PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Review

Check for updates

Article submitted to journal

Subject Areas:
92D40, 92D25, 05C90, 34C35, 34D05, 60J80, 60F17

Keywords:
random matrices; dynamical systems; faisability; stability; structured networks; population dynamics

Author for correspondence:
Jamal NAJIM
e-mail: jamal.najim@univ-eiffel.fr

Complex systems in Ecology: a guided tour with large Lotka-Volterra models and random matrices

Imane Akjouj[(1)], Matthieu Barbier[(2)], Maxime Clenet[(3)], Walid Hachem[(3)], Mylène Maïda[(1)], François Massol[(4)], Jamal Najim[(3)], Viet Chi Tran[(5)]

[1] Université de Lille,
[2] CIRAD and Institut Natura e Teoria - Pirenèus,
[3] Université Gustave Eiffel and CNRS,
[4] CNRS, Université de Lille, INSERM, CHU, Institut Pasteur Lille
[5] Université Gustave Eiffel

Ecosystems represent archetypal complex dynamical systems, often modelled by coupled differential equations of the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \phi_i(x_1, \cdots, x_N)\,,$$

where N represents the number of species and x_i, the abundance of species i. Among these families of coupled differential equations, Lotka-Volterra (LV) equations, corresponding to

$$\phi_i(x_1, \cdots, x_N) = r_i - x_i + (\Gamma \mathbf{x})_i,$$

play a privileged role, as the LV model represents an acceptable trade-off between complexity and tractability. Here, r_i is the intrinsic growth of species i and Γ stands for the interaction matrix: Γ_{ij} represents the effect of species j over species i. For large N, estimating matrix Γ is often an overwhelming task and an alternative is to draw Γ at random, parameterizing its statistical distribution by a limited number of model features. Dealing with large random matrices, we naturally rely on Random Matrix Theory (RMT).

© The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

The aim of this review article is to present an overview of the work at the junction of theoretical ecology and large random matrix theory. It is intended to an interdisciplinary audience spanning theoretical ecology, complex systems, statistical physics and mathematical biology.

1. Introduction: Complex networks, randomness and large dimension

(a) Goals of this study

Ecosystems can be seen as archetypal complex dynamical systems, as they usually consist of a large number of interacting components with heterogeneous properties. In the present article, these components are species or sub-populations that evolve according to their own demographic dynamics and interact through various mechanisms (such as competition, predation or facilitation). These dynamics generally take the form:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \phi_i(x_1, \cdots, x_N) \tag{1.1}$$

where $x_i(t)$ represents the (dimensionless) abundance or density of species i population, the function ϕ_i encapsulates all the sources of growth and mortality, and N is the (large) number of interacting species. We focus in this article on the ***Lotka-Volterra*** (LV) model [1,2] which writes:

$$\phi_i(\boldsymbol{x}) := \phi_i^{LV}(\boldsymbol{x}) = r_i - x_i + (\Gamma \boldsymbol{x})_i \ , \tag{1.2}$$

where r_i represents the intrinsic growth rate of species i and $\Gamma = (\Gamma_{ij})$ is a $N \times N$ matrix with Γ_{ij} representing the effect of species j on species i, see Section 2 for a detailed introduction of the LV model.

Historically, LV model was designed to understand the interaction between two species ($N = 2$), which abundances often show an oscillating behavior (think of the celebrated Hare-Lynx example [3]). In this study, we will focus on large foodwebs ($N \gg 1$), equilibria and their stability. Oscillating behaviors in large dimension are beyond the scope of this survey but are discussed in the "open problems" section (Section 6(b)). In the Supplementary Material File (referred to as *Supp-Mat file* in the sequel) [4, Section 2], we provide a brief presentation of the $N = 2$ case.

Notice that ϕ_i, often referred to as the ***net growth rate***:

$$\phi_i = \frac{1}{x_i} \frac{\mathrm{d}x_i}{\mathrm{d}t}$$

represents the growth rate of abundance x_i, i.e. its per capita rate of abundance change.

Equation (1.1) incorporates two essential properties of biological dynamics: first, a species can be extinct (equilibrium at $x_i = 0$), and second, a small population $x_i \to 0$ displays an exponential growth or decay with rate $\phi_i \,|_{x_i=0}$ referred to as the ***invasion growth rate*** (e.g. [5,6], see also the Supp-Mat file [4, part 4(d)]).

To analyze and model such complex dynamical systems, a set of formal tools, including network theory, dynamical systems and Random Matrix Theory (RMT), have proven successful across a variety of scientific disciplines. This review is meant to offer a constructive viewpoint on the connection between these mathematical tools, especially RMT, and ecological systems, addressed to an interdisciplinary audience spanning theoretical ecology, complex systems, statistical physics and mathematical biology. We propose to focus on formal problems inspired by complex ecosystems, with the ultimate aim of answering ecological questions regarding the conditions of species coexistence, community diversity and ecosystem stability.

Through the prism of the Lotka-Volterra model with random interactions, our aim here is to give a diverse overview of concepts and questions that have proven fruitful in that line of work. In particular, we wish to point out results that we conjecture may also hold qualitatively, or even

quantitatively, beyond this particular model, in the hope that such "universal"[1] behaviors might be shown in the future to capture some aspects of real ecosystems.

The connection of this model to empirical data is an important question which is beyond the scope of this review. We discuss this aspect in the discussion section (Section 6).

(b) Historical context

The first eminent proponent of applying results from complex dynamical systems to ecosystems was Robert May [7]. Inspired by the success of RMT for modelling unknown interactions in complex physical systems, such as large atomic nuclei [8], he argued as follows:

Obs Empirical observation: ecosystems with a large number of species appear to exist and persist for long times.

Hyp 1 If we assume that species abundances are poised at some dynamical equilibrium $\boldsymbol{x}^* = (x_1^*, \cdots, x_N^*)$ with all $x_i^* > 0$

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i^* \phi_i(\boldsymbol{x}^*) = 0 \tag{1.3}$$

Hyp 2 If we assume that their interactions are *sufficiently complex* to be modelled as random, i.e. more precisely, that the Jacobian matrix governing the linearized dynamics around the equilibrium

$$J_{ij} = \frac{\partial(x_i \phi_i)}{\partial x_j}(\boldsymbol{x}^*) \tag{1.4}$$

is modelled as

$$\begin{cases} J_{ii} &= -1 + M_{ii} C_{ii} \\ J_{ij} &= M_{ij} C_{ij} \qquad (i \neq j) \end{cases} \qquad i.e. \qquad J = -I + C \circ M$$

where the random variables M_{ij} are independent identically distributed (i.i.d.) centered random variables of variance V, C_{ij} are i.i.d. with Bernoulli distribution with parameter C, called the connectance (such that a species has $N \times C$ links on average), and $\circ$ denote the Hadamard product of matrices (pointwise multiplication of entries),

Res Then the equilibrium is stable for $NCV < 1$ and unstable for $NCV > 1$.

A formal presentation of May's result, in particular at a probabilistic level and emphasising the role of the large dimension of N, is provided in Supp-Mat file [4], (Section 1, Prop. 1.2).

May presented his conclusion as a paradox: earlier ecologists had imagined the observed complexity of natural ecosystems as a positive feature favoring their persistence (i.e. the more connected and/or strongly interacting species were, the more likely they would coexist), which could be contradicted by the loss of stability predicted in May's simplistic model. May's result, however, did not spell out where in the above argument lies the "paradox", i.e. which part of the argument should be reevaluated in the light of empirical evidence, or instead, should lead us to reevaluate earlier assumptions about ecosystems.

This result had a lasting influence on the theoretical ecology literature, yet it never received clear support from ecological data [9–12]. After that, theoretical and empirical work on this topic has been diverse in scope and focus, but can be interpreted as questioning each of the four points summarized above[2]. In response to **Obs**, ecosystems might be less diverse and complex than they appear [13] (e.g. few species may coexist at any location, or species may exist in large number N but with low interaction variance V). Contra **Hyp 1**, real ecosystem dynamics might be far from

[1]Beware that in RMT, universality has a different meaning: for a given random matrix, a result is universal if it does not depend on the distribution of the underlying random variables but only on its first moments.

[2]May's own interpretation was probably focused on the third (Nature had to employ "devious strategies", in his words, to allow high-diversity ecosystems to exist in a stable equilibrium).

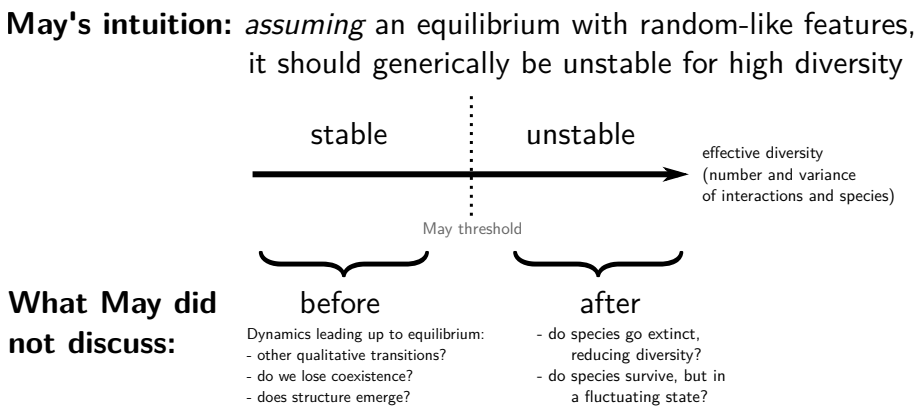


Figure 1: Schematic cartoon of early arguments from randomness in complex ecological systems. Robert May (1972) used a mathematical result to anchor a broader qualitative intuition: assuming a system with random features at dynamical equilibrium, there is an axis of effective diversity (equal to the product NCV in his model), combining system size (species number) and statistics (notably variance) of interaction, along which a qualitative transition from stability to instability is expected. Yet how likely is such an equilibrium to arise in the first place, and are there other qualitative transitions along this axis (or other axes that are as broadly relevant)?

equilibrium, e.g. dominated by transient excursions [14]. Contra **Hyp 2**, ecological interactions might be structured so that the resulting dynamics deviate importantly from predictions from full randomness [15] (suggesting a greater role of non-random structure in ecology). Finally, even admitting **Res**, the loss of stability when $NCV > 1$ does not imply the extinction of species: the ecosystem may still persist in a steady out-of-equilibrium state [16], and thus this mathematical statement may not be interpreted ecologically as a limit on a realistic number of species.

