \frac{\theta \Gamma(J_M + 1) \Gamma(J_M + \theta - k)}{k \Gamma(J_M + 1 - k) \Gamma(J_M + \theta)}, \quad (5)$$

where the key parameter θ is defined as

$$\theta = \frac{(J_M - 1)\nu}{1 - \nu}. \quad (6)$$

When $J_M \gg 1$ and $\nu \ll 1$, θ becomes $J_M\nu$, which reflects the biodiversity from the mechanism aspect, and is called as the “fundamental biodiversity number in [1].

For large J_M and finite θ , the density of species with relative abundance $\omega = k/J_M$ could then be expressed as

$$g_M(\omega) = \frac{\theta(1 - \omega)^{\theta-1}}{\omega}. \quad (7)$$

However, it is very hard to acquire any statistics of the metacommunity. We will need to assess the local communities in order to make an intuitive comparison with the experimental data.

2.3.2 Local Community

In a local community of size J ($J \ll J_M$), we shall assume the migration rate to be low that there is no migration out of the local community, and the speciation rate to be low so that we can ignore speciations in the local community scale. We now consider $P_n(t; \omega)$, which is the probability of finding a species with population n in the local community, given that its relative abundance in the metacommunity is ω . Here, the ω could be viewed as a constant from the evolutionary timescales of the much smaller local community. With respect to the master equation coefficients, we shall have:

$$A_n = \frac{(J - n)}{J} \left[\frac{n}{J - 1} (1 - m) + m\omega \right], \quad (8)$$

$$C_n = \frac{n}{J} \left[\frac{J - n}{J - 1} (1 - m) + m(1 - \omega) \right]. \quad (9)$$

Analytically solving for the stationary distribution $P_n(\omega)$, there would be

$$P_n(\omega) = \binom{J_L}{n} \frac{(\mu\omega)_n [\mu(1 - \omega)]_{J_L - n}}{(\mu)_{J_L}} \quad (10)$$

where another key parameter

$$\mu = \frac{(J - 1)m}{1 - m} \quad (11)$$

quantifies the migration, in the same way that θ quantifies speciation. $\binom{J_L}{n}$ is the binomial coefficient, and $(a)_n = \Gamma(a+n)/\Gamma(a)$ is the Pochhammer symbol.

It is then possible to calculate the distribution of species $\langle \phi_n \rangle_C$ in a local community:

$$\langle \phi_n \rangle_L = \sum_{k=1}^{J_M} P_n(k/J_M) \langle \phi_k \rangle_M \quad (12)$$

Similar to the metacommunity, in the continuum limit of large J , the density of species in the local community $g_L(\omega) = \lim_{J \rightarrow \infty} J \langle \phi_n \rangle_L$ could be derived as

$$g_L(\omega) = \mu \theta \int_0^1 \binom{\mu}{\mu u} (1-\omega)^{\mu u - 1} \omega^{\mu(1-u) - 1} u^\theta du. \quad (13)$$

It could be shown that this result, or the zero-sum multinomial distribution, is close to the lognormal distribution for a local community, and becomes asymptotically close to the logseries at large J limit.

3 Emergence of SAR Patterns(to be decided)

As shown in the last section, the emergence of the RSA distribution could be well explained by a “zero-dimensional” model, in the sense that it does not include any description of spatial inhomogeneity of species. However, space is also an essential element for understanding the ecosystem. Spatial structure of ecosystems has been known to control the shape of many macroecological patterns, and it is also a critical factor for understanding species’ coexistence[2]. Most empirical observations are also spatial.

Therefore, a neutral theory that addresses spatial descriptors is needed.

Currently, no coherent spatial NT has been proposed, instead there exists a collection of models that can explain some of the spatial patterns. Among the spatial descriptors, the species-area relationship (SAR) is the most commonly investigated one. It is defined as the average number of species $\langle S(A) \rangle$ sampled in an area A .

3.1 Experimental Results

In the 1920s, a power-law relationship $\langle S(A) \rangle = cA^z$ was postulated based on empirical data, and later experiments gave an inverse S-shaped curve[7][8]. As shown

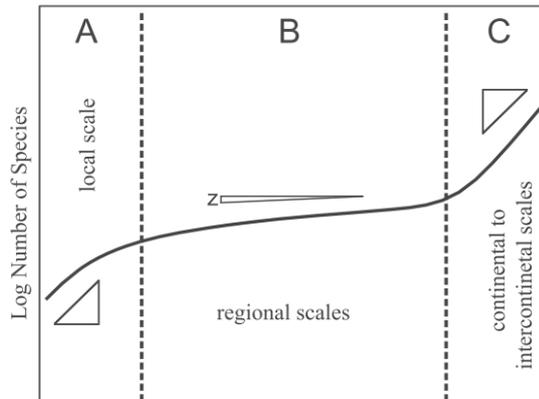


Figure 4: Three behaviors of the SAR on different scales[6].

in Fig. 4, the SAR follows a power law with the exponent z at intermediate scales, while on the small and large areas a linear behavior is displayed. Across different experiments, the value of the exponent z varies a lot, from 0.1 to 0.5, and has been observed to show dependence on multiple environmental factors[2].

3.2 Neutral Models

3.2.1 Phenomenological Models

Phenomenological models do not assume specific processes at the microscopic or individual level, instead they are based on phenomenological assumptions of the spatial distributions.

For example, the simplest way is to assume that all the individuals are randomly distributed in space. Therefore, if we denote there are $S(A_0)$ species in the region A_0 , and that species i has abundance n_i , then for a subsection A of the area A_0 , we shall have the following SAR:

$$\langle S(A) \rangle = S(A_0) - \sum_{i=1}^S \left(1 - \frac{A}{A_0}\right)^{n_i}. \quad (14)$$

From the results of RSA in the last section, one could show that the species abundance distribution could be asymptotically described by the Fisher's log series:

$$\langle \phi(n) \rangle = SP(n) = \theta r^n / n \quad (15)$$

where r is a parameter related to birth and death rates. In this way, the SAR could be then explicitly written as

$$\begin{aligned} \langle S(A) \rangle &= S(A_0) - \sum_{n=1}^{\infty} \theta \frac{r^n}{n} \left(1 - \frac{A}{A_0}\right)^n \\ &= \theta \log \left(1 + \frac{r}{1-r} \frac{A}{A_0}\right). \end{aligned} \quad (16)$$

To reproduce the inverse S-shaped SAR curve observed in experiments, we will also need to take local conspecific clusters into consideration. In real ecosystems, due to dispersal limitation, individuals of the same species tend to be clustered. This could be described by a Poisson cluster process, where the center of clusters are randomly distributed in space, and a random number of individuals are positioned around them by a given spatial distribution $\phi(\mathbf{r})$ with respect to the center. By this process, the SAR could be calculated as[6]:

$$S(A) = S(A_0) \int d^2\mathbf{r} \left[1 - \int_0^{\infty} d\lambda p(\lambda) e^{-\lambda \int_{A(\mathbf{r})} d^2\mathbf{r}' \phi(\mathbf{r}')} \right] \quad (17)$$

where $p(\lambda)$ is determined by the RSA of the metacommunity. This approach reproduces the inverted S-shaped curve, showing that this shape can be explained simply from spatial correlations of individuals of the same species.

3.2.2 Stochastic Models

There are several models that aim to reproduce the empirical inverted-S curve from spatially explicit processes at the individual level. These models typically contain births, deaths and diffusions. They are analytically intractable and could only have asymptotic or computational results.

In [9] a model called the multispecies voter model (MVM) with speciation is introduced. It describes a 2D lattice where at each time step one randomly chosen individual is removed, and then replaced with a copy of one of other individuals with a probability $1 - \nu$, or with an individual of a species that is not currently absent in the system with a speciation probability ν .

Specifically, the replacement probability by a copy of the present individual is determined by the distance to the removed individual, rather than the nearest-neighbor diffusion. This dependence is characterized by a dispersal kernel[2]. Up to the choice of the dispersal kernel, the SAR could be described by two parameters: the speciation rate ν , and the dispersal length ξ . Computational results have shown that in an infinite landscape, the SAR from this model reproduces the characteristic inverted-S shape curve, and asymptotically it scales as[9]:

$$\langle S(A, \xi, \nu) \rangle = \xi^r S(A\xi^{-r}, \nu) \quad (18)$$

where the exponent r is independent of the two parameters.