The discussion that May's result generated [17,18] also highlights the issues arising from misunderstanding mathematical notions of equilibrium existence, equilibrium stability, and system persistence, as well as the gap between measurable quantities in ecology and those relevant to the analysis of ODEs [19–21]. However, the "May-Wigner transition" was an influential result because, while it was proven for a particular case (a random linearized dynamical system), there were many reasons – both mathematical and empirical, from successes of RMT in other fields – to conjecture that it could have far broader generalizations.

(c) Qualitative questions

May's original work did not actually state what happens in arbitrary nonlinear ecological dynamics, besides the fact that it may exhibit a transition to instability (cf. Fig. 1). For instance, can some species go extinct without a loss of stability? [22–24].

One aim of this review is to show that such questions, which cannot be addressed in May's setting, can be and have been fruitfully studied in subsequent decades using the LV model (2.1). A model that is simple, can be derived as an approximation of individual-based models (see Supp-Mat file [4, Section 3]), and is rich enough to demonstrate various dynamical phenomena.

Another important line of inquiry that has grown in response to May's work on stability has focused on *imposed structure*, searching for various deviations from full randomness that might possibly allow to restore feasibility, stability, or both. These structural features can be motivated by empirical networks, such as reproducing in broad strokes the hierarchical or group structure of a foodweb [25], the size-dependence of trophic interactions [26], or accounting for the spatial structure of ecosystems [27,28]. They can also arise from theoretical mechanisms, such as

interactions deriving from underlying ecological traits and niches [29], or their evolution [30,31]. We will discuss results and questions arising when we allow a combination of deterministic structure and randomness in Lotka-Volterra systems, see Sections 4 and 6(a).

(d) Outline of the article

In Section 2, we introduce the LV system of coupled differential equations. We discuss and motivate the large random interactions model, and introduce two such models, with independent interactions (i.i.d. model), and correlated reciprocal interactions (elliptic model). In Section 3, we present results for the two reference models (i.i.d. and elliptic) and address the question of the existence of a unique equilibrium $\boldsymbol{x}^* = (x_i^*)$ for a LV system with random interactions and its stability, that is the conditions for which $\boldsymbol{x}(t) \to \boldsymbol{x}^*$. We also focus on the feasibility of the equilibrium, that is the conditions for which no species vanishes at the equilibrium: $x_i^* > 0$ for all i. If feasibility is not reached, we describe the composition of the equilibrium in terms of surviving species $x_i^* > 0$ and vanishing species $x_i^* = 0$, together with various properties of the equilibrium. Beyond i.i.d. and elliptic models, there is a need to consider more realistic structures of interaction matrices which take into account important features of real foodwebs. In Section 4, we focus on sparse models and kernel matrices. We present a quick survey of the related literature in theoretical ecology and recall associated basic RMT results. In Section 5, we extend Lotka-Volterra models by adding extra randomness yielding to stochastic differential equations (SDEs).

Finally, we summarize in Section 6 what has been covered in this review and what has not. We also present some directions of investigation and open problems of interest.

We briefly describe the Supplementary Material File [4] which contents complete the main exposition. In [4, Section 1], we present May's model. In [4, Section 2], we present the deterministic LV model of size 2, for pedagogical purpose. [4, Section 3] establishes the connexion between individual-based models and LV model, while [4, Section 4] shows how individual-based models studied at different asymptotics yield community models with noise. In [4, Section 5], we provide mathematical details related to Section 4, in particular a certain type of deterministic networks and kernel matrices. More precisely, concerning the latter, we emphasize the connexion between general kernel matrices and Marchenko-Pastur distribution associated to large random covariance matrices.

2. The Lotka-Volterra model

(a) The Lotka-Volterra system of differential equations

We are interested in many-species ecological dynamics and our main object of study will be the following LV system of differential equations:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left(r_i - x_i + (\Gamma \boldsymbol{x})_i \right) , \tag{2.1}$$

where $i \in [N] := \{1, \cdots, N\}$ and $\boldsymbol{x} = (x_1, \cdots, x_N)$. The parameter N represents the number of species, supposed large, $x_i = x_i(t)$ is a dimensionless quantity evolving in time t, in relation with the abundance of species i, r_i represents the intrinsic growth of species i and $\Gamma = (\Gamma_{ij})$ is a $N \times N$ matrix reflecting the interaction effect of species j on the growth of species i:

$$(\Gamma \boldsymbol{x})_i = \sum_j \Gamma_{ij} x_j \, .$$

Notice that if there are no interactions, i.e. $\Gamma = 0$, we recover for each species the standard logistic (Verhulst) differential equation which in particular reflects the limited amount of available resources.

Using $\phi_i^{LV}(\boldsymbol{x}) := r_i - x_i + (\Gamma \boldsymbol{x})_i$ introduced in (1.2), Eq. (2.1) follows the generic form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \phi_i^{LV}(\boldsymbol{x}),$$

introduced in (1.1) with ϕ_i^{LV} representing the net growth rate of species i.

As mentioned above, we detail in [4, Section 3] how this LV model naturally emerges from an individual-based model when we consider a certain asymptotics of the birth, death and interaction rates. Let us also formalise in Remark 2.1 below how one can obtain the dimensionless, normalized system of equations (2.1), where each species has a comparable importance, from a similar system dealing with the actual abundances of species.

Remark 2.1 (Obtaining the dimensionless LV model (2.1))**.** *In ecology, a natural formalization of the LV model involving the actual abundances or densities X_i of the modelled species reads as:*

$$\frac{\mathrm{d}X_i}{\mathrm{d}t} = X_i \left[r_i - D_i X_i + (M\boldsymbol{X})_i \right], \tag{2.2}$$

where r_i is species i intrinsic growth rate, D_i is species i density-dependent term and $\boldsymbol{M}$ is the matrix of interaction coefficients. In order to obtain the dimensionless version of the Lotka-Volterra model presented in Equation (2.1), the following changes of variables are needed:

$$x_i := D_i X_i, \quad \Gamma_{ij} := M_{ij}/D_j. \tag{2.3}$$

Plugging these new variables into Eq. (2.2), one obtains Eq. (2.1).

In a large ecosystem consisting of N species ($N \gg 1$), the precise knowledge of the interaction matrix $\Gamma = (\Gamma_{ij})$ among these species is often out of reach. An interesting alternative is to model the $N \times N$ matrix Γ with random entries and to rely on RMT. The statistical properties of the entries may then reflect a partial knowledge on the ecological interaction network.

(b) Two random interactions models

In this section, we precisely describe two models for the matrix Γ with random interactions. The first is the simplest theoretical baseline of i.i.d. entries with zero mean, while the second model is a natural extension that allows to represent more types of ecological interactions [32]. Other (more involved) models will be discussed in Section 4.

Independent and Identically Distributed entries.

In this model, the entries Γ_{ij} are i.i.d., with a N-dependent common distribution and can be:

$$(i) \quad \Gamma_{ij} = \frac{A_{ij}}{\sqrt{N}} \qquad \text{or} \qquad (ii) \quad \Gamma_{ij} = \frac{A_{ij}}{\alpha_N \sqrt{N}} \tag{2.4}$$

where the A_{ij}'s are i.i.d., $\mathbb{E}\, A_{ij} = 0$, $\mathbb{E}\, A_{ij}^2 = 1$ and a distribution independent from N.

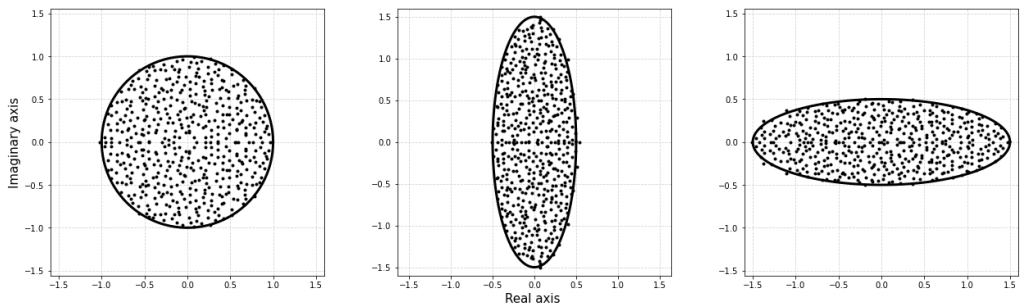
In the case (i), the $N^{-1/2}$-normalization casts matrix $\Gamma = \frac{A}{\sqrt{N}}$ into the framework of RMT, where the limiting properties of the spectrum and eigenvectors of matrix Γ are well described.

The circular law (cf. [33]) asserts that the spectrum of Γ converges toward the uniform distribution on the disk of radius 1, see Fig. 2-(a). Moreover, denote by $\rho(\Gamma)$ Γ's spectral radius:

$$\rho(\Gamma) := \max\{|\lambda(\Gamma)|,\ \lambda(\Gamma) \in \mathbb{C} \text{ eigenvalue of } \Gamma\},$$

then its asymptotic behaviour is well-known: $\rho(\Gamma) \xrightarrow[N\to\infty]{a.s} 1$, where $\xrightarrow{a.s}$ stands for the almost sure convergence. As a consequence, matrix Γ has a non-vanishing macroscopic effect on the dynamical system (2.1) even for large N.