Simulations also show that the exponent z in the SAR curve could be fitted with a logarithmic relationship with ν [2]:

$$z = \frac{1}{q + m \log(\nu)}. \quad (19)$$

where q and m are real parameters. This has also been observed in other spatially explicit models with different mechanisms.

4 Reconciling Niche and Neutrality

Neutral models have successfully described many macroecological patterns as the result of emergence from identical individuals governed by simple mechanisms. However, the assumption of neutrality is not always necessarily true, and there are plenty of empirically observed patterns that cannot be explained by NT alone. It is generally believed that most patterns of natural communities are generated by a “cocktail” of processes, involving both niche-based processes that emphasize the different across species, and stochastic demographical processes that feature neutrality.

From the perspective of NT, it is informative, then, to study under what circumstances can the ecology drive the community to a state, where stochasticity plays a crucial role in determining the community observables, despite the strong difference across the species, and how the state is reached. There have been several approaches along this direction of reconciling the NT and the classical niche theory under one framework[2]. I will present two of them here.

4.1 Emergent Neutrality

Information in this section comes from [10].

The most popular phenomenological model of describing species competition is the Lotka-Volterra (LV) equations:

$$\frac{dN_i}{dt} = rN_i \left(K_i - \sum_j \alpha_{i,j} N_j \right) / K_i \quad i = 1, 2, \dots, n; \alpha_{i,i} = 1, \quad (20)$$

where N_i is the density of species i , r is the universal maximum per capita growth rate, K_i is the carrying capacity of species i , and $\alpha_{i,j}$ is the competition coefficient that quantifies the effect of species j on species i .

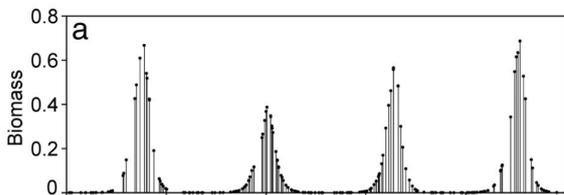


Figure 5: A stable pattern of species forming lumps. The x-axis represents the niche (body size).

Suppose the competition coefficients were set to have species lying on an one-dimensional niche axis, such as body size. In this way, an intuitive expectation based on competitions would be that the surviving species at equilibrium would be equally spread on the niche axis. Instead, numerical simulations showed a transient state of equally spread lumps, and within each lump there are multiple very similar species, as shown in Fig. 5.

The lumps are formed, because the niche similarity of species within each single lump prevents competitive exclusion from quickly selecting the best competitor among a group of similar species, allowing their coexistence for very long times, even though in the final equilibrium state only the superior species will persist.

There are mechanisms that can maintain this pattern of lumps. For example, the LV equations can be adapted to include a regulation as follows: The reproduction is reduced up to g when population density exceeds a threshold H :

$$\frac{dN_i}{dt} = rN_i \left(K_i - \sum_j \alpha_{i,j} N_j \right) / K_i - g \frac{N_i^2}{N_i^2 + H^2} \quad (21)$$

In this way, the lump patterns become permanent. This model shows a possibility where species that are initially different in ecological traits, and that therefore behave in a non-neutral fashion, can be driven by evolutionary processes to form states that can be well approximated by neutral models over appropriate spatial and temporal scales.

4.2 Phase Transition Between Niche and Neutrality

Information in this section comes from [11].

To construct permanent neutrality from the LV system, apart from the adaptation mentioned in the last section, there's also an intuitive way, which is to add a stochastic term — a community could be seen as “statistically neutral” if its multivariate distribution of species abundances cannot be distinguished from a distribution constructed under the assumption of ecological neutrality. This could be compared to a physical system: as higher temperature brings in a higher degree of stochasticity, the energetic differences caused by interactions would have less influence in the final states. Therefore, by including a Gaussian noise as the stochastic factor, and introducing the immigration λ , the LV equations become:

$$\frac{dN_i}{dt} = \lambda + rN_i \left(K_i - \sum_j \alpha_{i,j} N_j \right) / K_i + \sqrt{\omega N_i} \eta_i(t) \quad i = 1, 2, \dots, n, \quad (22)$$

where the noise has $\langle \eta_i(t) \rangle$ and $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$.