In the case (ii), there is an extra normalization term α_N which may or may not depend on N. If it does not depend on N, then it is simply a parameter that allows to tune the variance of the



(a) $\xi = 0$ (b) $\xi = -0.5$ (c) $\xi = 0.5$

Figure 2: Spectrum of 500×500 centered random matrices with (a) $N^{-1/2}$-normalized i.i.d. standard Gaussian entries, (b) elliptic distribution $\xi = -0.5$, (c) elliptic distribution $\xi = 0.5$. The points represent the eigenvalues. The solid curve represents the boundary of the support of (a) the circular law (uniform law on the disc), (b) and (c) the elliptic distribution with parameter ξ.

entries since $\mathrm{var}(\Gamma_{ij}) = \frac{1}{\alpha^2 N}$. If α_N grows to infinity as $N \to \infty$, it has the effect of asymptotically squeezing to zero the contribution of the interaction matrix Γ as

$$\rho(\Gamma) = \frac{\rho(A/\sqrt{N})}{\alpha_N} \xrightarrow[N\to\infty]{} 0\,.$$

Despite this, we will see in Section 3(d) that $\alpha_N^* = \sqrt{2\log(N)}$ is the threshold to reach feasibility (no vanishing species - see the formal definition in Section 3(a)-(ii)).

The elliptic model.

Two assumptions of the i.i.d. model are commonly relaxed to describe a wider range of ecological scenarios. First, while the i.i.d. model assumes that interactions have zero mean, ecological networks often contain interactions with a prescribed sign. Second, the i.i.d. model enforces the reciprocal interactions Γ_{ij} and Γ_{ji} to be uncorrelated. However, a large literature, for random as well as for deterministic interactions, deals with symmetric matrices $\Gamma_{ij} = \Gamma_{ji}$, which can for instance arise in ecology in the case of competition [34], or skew-symmetric matrices $\Gamma_{ij} = -\Gamma_{ji}$ [35], which were originally proposed for predator-prey interactions [2]. The random symmetric case is well-known in RMT and is referred to as the Wigner model, see for instance [36, Chap. 2].

These various cases can be unified into the elliptic model [37–39], which displays a richer statistical structure than the i.i.d. model. The entries of matrix Γ write

$$\Gamma_{ij} = \frac{A_{ij}}{\sqrt{N}} + \frac{\mu}{N}\,, \tag{2.5}$$

where μ is a deterministic quantity, A_{ij} is centered with variance equal to 1. In particular, $\frac{\mu}{N}$ stands for Γ_{ij}'s expectation. The second feature of the model is the existence of a correlation between the random variables A_{ij} and A_{ji} $(i \neq j)$:

$$\mathrm{corr}(A_{ij}, A_{ji}) = \xi \in [-1, 1] \tag{2.6}$$

while A_{kk} and $\{A_{ij}, A_{ji}\}$ are independent for k, i, j and $i < j$.

Under the elliptic model and in the case where $\mu = 0$ and $|\xi| \neq 1$, the spectrum of Γ converges toward the uniform law on the ellipse

$$\mathcal{E}_\xi = \left\{ z = x + \mathbf{i}y,\ \frac{x^2}{(1+\xi)^2} + \frac{y^2}{(1-\xi)^2} \leq 1 \right\}\,,$$

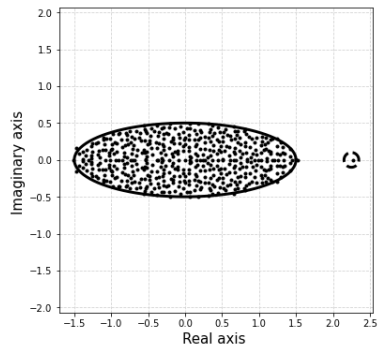


Figure 3: Spectrum of a deformed 500×500 elliptic matrix with parameters $\xi = 0.5, \mu = 2$. The solid line represents ellipse $\mathcal{E}_\xi$, the boundary of the support of the limiting spectral distribution for an elliptic model. Notice the extra outlier near $2 + \frac{\rho}{2}$ as expected.

see Figure 2, (b) and (c) (hence the name). If $\mu > 1$ and does not belong to $\mathcal{E}_\xi$, then it has been shown in [40] that we witness an extra outlier located near $\mu + \frac{\xi}{\mu}$ (i.e. a single random eigenvalue of Γ will converge to $\mu + \frac{\xi}{\mu}$ as $N \to \infty$), see Figure 3.

We may also consider this model with an extra normalization, as in (2.4)-(ii):

$$\Gamma_{ij} = \frac{A_{ij}}{\alpha_N \sqrt{N}} + \frac{\mu}{N} \,. \tag{2.7}$$

Among the key issues we will address is the study of the various regimes of the dynamical system (2.1) with random interactions: existence of an equilibrium, study of its stability, etc.

Remark 2.2. *Using random interaction models means that we hope to predict essential aspects of an ecosystem only on the basis of statistical features of the network, rather than its detailed structure. In May's framework (see Sec. 1(b)), the main parameter turned out to be the complexity NCV, combining the number N of species, the connectance C and the variance V of the interactions. The system's equilibrium is predicted to go from stable to unstable as its complexity parameter NCV crosses a threshold.*

Our LV models are parametrized by the variance $\frac{1}{\alpha^2}$ of the interactions, their bias μ and their correlations ξ etc. One of our goal in the sequel will be to investigate how these parameters can be combined to characterize the behavior of the corresponding dynamical system.

3. Equilibrium, coexistence and stability

Due to the form of the dynamical system (2.1), standard ODE results yield that if $\boldsymbol{x}(t=0) = \boldsymbol{x}_0 > 0$ (resp. $\boldsymbol{x}(t=0) = \boldsymbol{x}_0 \geq 0$) componentwise, then $\boldsymbol{x}(t) > 0$ (resp. $\boldsymbol{x}(t) \geq 0$) for all $t > 0$. We are thus interested in equilibrium points $\boldsymbol{x}^* = (x_i^*)_{i \in [N]}$ satisfying

$$x_i^* \phi_i^{LV}(\boldsymbol{x}^*) = x_i^*(r_i - x_i^* + (\Gamma \boldsymbol{x}^*)_i) = 0 \,, \quad i \in [N] \quad \text{and} \quad x_i^* \geq 0 \,.$$

Let us present the main properties of the equilibria that have been investigated in the literature.

Outline of the section

In Section 3(a), Paragraph (iii), we list the main questions we want to address. In Section 3(c), we provide general criteria to assess the existence of a unique stable equilibrium. In Section 3(d), we address the question of feasibility. In Section 3(e), we provide heuristics on the number of surviving species in the case where the equilibrium is not feasible. In Section 3(f), we present in Table 1 an overview of the main results and remaining open questions. The phase diagram in Fig. 9 may also help the reader to navigate among the various results.

(a) Terminology and main questions

Consider the general system (1.1). If needed, we will use ϕ_i^{LV} instead of ϕ_i.

(i) Various notions of stability

The most common notion of stability in dynamical systems is the so-called ***Lyapounov stability***: The equilibrium $\boldsymbol{x}^*$ is Lyapounov stable if for any neighborhood U of $\boldsymbol{x}^*$, there exists a neighborhood W of $\boldsymbol{x}^*$ such that

$$\boldsymbol{x}(0) \in W \quad \Longrightarrow \quad \boldsymbol{x}(t) \in U \qquad \text{for all } t \geq 0\,.$$

From now on, we will simply refer to ***stability*** instead of Lyapounov stability.

The equilibrium $\boldsymbol{x}^*$ is said to be ***asymptotically stable***, if and only if it is stable and the neighborhood W can be chosen so that

$$\boldsymbol{x}(0) \in W \quad \Longrightarrow \quad \boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*\,.$$

One can also get interested in ***global stability***, in the sense that $\boldsymbol{x}^*$ is stable and

$$\forall\, \boldsymbol{x}(0) \in (0,\infty)^N\,, \qquad \boldsymbol{x}(t) \xrightarrow[t\to\infty]{} \boldsymbol{x}^*\,.$$

The following theorem provides a necessary condition for stability to hold (notice that the result below holds for general functions ϕ_i).

Theorem 3.1 (see Takeuchi [41], Theorem 3.2.5)**.** *Consider the system*

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i\phi_i(x_1,\cdots,x_N), \quad i \in [N]\,, \tag{3.1}$$

with all ϕ_i continuous. If an equilibrium point $\boldsymbol{x}^* \geq 0$ of (3.1) is stable, then $\phi_i(x_1^*,\cdots,x_N^*) \leq 0$, $i \in [N]$.

We will explain below how to relate Γ's properties and the stability of the equilibrium.

(ii) Invasion, extinction, feasability and uninvadability

An important feature of Lotka-Volterra dynamics (2.1) is the fact that species can be extinct: $x_i = 0$ entails $\mathrm{d}x_i/\mathrm{d}t = 0$ and is always a possible equilibrium value for species i. On the other hand, if at some time $x_i(t) > 0$, then positivity is maintained at all later times [42]. However, the solution can tend to zero asymptotically (see also Section 5 for finite populations).

The fact that, for all species i, $x_i = 0$ is always associated to some possible equilibria has motivated the notion of ***species invasion*** where one takes a system where $x_i(0) = 0$, and perturbs it at time t by setting x_i to a nonzero value. In particular, we often ask whether a species can invade *from rare*, i.e. what is the asymptotic behavior when $x_i(t)$ is set to a small positive value? This can be answered by considering the net growth rate ϕ_i^{LV}, defined for species i in Eq. (1.2) and representing its per capita rate of abundance change. The constant

$$\phi_i^{LV}\,|_{x_i=0} = \phi_i^{LV}(x_1,\cdots,x_{i-1},0,x_{i+1},\cdots,x_N) = r_i - \sum_{j\neq i}\Gamma_{ij}x_j \tag{3.2}$$

defines the rate of exponential growth or decay of a small population $x_i \approx 0$ in an environment where the other species start with sizes given by $\boldsymbol{x}$. This quantity is called the ***invasion growth rate*** in the community ecology literature, or invasion fitness (or just fitness) in evolution. The net growth rate ϕ_i^{LV} also has a probabilistic individual-based interpretation since it is related to the survival probability of a birth and death process that approximates locally $x_i(t)$ when x_i is small and considering that the other sizes also do not vary much (see [4, Section 4(c)]). The species is said to be deterministically (or asymptotically) extinct if $x_i \to 0$, and permanent otherwise [43–45].

Ecologists have investigated *invasion sequences*, e.g. whether different dynamics and attractors are reached depending on the order in which species are introduced [46]. The same notions

arise in an evolutionary context, where mutant types can be modelled as initially absent species. Champagnat et al. [47] construct the *polymorphic evolution sequence* that alternates phases where the dynamics is described by the LV system (2.1) and phases of invasion of new arriving species, generalizing the trait substitution sequence process introduced in [45]. Merging coevolution and invasion sequences in simulation studies has also been a long-standing endeavour among ecologists [48–52].

Beyond deterministic extinctions and invasions, we explain in Sec. 5 and [4, Section 3] how the expression for fitness arises in a probabilistic individual-based description and connects to survival probability in such a context.

The notion of ***feasibility*** will refer to an equilibrium $\boldsymbol{x}^* = (x_i^*)$ of (1.1) where all species coexist:

$$x_i^* > 0 \quad \text{for} \quad i \in [N]\,.$$

Such an equilibrium will be called ***feasible***.

The notion of ***uninvadability*** will refer to an equilibrium $\boldsymbol{x}^* = (x_i^*)$ of (1.1) satisfying:

$$x_i^* \phi_i(\boldsymbol{x}^*) = 0 \quad \text{for} \quad i \in [N]$$

where either

(a) $\phi_i(\boldsymbol{x}^*) = 0$ and $x_i^* > 0$, in which case the species x_i^* is said to ***survive*** or
(b) $x_i^* = 0$ and $\phi_i(\boldsymbol{x}^*) \leq 0$, in which case the species is said to ***vanish*** (or to be extinct).

Such an equilibrium will be called ***uninvadable***. Notice that it is stated in Theorem 3.1 that an equilibrium, to be stable, must be uninvadable.

(iii) Main questions we want to address

This raises a number of fundamental questions, in particular:

(i) Given parameter $\boldsymbol{r}$ and a random model for matrix Γ, is there a unique equilibrium? Is it locally stable? globally stable ? What is the proportion of surviving species and what are their statistical features?
(ii) Given parameter $\boldsymbol{r}$ and a random model for matrix Γ, what are the conditions to get a feasible equilibrium? Is it unique? stable?
(iii) Given an interaction matrix Γ, can we characterize the domain of growth rates $\boldsymbol{r}$ that allow feasibility?

We focus hereafter on questions (i) and (ii). We will not develop answers to question (iii), which assumes that $\boldsymbol{r}$ is likely to vary with environmental conditions while Γ is biologically fixed, since the main object of this study is LV systems when N is large and the interaction matrix Γ contains a random component. However, extensive works on question 3 can be found elsewhere, in the context of structural stability of LV equilibria [53,54].

There are several levels of answers to these questions, depending of the expected level of rigor. As we will see, we can provide fairly precise mathematical answers for a narrow range of assumptions (tight assumptions on the interactions Γ). Theoretical physics tools and computer experiments will substantially relax these assumptions and widen our understanding.

(b) Linear Complementarity Problem: an important concept to study equilibrium dynamics

In this section, we provide the definition of the Linear Complementarity Problem (LCP), which is part of the theory of mathematical programming (see [55,56] for standard references). LCP has already been used in ecological contexts in [41, Chapter 3].

Given a $N \times N$ matrix M and a $N \times 1$ vector q, we say that the $LCP(M,q)$ admits a solution $(z,w) \in \mathbb{R}^N \times \mathbb{R}^N$ if there exist two such vectors satisfying the following set of constraints:

$$\begin{cases} w & = & Mz + q \geq 0\,, \\ z & \geq & 0\,, \\ w^T z & = & 0\,. \end{cases}$$

In this case, we simply write $z \in LCP(M,q)$ since $w = Mz + q$ can be inferred from z.

Consider the LV dynamics (2.1). An uninvadable equilibrium $\boldsymbol{x}^* = (x_i^*)$ (if it exists) will satisfy

$$\begin{cases} x_i^* \phi_i^{LV}(\boldsymbol{x}^*) & = & 0\,, \\ x_i^* & \geq & 0 \\ \phi_i^{LV}(\boldsymbol{x}^*) & \leq & 0 \end{cases}, \qquad \text{for all } i \in [N]\,. \tag{3.3}$$

With the explicit form of $\phi_i^{LV}(\boldsymbol{x}^*)$, see (1.2), this exactly means that $\boldsymbol{x}^* \in LCP(I - \Gamma, -\boldsymbol{r})$.

(c) Criteria for existence and uniqueness of a globally stable equilibrium.

As mentioned above, an equilibrium point $\boldsymbol{x}^* = (x_i^*)$, if it exists, should satisfy

$$x_i^*(r_i - x_i^* + (\Gamma \boldsymbol{x}^*)_i) = 0, \quad i \in [N]\,. \tag{3.4}$$

Hence, either $x_i^* = 0$, and the species i vanishes at equilibrium or $r_i - x_i^* + (\Gamma \boldsymbol{x}^*)_i = 0$. We a priori do not know beforehand which species vanish and which ones remain. Moreover, uniqueness of the equilibrium may not be guaranteed.

A systematic way to find all the solutions is to arbitrarily fix the vanishing species $x_i^* = 0$ for $i \in A$, A being any subset of $[N]$ (2^N possibilities), then to solve the remaining set of equations

$$r_j - x_j^* + (\Gamma \boldsymbol{x}^*)_j = 0, \quad j \in A^c := [N] \setminus A\,. \tag{3.5}$$

If the obtained x_j^* are positive, then $(0, i \in A; x_j^*, j \in A^c)$ is a possible solution.

Adding an uninvadability condition may considerably reduce the number of solutions, which can then be analyzed in the LCP framework.

The following result due to Takeuchi and Adachi [41] provides a sufficient condition for the existence of a unique equilibrium and the global stability of the LV system. It is the cornerstone to establish single equilibrium/stability conditions for interaction matrices with random entries and is based on the explicit construction of a Lyapunov function.

Theorem 3.2 (Takeuchi and Adachi 1980, see [41] Th. 3.2.1)**.** *Consider the system* (2.1)*:*

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i\,(r_i - x_i + (\Gamma \boldsymbol{x})_i) = x_i\,(r_i + [(-I + \Gamma)\boldsymbol{x}]_i)$$

and assume that there exists a diagonal matrix Υ with positive diagonal elements such that matrix $\Upsilon(I - \Gamma) + (I - \Gamma)^* \Upsilon$ is positive definite. Then there exists a unique equilibrium point $\boldsymbol{x}^*$ solution of the $LCP(I - \Gamma, -\boldsymbol{r})$ and this equilibrium is globally stable in the sense that it is stable and

$$\forall\, \boldsymbol{x}(0) \in (0,\infty)^N, \quad \boldsymbol{x}(t) \xrightarrow[t \to \infty]{} \boldsymbol{x}^*\,.$$

In Takeuchi's language, $-I + \Gamma$ is said to be *Volterra-Lyapounov stable* (VL-stable). As explained in [41], it is in general difficult to determine whether a given matrix is VL-stable. But one can compare it to other better known set of matrices : any negative diagonal dominant matrix, or negative semidefinite matrix is VL-stable and conversely, if A is VL-stable then A has all its principal minors positive. It is possible to be more precise when all off-diagonal elements have the same sign but not in general.

Combining Theorem 3.2 and standard results in RMT on the limit of the largest eigenvalue of a Wigner matrix, one can determine a set of parameters for which our random interaction models are VL-stable. We prove the following result:

Theorem 3.3 (Unique equilibrium and stability under uninvadability condition)**.** *Let $\alpha_N = \alpha$ be fixed. If one of the following conditions is satisfied:*

(i) Matrix Γ is given by model (2.4)-(ii) and $\alpha > \sqrt{2}$,
(ii) Matrix Γ is given by model (2.7) and the parameters $(\alpha, \mu, \xi) \in \mathbb{R}^+ \times \mathbb{R} \times [-1, 1]$ satisfy

$$\alpha > \sqrt{2(1+\xi)} \quad \text{and} \quad \mu < \frac{1}{2} + \frac{1}{2}\sqrt{1 - \frac{2(1+\xi)}{\alpha^2}}\,.$$

Then almost surely, there exists N large enough such that the system (3.3) admits a unique solution $\boldsymbol{x}^* = (x_i^*)$, $x_i^* \geq 0$. Moreover, this equilibrium is globally stable (see Section 3(a)-(i) for the definition).

We provide the proof under Assumption (i). For the general proof, see [57, Prop. 2.6].

Elements of proof. Assume (i) and let Υ in Th. 3.2 be given by $\Upsilon = I$. Compute

$$I - \Gamma + I - \Gamma^* = 2I - (\Gamma + \Gamma^*) = 2I - \frac{\sqrt{2}}{\alpha}\left(\frac{A_{ij} + A_{ji}}{\sqrt{2N}}\right). \tag{3.6}$$

Notice that $W = \left(\frac{A_{ij}+A_{ji}}{\sqrt{2N}}\right)$ is a symmetric matrix with independent centered entries on and above the diagonal with variance $\mathrm{var}\left(\frac{A_{ij}+A_{ji}}{\sqrt{2}}\right) = 1 + \delta_{ij}$, where δ_{ij} is the Kronecker symbol with value 1 if $i = j$ and zero else. In RMT, W is referred to as a Wigner matrix, and its properties are well-studied. In particular, it is known that its largest eigenvalue $\lambda_{\max}(W)$ behaves as follows:

$$\lambda_{\max}(W) \xrightarrow[N\to\infty]{a.s.} 2\,,$$

see for instance [36, Theorem 5.1]. Going back to (3.6), we get:

$$I - \Gamma + I - \Gamma^* = 2I - \frac{\sqrt{2}}{\alpha}W$$

and the smallest eigenvalue of this matrix is

$$\lambda_{\min}\left(2I - \frac{\sqrt{2}}{\alpha}W\right) = 2 - \frac{\sqrt{2}}{\alpha}\lambda_{\max}(W) \xrightarrow[N\to\infty]{a.s.} 2\left(1 - \frac{\sqrt{2}}{\alpha}\right).$$

Taking $\alpha > \sqrt{2}$ yields the desired result.