To measure the degree of statistical neutrality, the distance from the adapted LV-simulated species abundance distribution ($P_{LV}(\vec{x})$) and the one generated by purely neutral dynamics ($P_N(\vec{x})$) is calculated. The results are shown in Fig. 6, where the neutral and niche regimes become two macroscopic regimes divided by a phase transition.

5 Discussions

In this essay, I presented how the patterns of two of the most important macroecological descriptors can be derived from NT, which can be viewed as success of NT. I also discussed two approaches to integrate the niche theory and NT into one framework, which is a future direction of NT.

These are far from the whole picture of the field. For example, the above discussions on RSA and SAR are examples of NT at stationarity. There has been both classical and recent work on the temporal behaviors of the NT, and they have been able to explain some of the dynamical patterns in ecological communities in reality. Another active direction is to understand the link of different macroecological patterns that has been found from the data. In particular, we still lack a mechanistic understanding of the effect of spaces and spatial patterns, and no spatially explicit models have been solved.

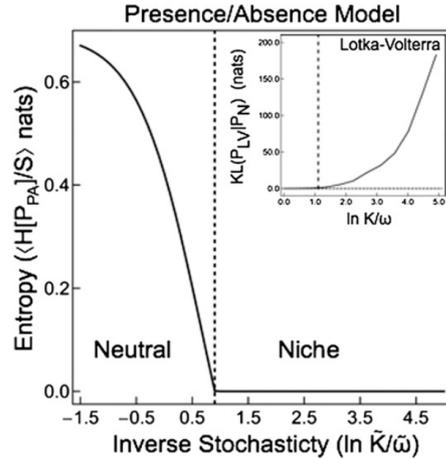


Figure 6: Phase diagram displaying Neutral and Niche regimes. The inset shows the behavior of the LV model.

There has been plenty of work in marking the validity of the neutrality assumption, while another strong assumption of NT has been less discussed, which is that the system can only have one trophic layer. The NT is then limited to species that compete with each other on the same pool of limited resources. However, the architecture of ecological interaction networks plays a crucial role in shaping and regulating community dynamics, and it has become a bubbling field of study. It is a major challenge to connect stochastic NT and ecological networks structure within a unified theoretical framework. Therefore, the NT remains to be an active field with many questions to be answered.

References

- [1] Stephen P Hubbell. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. Princeton University Press, 2011.
- [2] Sandro Azaele, Samir Suweis, Jacopo Grilli, Igor Volkov, Jayanth R Banavar, and Amos Maritan. Statistical mechanics of ecological systems: Neutral theory and beyond. *Reviews of Modern Physics*, 88(3):035003, 2016.
- [3] Ronald A Fisher, A Steven Corbet, and Carrington B Williams. The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, pages 42–58, 1943.
- [4] Frank W Preston. The commonness, and rarity, of species. *Ecology*, 29(3):254–283, 1948.
- [5] M Vallade and B Houchmandzadeh. Analytical solution of a neutral model of biodiversity. *Physical Review E*, 68(6):061902, 2003.
- [6] Jacopo Grilli, Sandro Azaele, Jayanth R Banavar, and Amos Maritan. Spatial aggregation and the species–area relationship across scales. *Journal of theoretical biology*, 313:87–97, 2012.
- [7] Olof Arrhenius. Species and area. *Journal of Ecology*, 9(1):95–99, 1921.
- [8] FW Preston. Time and space and the variation of species. *Ecology*, 41(4):611–627, 1960.
- [9] James Rosindell and Stephen J Cornell. Species–area relationships from a spatially explicit neutral model in an infinite landscape. *Ecology letters*, 10(7):586–595, 2007.
- [10] Marten Scheffer and Egbert H van Nes. Self-organized similarity, the evolutionary emergence of groups of similar species. *Proceedings of the National Academy of Sciences*, 103(16):6230–6235, 2006.

- [11] Charles K Fisher and Pankaj Mehta. The transition between the niche and neutral regimes in ecology. *Proceedings of the National Academy of Sciences*, 111(36):13111–13116, 2014.