□

It is worth noticing that the (i) and (ii) respectively are sufficient conditions to get a unique globally stable equilibrium but simulations and heuristic arguments (see e.g. [58]) indicate that these conditions are not tight and one could observe a unique equilibrium for smaller α's.

Beyond Theorem 3.2, there exists other stability criteria of a similar flavour that may be useful. One can cite for example Goh [59] that provides stability results in the case of the existence of a feasible equilibrium, or Champagnat et al. [60] that provide a variant to Theorem 3.2. Those results could have similar RMT interpretations as Theorem 3.3.

(d) Unique feasible equilibrium.

Once the conditions are met so that the LV system of coupled equations admits a unique equilibrium, a natural question arises: is this equilibrium feasible, in the sense that $\boldsymbol{x}^* > 0$ componentwise? A negative answer has been brought by Dougoud et al. [61] in the case where α is fixed. As a first conclusion, getting feasibility requires $\alpha = \alpha_N \nearrow \infty$. This implies a qualitative change in the nature of the interactions since the random part of the interaction matrix would

have a (macroscopic) vanishing effect:

$$\rho(\Gamma) = \mathcal{O}\left(\frac{1}{\alpha_N}\right) \to 0 \quad \text{(i.i.d. model)}\,, \quad \rho(\Gamma) = |\,\mu\,| + \mathcal{O}\left(\frac{1}{\alpha_N}\right) \quad \text{(elliptical model)}\,,$$

where $\mathcal{O}$ stands for the standard big O notation. In Bizeul and Najim [62], the feasibility threshold $\alpha_N \sim \sqrt{2\log(N)}$ is established. We start with an argument of Dougoud et al. [61].

No feasibility if α is fixed.

Assume that the equilibrium point $\boldsymbol{x}^*$ is feasible, then $x_i^* > 0$ for $i \in [N]$ and the equations (3.4) are equivalent to the linear system

$$(I - \Gamma)\boldsymbol{x}^* = \boldsymbol{r}.$$

Let Γ be given by the i.i.d. model (2.4). It is well-known in RMT that the spectral radius $\rho(A/\sqrt{N})$ almost surely converges to 1 (see e.g. [63]). As a consequence, for every $\alpha > 1$, $\rho(\Gamma) < 1$ eventually (i.e. almost surely for large N) and matrix $(I - \Gamma)$ is almost surely invertible for large N. Hence, the following algebraic representation of the equilibrium:

$$\boldsymbol{x}^* = \left(I - \frac{A}{\alpha\sqrt{N}}\right)^{-1} \boldsymbol{r}\,. \tag{3.7}$$

In the simpler case where $\boldsymbol{r} = \mathbf{1}_N$, the $N \times 1$ vector of ones, Geman and Hwang [64] have proved that asymptotically, for every finite M,

$$(x_1^*, \cdots, x_M^*) \xrightarrow[N\to\infty]{\mathcal{L}} \mathcal{N}_M(\mathbf{1}_M, \sigma_\alpha^2 I_M)\,,$$

where $\sigma_\alpha^2 = \frac{1}{4\alpha^2 - 1}$ depends on α, $\mathcal{N}_M(\boldsymbol{a}, C)$ $\mathcal{N}_M(\mathbf{a}, C)$ is the multivariate normal distribution with mean $\mathbf{a}$ and covariance matrix C, and $\xrightarrow{\mathcal{L}}$ stands for the convergence in distribution. As a consequence, Dougoud et al. [61] argued that under this interaction regime (fixed α) observing a feasible equilibrium was unlikely. In fact, the theoretical result by Geman and Hwang [64] asserts that each component x_i^* of the equilibrium asymptotically behaves as an independent Gaussian random variable centered at 1, with a variance independent from N, hence the heuristics

$$\mathbb{P}(\min_{i\in[N]} x_i^* > 0) \simeq \prod_{i\in[N]} \mathbb{P}(x_i^* > 0) \xrightarrow[N\to\infty]{} 0\,.$$

Otherwise stated, the initial assumption that $\boldsymbol{x}^* > 0$ is very unlikely to happen and is asymptotically a large deviation. This a priori analysis motivates the study of a feasible equilibrium under the regime $\alpha = \alpha_N \xrightarrow[N\to\infty]{} \infty$.

Feasibility when α_N grows to infinity.

In the case where $\boldsymbol{r} = \mathbf{1}$, there is a sharp phase transition around the threshold value $\alpha_N^* \sim \sqrt{2\log(N)}$ for both models (2.4)-(ii) and (2.7). Below the threshold, there is no feasibility with very high probability while above the threshold, feasibility occurs with probability growing to 1.

Theorem 3.4. *Let Γ be given by model (2.4)-(ii) or (2.7); in the latter case assume that $\mu < 1$. Assume $\alpha_N \xrightarrow[N\to\infty]{} \infty$ and denote by $\alpha_N^* = \sqrt{2\log(N)}$. Let $\boldsymbol{r} = \mathbf{1}$. Then $\boldsymbol{x}^*$ given by (3.7) is well-defined and*

(i) If $\exists\,\varepsilon > 0$ such that $\alpha_N \le (1-\varepsilon)\alpha_N^*$ then $\mathbb{P}(\min_{i\in[N]} x_i^* > 0) \xrightarrow[N\to\infty]{} 0$,
(ii) If $\exists\,\varepsilon > 0$ such that $\alpha_N \ge (1+\varepsilon)\alpha_N^*$ then $\mathbb{P}(\min_{i\in[N]} x_i^* > 0) \xrightarrow[N\to\infty]{} 1$.

Although the full proofs of the theorem are involved, simple heuristics capture the phase transition in the i.i.d. case and is presented hereafter.

Remarks

- These results are established in [62] and [57] for the i.i.d. case and the elliptic case.
- In the case where $\boldsymbol{r} \neq \mathbf{1}$ is still positive componentwise, there is not a sharp threshold at $\alpha_N^* = \sqrt{2\log(N)}$ but rather a *transition buffer* $[\alpha^*_{\min,N}, \alpha^*_{\max,N}]$ from non-feasibility ($\alpha_N < \alpha^*_{\min,N}$) to feasibility ($\alpha_N > \alpha^*_{\max,N}$). Details can be found in [62, Section 4.2].

Heuristics to understand the phase transition in Theorem 3.4. Assume that Γ is given by model (2.4)-(ii). In the representation (3.7) of the equilibrium, expand the inverse matrix as a Neumann series and only consider the first order expansion:

$$\boldsymbol{x}^* = \left(I - \frac{A}{\alpha_N\sqrt{N}}\right)^{-1}\mathbf{1} = \mathbf{1} + \frac{A}{\alpha_N\sqrt{N}}\mathbf{1} + \cdots$$

Every component x_k^* of $\boldsymbol{x}^*$ writes

$$x_k^* = 1 + \frac{Z_k}{\alpha_N} + \cdots \quad \text{where} \quad Z_k = \frac{\sum_{j\in[N]} A_{kj}}{\sqrt{N}}\,.$$

Notice that the Z_k's are i.i.d. $\mathcal{N}(0,1)$. Pushing the approximation and taking the minimum yields

$$\min_{k\in[N]} x_k^* \simeq 1 + \frac{\min_{k\in[N]} Z_k}{\alpha_N}\,.$$

Now standard results from extreme value theory yield $\min_{k\in[N]} Z_k \sim -\sqrt{2\log(N)}$, hence

$$\min_{k\in[N]} x_k^* \simeq 1 - \frac{\sqrt{2\log(N)}}{\alpha_N}\,.$$

The relative position of α_N with respect to $\alpha_N^* = \sqrt{2\log(N)}$ yields the desired result. □

(e) Unique equilibrium with vanishing species

In the case of a unique equilibrium with species vanishing when $t \to +\infty$, it is interesting to understand some properties of the survivors such as the individual distribution of the abundance of a given species, the number of vanishing species, etc. Various techniques (such as the replica method from theoretical physics) yield quantitative heuristics validated by simulations. A full mathematical analysis remains currently out of reach.

(i) Number of vanishing species.

We mentioned earlier that should the parameter α (related to the strength of the interaction) be constant or less than $\sqrt{2\log(N)}$, the equilibrium $\boldsymbol{x}^*$ will feature vanishing components $x_i^* = 0$ representing disappearing species.

In this section, we address the question of estimating the proportion of surviving species $p = p(\alpha)$ as a function of parameter α. In [58], Bunin provides heuristics based on the cavity method to address this question, while in [65], Galla establishes equations comparable to those of Bunin based on dynamical generating functionals techniques. Both heuristics apply for the elliptical model (2.7). For the i.i.d. model, a simple order statistics argument can be found in [66,67]. These a priori different methods yield the same equations from which one can extract $p(\alpha)$.

Given the random equilibrium $\boldsymbol{x}^*$, we introduce the following quantities:

$$\mathcal{S} = \{i \in [N], x_i^* > 0\}\,, \quad \hat{p} = \frac{|\mathcal{S}|}{N}\,, \quad \hat{m}^2 = \frac{1}{|\mathcal{S}|}\sum_{i\in[N]} (x_i^*)^2\,.$$

Denote by $Z \sim \mathcal{N}(0,1)$ and by Φ the cumulative Gaussian distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}}\,du\,.$$

Conjecture 3.5 (Bunin [58], Galla [65], Clenet et al. [66,67]). *Let $\alpha \in \left(\sqrt{2}, \sqrt{2\log(N)}\right)$ and assume that Γ follows Model (2.4)-(ii). The following system of two equations and two unknowns (p, m)*

$$m\sqrt{p}\,\Phi^{-1}(1-p) + \alpha = 0\,, \tag{3.8}$$

$$1 + \frac{2m\sqrt{p}}{\alpha}\mathbb{E}\left(Z \mid Z > -\frac{\alpha}{m\sqrt{p}}\right) + \frac{m^2 p}{\alpha^2}\mathbb{E}\left(Z^2 \mid Z > -\frac{\alpha}{m\sqrt{p}}\right) = m^2 \tag{3.9}$$

admits a unique solution (p^*, m^*) and the following convergence holds

$$\hat{p} \xrightarrow[n\to\infty]{a.s.} p^* \qquad \text{and} \qquad \hat{m} \xrightarrow[n\to\infty]{a.s.} m^*\,.$$

Remark 3.6. *Notice that the condition $\alpha > \sqrt{2}$ guarantees by Theorem 3.3 that a.s. eventually there exists a unique equilibrium. This condition is sufficient but might not be necessary. In the simulations hereafter, we also test the case where $\alpha \in (1, \sqrt{2}]$ and observe that with high probability, there exists a unique equilibrium and a good matching with equations (3.8)-(3.9).*

Simulations. We describe the setup of simulations to be compared with the theoretical values announced in Conjecture 3.5. Similar simulations can be found in [58].

We fix $N = 1000$ and draw L independent realizations of matrices $A^{(i)}$. We then compute the corresponding equilibria $\boldsymbol{x}^{*(i)}(\alpha)$ and their related quantities $(\hat{p}^{(i)}(\alpha), \hat{m}^{(i)}(\alpha))$ for a given $\alpha > 0$. We finally compare the empirical Monte Carlo averages:

$$\hat{p}_L(\alpha) = \frac{1}{L}\sum_{i=1}^{L}\hat{p}^{(i)}(\alpha) \quad \text{and} \quad \hat{m}_L(\alpha) = \frac{1}{L}\sum_{i=1}^{L}\hat{m}^{(i)}(\alpha)$$

to their theoretical counterparts $p^*(\alpha), m^*(\alpha)$, solutions of (3.8) and (3.9). For $\alpha \in (\sqrt{2}, \sqrt{2\log(1000)})$, we consider $L = 500$ repeated samplings and for $\alpha \in (1, \sqrt{2}]$, $L = 100$. As shown in Figure 4, the matching is remarkable, even for α below $\sqrt{2}$.

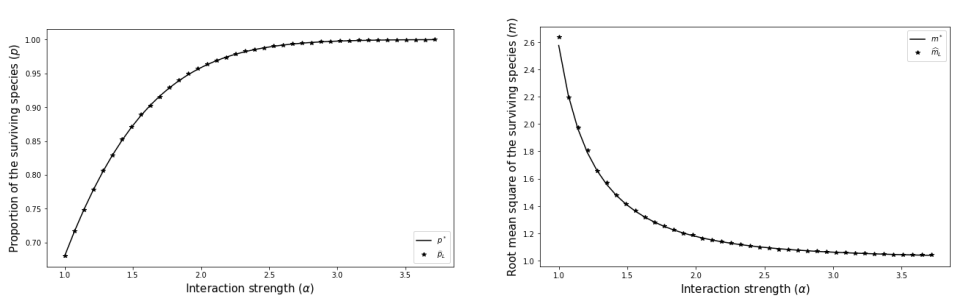
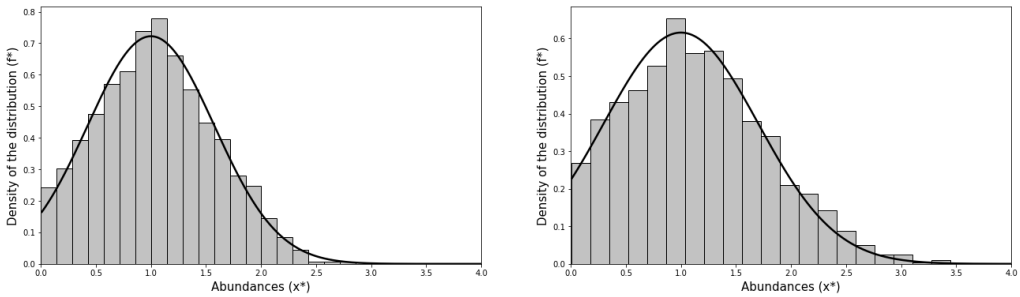


Figure 4: The plot represents a comparison between the theoretical proportion of surviving species $p^*(\alpha)$ (left) and second moment $m^*(\alpha)$ (right) computed as solutions of (3.8)-(3.9), and their empirical Monte Carlo counterpart $(\hat{p}_L(\alpha), \hat{m}_L(\alpha))$. The parameter α on the x-axis ranges from 1 to $\sqrt{2\log(N)} \simeq 3.71$. The value $\alpha = \sqrt{2}$ corresponds to the theoretical lower bound provided by Theorem 3.3, which guarantees a stable equilibrium; the value $\alpha = \sqrt{2\log(N)}$ is the upper-limit above which we have no extinction ($p^* = 1$). Notice that for $\alpha \in (1, \sqrt{2}]$, the simulations show a remarkable matching with the heuristics despite no theoretical guarantee.



(a) Gaussian entries. (b) Uniform entries.

Figure 5: Distribution of surviving species. The solid line represents the theoretical distribution f^* as given by Conjecture 3.7. The histogram is built by solving the LCP problem with an interaction matrix of size $N = 2000$ and parameter set to $\alpha = \sqrt{3}$. In plot (A), the entries are Gaussian $\mathcal{N}(0,1)$ and the interaction strength is fixed to $\alpha = 2$. In plot (B), the entries are uniform $\mathcal{U}([-\sqrt{3}, \sqrt{3}])$ which implies that the entries are centered with variance one.

(ii) Single species distribution.

The previous conjecture provides an estimation of the proportion of surviving species $p^*(\alpha)$. We go here one step further and describe the distribution of a given abundance x_i^* where index i corresponds to a surviving species.

Conjecture 3.7. *Let $\alpha \in \left(\sqrt{2}, \sqrt{2\log(N)}\right)$ and let $i \in \mathcal{S}$, i.e. i corresponds to a surviving species. Let p^*, m^* be the solutions of (3.8) and (3.9) and $Z \sim \mathcal{N}(0,1)$. Then the distribution of x_i^* is a truncated Gaussian:*

$$\mathcal{L}(x_i^*) = \mathcal{L}\left(1 + \frac{m^*\sqrt{p^*}}{\alpha} Z \,\middle|\, Z > -\frac{\alpha}{m^*\sqrt{p^*}}\right) .$$

Otherwise stated, x_i^* admits the following density:

$$f^*(v) = \frac{1_{(v>0)}}{\Phi(-\delta)} \frac{\delta}{\sqrt{2\pi}} \exp\left(-\frac{\delta^2(v-1)^2}{2}\right) \qquad \text{where} \qquad \delta = \frac{\alpha}{m^*\sqrt{p^*}}$$

and Φ stands for the cumulative Gaussian distribution.

The matching between the theoretical density f^* given in Heuristics 3.7 and a histogram of a given equilibrium $\boldsymbol{x}^*$ is illustrated in Figure 5. In particular, the theoretical distribution matches, even with non-Gaussian entries (see Fig. 5b).

(iii) Interactions between survivors

When only a fraction of species survive in the unique LCP equilibrium, one can also ask how the interactions restricted to the survivors are modified. Mathematically, this boils down to consider the submatrix $(\Gamma_{ij})_{i,j\in\mathcal{S}}$. Of course, the lines and columns that are selected depend on the initial realisation of the matrix Γ and it is not an easy task to predict the new statistical features of the entries. Nevertheless, heuristics for these quantities have been given in [68], using the cavity method. These authors have obtained general formulas for the model (2.5) that can be found in [68], but for the sake of simplicity, we present it here in the case $\xi = 0$.

Conjecture 3.8. *Let Γ be a non-centered Ginibre matrix, that is obeying model (2.5) with $\xi = 0$. Assume that $\alpha \in \left(\sqrt{2}, \sqrt{2\log(N)}\right)$ and, for any $i \in [N]$, $r_i = 1$. Then, for any $i, j \in \mathcal{S}$, i.e. i, j correspond to*

surviving species, the entry Γ_{ij} is still Gaussian but with the following bias and correlation:

$$\mathbb{E}[\Gamma_{ij} \mid \boldsymbol{x}^*] - \frac{\mu}{N} = -\frac{x_i^* - 1 + \mu\overline{\boldsymbol{x}^*}}{N \times \overline{(\boldsymbol{x}^*)^2}} x_j^* \qquad \textit{and} \qquad \mathrm{Corr}\big[\Gamma_{ij}, \Gamma_{ik} \mid \boldsymbol{x}^*\big] = -\frac{x_j^* x_k^*}{N \times \overline{(\boldsymbol{x}^*)^2}}, \tag{3.10}$$

where, for any vector $\boldsymbol{x} = (x_i,\ i \in [N])$, $\boldsymbol{x}^2 = (x_i^2, i \in [N])$ and $\overline{\boldsymbol{x}} := \frac{1}{N}\sum_{i=1}^N x_i$.

In [69], a slightly different *maximum likelihood* viewpoint is adopted, where the elliptic model (2.5) is considered as a prior distribution and the corresponding posterior distribution is computed knowing the equilibrium, formulated as a linear constraint as in (3.5). Formulas (3.10) are recovered in this context. It is possible to study how these correlations impact spectral properties of the restricted matrix [69,70]. Rigorous proofs remain out of reach.

(f) An overview of the results and some open questions

In the next table, we summarize some of the results presented here and list a few open questions.

α	α fixed		$\alpha = \alpha_N \nearrow \infty$	
value	$\sqrt{2}$		$\sqrt{2\log(n)}$	
Equilibrium	OQ?	`unique`		
Feasibility	OQ?	`no`		`yes`
Single species distribution	OQ?	*truncated gaussian*		`vanishing gaussian`
proportion of vanishing species	OQ?	*provided by accurate heuristics*		`all species are present`

Table 1: The table above summarizes the different regimes (depending on α) of the LV system (2.1) where $r_i = 1$ and Γ follows i.i.d. model (2.4)-(ii) with standard Gaussian entries. By "OQ?", we mean open question; in `typewriter font`, the result is proven mathematically; in *italics*, the result is established at a physical level of rigor.

4. Structured models

In the previous section, we have studied at large the LV system (2.1) where the matrix Γ, supposed random, has either i.i.d. entries or follows the elliptic model, see Section 2(b). Another line of research focused on a model of the Jacobian of the system near equilibrium:

$$J = -I + M \tag{4.1}$$

where M is random, the question being then to understand the relative localisation of the spectrum of M with respect to -1 to conclude on the stability of the system. This approach is historically the first with May's paper [7], a presentation of which is provided in [4, Section 1].

To progress toward a more realistic description of the reality, one is tempted to consider more involved models of random matrices to take into account more properties of the complex systems such as sparsity, existence of underlying structures, randomness beyond independence, etc.

For instance, the question of the effect of the structure of the ecological network on its feasibility and stability already appeared in [71], where Pimm argued that connectance is not

the only parameter that can influence the feasibility and stability of the networks, and started a theoretical study of structured (both deterministic and random) networks.

In this section, we present a variety of random matrix models beyond the i.i.d. and elliptic ones, highlighting their use in theoretical ecology and listing mathematical results and questions of interest. Often, mathematical results are not directly available and massive simulations remain the main approach to exploit the potentialities of such models. We start this collection of structured models with a quick reminder of what ecologists understand by "sparsity", since most of the model structures revolve around the idea of structuring where the zeros are in matrix Γ.

Sparsity in ecological networks. Empirically, in an ecosystem with N species, even if the maximal number of interactions is N^2, the real number L of nonzero interactions is often much smaller. We define the ***connectance*** as

$$C = \frac{L}{N^2} \,. \tag{4.2}$$

For a LV system, this means that matrix Γ has $N^2 - L$ null entries. Although the interpretation is less obvious for random jacobian matrices (4.1), the notion of connectance already appeared in [7,22] and is important to model sparsity. Recently, Grilli et al. [72] worked explicitly on the interaction matrix of a LV system and studied the stability and feasibility of the equilibrium as a function of various parameters, among which the connectance (see also [73,74]). Based on empirical evidence, [75] suggested that food webs can have very low connectance.

The notion of sparsity used thereafter is different from the one usually employed in mathematics, where matrices or networks are said to be *sparse* when the ratio L/N^2 goes to zero with N and *dense* when C is of the order of a positive constant for large N, the graph is said to be dense. In ecology, "sparse networks" are understood more broadly as networks in which not all interactions exist (i.e. matrices with zeros). In the following, we will consider sparse networks, in the ecological sense, with a wider range of connectance including the regime when $C = O(1)$, which is called dense in mathematics.

Beyond the connectance, it is possible to take into account the structure of the network by setting some of the entries to zero, thus enforcing ecological sparsity, but in a structured manner. For this, we may use a matrix $\Delta = (\Delta_{ij})$ where Δ_{ij} equals 1 if species j has an effect on species i and 0 otherwise. If one draws the system interactions as a graph, then Δ can be interpreted as the adjacency matrix of this graph and the interaction matrix Γ or the community matrix M can then be represented as proportional to $\Delta \circ A$, with $(\Delta \circ A)_{ij} = \Delta_{ij} A_{ij}$ and A is random either i.i.d. or elliptic. In such a model, Δ represents the structure of the system summarizing ecological sparsity and A the (random) intensity of the interactions.

(a) The simplest model for sparsity: Erdős-Rényi graphs

When all species play the same role in the foodweb and the only parameter of interest is the average number of interactions for a given species, it is natural to choose Δ as the adjacency matrix of an Erdős-Rényi (ER) graph of size N : each coefficient of the random matrix Δ has probability p to be nonzero, equal to 1, and probability $1-p$ to be put to zero, independently of the others. The average number of edges in the graph is pN^2, hence the connectance C equal to p.

ER in the mathematical literature

ER graphs are reference models in random graphs and their geometric properties are well-known (see e.g. [76–78]), together with the spectral properties of their adjacency matrices. In the regime when $C = O(1)$, which is called *dense* by mathematicians, the ER matrix is a rank one deformation

of a matrix with centered i.i.d. entries, so that we observe a circular law and one outlier. In fact,

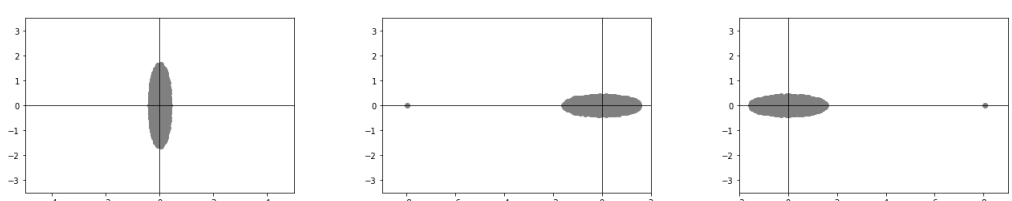
$$\frac{\Delta}{\sqrt{N}} = \frac{1}{\sqrt{N}}(\Delta - \mathbb{E}\Delta) + \frac{1}{\sqrt{N}}\mathbb{E}\Delta \quad \text{with} \quad \left\| \frac{1}{\sqrt{N}}\mathbb{E}\Delta \right\| = \sqrt{N}\, C\,,$$

where $\|\cdot\|$ refers to the spectral norm when applied to a matrix. Notice that the precise understanding of the extreme eigenvalues of $\frac{\Delta}{\sqrt{CN}}$ in sparse or very sparse regimes is still an active subject in RMT. A concise overview can be found in the introduction of [79].

ER in the ecological literature : sparsity increases stability

As developed in [4, Section 1], the case when $M_{ij} = \Delta_{ij} A_{ij}$, with Δ the adjacency matrix of a dense ER graph and A has i.i.d. centered entries with variance V has been already considered by May. This model is equivalent to the full model, where the entries have variance CV. In this case sparsity increases stability: the stability condition $NVC < 1$ is easily satisfied for small C.

The case when Δ is the adjacency matrix of an ER graph but the model for the matrix A is more involved has been studied in particular in [15]. They use models for A that are of the same flavour as the elliptic model - for example, $(A_{ij}, A_{ji})_{i<j}$ both positive to model mutualistic systems or with opposite sign to model a prey-predator situation. As illustrated in Figure 6, in the mutualistic case, outliers with a large real part may strongly affect the stability. In [15], the authors also establish an explicit stability criterion adapted to each case, generalizing May's criterion and emphasize again that sparsity increases stability.



(a) Predator-prey. (b) Competitive. (c) Mutualistic.

Figure 6: Spectrum of the interaction matrix for $N = 1000$ species. Δ is ER symmetric with $C = 0.1$. For the competitive (resp. mutualistic) model, $A_{ij} = A_{ji}$ with distribution $-|\mathcal{N}(0,1)|$ (resp. $|\mathcal{N}(0,1)|$) variables. For predator-prey $A_{ij} = -A_{ji}$, distribution $|\mathcal{N}(0,1)|$.

(b) Sparsity with a deterministic structure

Consider a deterministic d-regular oriented graph with N vertices, that is a graph where each vertex i has exactly d oriented edges exiting from i and d edges coming to i. Let Δ be the adjacency matrix of such a graph, then Δ is deterministic and has d non-null entries per row and per column and $L := d \times N$ non-null entries overall. Parameter d which may depend on N accounts for the sparsity of the system and in the framework of a LV system, consider the interaction matrix:

$$\Gamma = \frac{1}{\alpha}\frac{\Delta \circ A}{\sqrt{d}} = \frac{1}{\alpha}\left(\frac{\Delta_{ij} A_{ij}}{\sqrt{d}}\right)_{ij}\,,$$

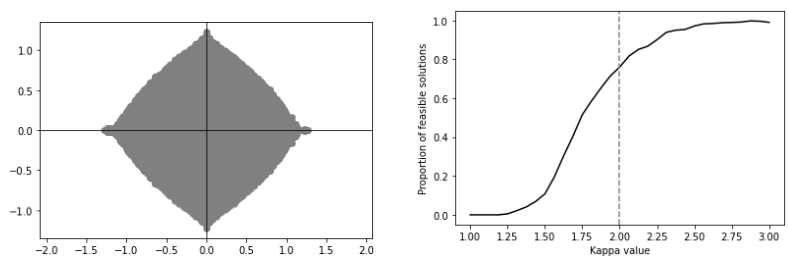
where the A_{ij}'s are i.i.d. and α is an extra normalization which may depend on N. Notice that the normalization is no longer $\sqrt{N}$ but $\sqrt{d}$ accounting for the fact that there are exactly d non-null entries per row. For such a model the connectance C equals:

$$C = \frac{d}{N}$$

and the interest lies in "small" values of d.

This model has been studied by Akjouj and Najim in [80] where specific assumptions on d and Δ are considered, namely either d is proportional to N, or $d \gg \log(N)$ and Δ has a specific block structure, cf. Model (A) in [80] and [4, Section 5(a)]. In this article, it is shown that the same phase transition as in Theorem 3.4 occurs: Feasibility and stability hold iff $\alpha = \alpha_N \gg \sqrt{2\log(N)}$.

The spectrum of matrix $\frac{\Delta \circ A}{\sqrt{d}}$ together with the proportion of equilibria near the phase transition thershold are plotted in Figure 7.



(a) (b)

Figure 7: Deterministic model with $N = 1000, d = 10$. Each species interacts with d species. Non-null interactions are i.i.d. $\mathcal{N}(0,1)$. (a) Spectrum of the $\frac{1}{\sqrt{d}}$-normalized interaction matrix. (b) Proportion of feasible equilibrium: each point represents the proportion of feasible solutions $\boldsymbol{x}$ over 200 realizations of random matrices Γ_N for different values of κ, with $\alpha_N = \sqrt{\kappa \log(N)}$.

(c) Introducing modularity through Stochastic Block Model (SBM)

Beyond the ER case, when every species equally interacts with another, it is often more realistic to consider *communities* within the ecosystem (also called ***modules***), that is groups of species sharing the same connexion patterns. This leads to the celebrated Stochastic Block Model (SBM), introduced in [81], (see also [82,83] for reviews). Let $r \in \mathbb{N}$ be the number of communities. Given

- a vector of positive real numbers $(\pi_1, \cdots, \pi_r)$ such that $\sum_{i=1}^{r} \pi_i = 1$,
- an $r \times r$ matrix P,

the corresponding SBM is a random graph whose vertices are partitioned into r communities $C_1, \cdots, C_r$, where each node belongs to the community C_i, $i \in \{1, \cdots, r\}$, with probability π_i. Then, an edge between a vertex $u \in C_i$ and a vertex $v \in C_j$ exists with probability p_{ij}, independently of all other edges.

SBM in the mathematical literature

There exists a huge literature on the SBM, initially introduced to analyze social networks, and extensively used in machine learning for modelling complex networks and address the community detection problem. The goal there is to design algorithms to cluster the different communities and estimate accurately matrix P, see for example [84,85].

Again using the Hadamard product $\Delta \circ A$, the spectrum of the adjacency matrix Δ associated to a SBM can be described, at least in simple cases. Consider for example a SBM with $r = 2$ communities of equal size ($\pi_1 = \pi_2 = 1/2$) and let

$$\begin{pmatrix} p & q \\ q & p \end{pmatrix},$$

with p and q of order $0(1)$ (dense case). Then Δ is a rank-two perturbation of a matrix with centered independent entries. Depending on the values of $\frac{p+q}{2}$ and $\frac{p-q}{2}$, there can be up to two outliers in its spectrum. As in the ER case, sparse cases has also been recently considered [86].

SBM in the ecological literature : modularity increases stability

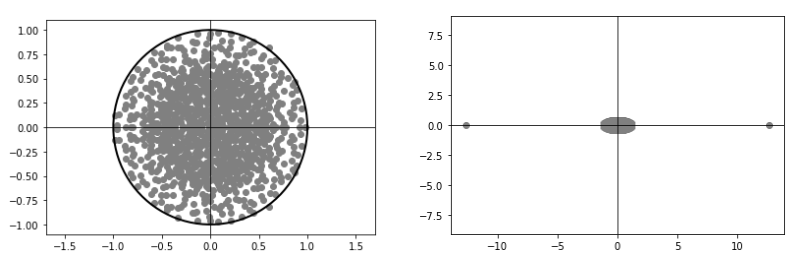
In the seventies, May and Pimm already considered rudimental forms of the SBM into the framework of the Jacobian model (4.1), to take into account some features of ecosystems such as modularity and ***compartmental models***.

In [7], May presents a simple occurrence of SBM. He considers a SBM with r modules and a probability vector $(c_1, \ldots, c_r)$. This SBM corresponds to modules with no interactions, while within the ith block made of d_i species, the interactions behave like an ER graph with connectance c_i and variance V_i. The stability condition reads :

$$\max_{i \in [r]} c_i V_i d_i < 1,$$

hence modularity increases stability. This phenomenon is illustrated in Figure (8).

In [71], Pimm addresses the following question *"should model systems be organized into compartments of species characterized by strong interactions within compartments, but weak interactions among the compartments ?"* A random version of such a model would correspond to a SBM with a matrix P having large diagonal coefficients and small off-diagonal ones.



(a) (b)

Figure 8: SBM with two communities of 500 species, $N = 1000$. The probability of interactions inside the first (resp. second) community is c_1 (resp. c_2) and the probability of interaction with species of the other community is equal to q. (a) Spectrum of the interaction matrix normalized by $\frac{1}{\sqrt{(c_1+q)N}}$, where Δ is a SBM with $c_1 = 0.5 > c_2 = 0.2 > q = 0.02$ and A has i.i.d. Gaussian $\mathcal{N}(0,1)$ entries. (b) Spectrum of the interaction matrix normalized by $\frac{1}{\sqrt{qN}}$ where Δ is a bipartite symmetric model, that is a SBM with $q = 0.5$ and $c_1 = c_2 = 0$ and A has i.i.d $|\mathcal{N}(0,1)|$ entries.

More recently, the effect of modularity on the stability of the networks is extensively explored in [87] in the framework of a tamed version of LV equations. They evaluate modularity through an index, introduced by [88] and, through simulations, illustrate that persistence, that is the number of surviving species, increases with modularity in trophic networks (see also [89]) but decreases with modularity in mutualistic networks. It would be interesting to investigate whether mathematical results on SBM could help to understand their observations.

The question of stability can be important for plant-pollinator ecosystems, corresponding to bipartite mutualistic networks (see e.g. [87,90]). In [90], the evolution of abundances can be approximated, when the number of species N tends to infinity, by a kinetic integro-differential equation where the dense graphs are replaced by ***graphons***. The theory of graphons is mathematically well developed but beyond the scope of this review (see for example [91]).

(d) Nested models : a few generalist and many specialist species.

In the Erdős-Rényi model or in SBM, the network is determined by considering the absence or presence of edges for each pair of vertices independently of the others. Other models of random graphs are defined by specifying the degree distributions. For example in the configuration models (also known as Molloy-Reed-Bollobás, see e.g. [76,78,92]), independent random variables distributed with the target degree distribution are associated to each vertex and edges are formed by pairing at random the half-edges.

By choosing heavy-tailed degree distributions, one can create a few vertices with very high degrees (corresponding to generalist species) and a majority of vertices with low degree (corresponding to specialists). Such ecosystems are called ***nested***. They have been modelled and studied, at least through simulations, see for instance [87] following [93].

Nested ecosystems can also be described through random graphs with given *expected degrees*. This model is known as the *Chung-Lu model* : take a deterministic sequence $w = (w_1, \dots, w_n)$, that will correspond to the expected degrees and draw an edge between vertex i and vertex j with probability $w_i w_j / \sum_{i=1}^n w_i$ independently of all other edges. If we choose all the weights to be equal to pn, we are obviously back to the ER model with connectance p but nested ecosystems can be modeled by choosing a power-law distribution for the weights, that is $w_i = c_i^{-\frac{1}{\beta-1}}$, for i greater or equal to some i_0. In this case, the number of species interacting with k others is proportional to $k^{-\beta}$. The spectrum of the adjacency matrix of such a graph has been studied in [94] where a phase transition is shown to occur at $\beta = 2.5$: for $\beta > 2.5$, the largest eigenvalue behaves like $\sqrt{m}$, the maximal degree in the graph, whereas for $\beta < 2.5$, the largest eigenvalue behaves like $\bar{d}$, the weighted average of the square of the expected degrees. It would be interesting to investigate whether these mathematical results could be effectively used in the study of nested ecosystems.

(e) Kernel matrices

Definition of the model. Part of the literature on ecological networks considers that the interactions between two species depends on a distance between their respective values of some functional traits. The examples that we will present below, fit into the mathematical framework of **kernel matrices**. We have

$$\Gamma_{ij} \text{ or } M_{ij} = f(g(\boldsymbol{x}_i, \boldsymbol{x}_j)), \tag{4.3}$$

where $\boldsymbol{x}_i$ is a vector modelling the traits of species i, $g : \mathbb{R}^p \times \mathbb{R}^p \longrightarrow \mathbb{R}$ is a symmetric function, denoted as the kernel (corresponding to the measure of the distance), and $f : \mathbb{R} \longrightarrow \mathbb{R}$ a function called the envelope. Examples are $g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{y}$, or $\|\boldsymbol{x} - \boldsymbol{y}\|^2$ and $f(x) = \exp(ex)$ or $(1 + x)^a$ etc.

Kernel matrices in the mathematical literature. Among these models, the first interesting and well studied case is the so-called Wishart case[3], when $g(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{y}$ and $f(x) = x$. If the entries of the vectors are i.i.d. centered and normalized, then it is well known (see [95]) that, if the ratio p/N of the number of traits over the number of species converges to τ, the spectrum converges almost surely to the Marcenko-Pastur distribution. We will develop the mathematical theory of kernel matrices in [4, Section 5(b)]. The main message is that in the RMT regime, that is when the number of traits is large and of the same order as the number of species, and if g and f are reasonable, they have almost no influence on the spectrum, meaning that, in the models, "any" kernel matrix could be without harm replaced by a Wishart matrix or a Gaussian kernel matrix.

Kernel matrices for ecological networks. Dieckmann and Doebeli consider a simple co-evolutionary model [96] (see also [6,97]) where interactions are of the form (4.3) with $f(x) = \eta_c \exp(-x^2/2\sigma_c^2)$ and $g(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|^2$, $\eta_c \in \mathbb{R}$ being the strength of the competition of mutualism and $\sigma_c > 0$ being the width of the ecological niches. The function f corresponds to a Gaussian kernel (also used in [98]). In [99], they develop a *phenotype matching model*, where

[3]That is the empirical covariance matrix of the vectors $X_1, \dots, X_N$.

the interaction is stronger when the traits of two species are close but also *phenotype difference (or threshold) model*, in which successful interaction depends on the degree to which the trait of the second species surpasses the trait of the first one[4] (see also [100,101]). Other models involving thresholds can be found in [102] and in [103], the models involving a combination of characteristics of the species taken separately and a measure of the similarity between the traits.

Let us finally mention the work of [104], which lies in the LV framework, with a kernel matrix Γ for which the distance between the traits, is determined through a distance between species in their phylogenetic tree. The author addresses the questions of feasability and stability as we have detailed it for the elliptic case and explicitly uses the link with Wishart matrices mentioned above. It would be interesting to investigate how this point of view applies to other contexts.

5. Other community models with noise

The models presented in the previous sections can be complexified. For instance, in [105,106], a dynamical system allowing to incorporate homogeneous Gaussian vector field with gradient and solenoidal contributions (resuming to May's equation when linearizing around critical points) is presented. It is shown that in this case there can be a exponential number of equilibria, the majority being unstable. Also, in [107], interactions corresponding to locally tree-like networks are modeled, with a focus on transient early phases in Equations 2.1.

Stability questions as well as out of equilibrium studies have also been investigated in other kinds of stochastic Lotka-Volterra systems, when not only the interaction matrix is random but when also noise is added to the dynamics. We review different models with noise that are obtained from individual-based stochastic models by various rescalings. The evolution is described by stochastic differential equations (SDEs), of Feller or Ornstein-Uhlenbeck types, see [108] for a review. Other possible equations are presented in [4, Section 3] with time-scales separations and [109] for discussions in view of ecological systems. For such systems of stochastic differential equations (such as the ones in Prop. 5.2 and 5.3), the notion of stability for ODEs is replaced by the notion of equilibrium distribution – the only stationary state being often zero, corresponding to the extinction of all species. More precisely, the solution $\boldsymbol{X}(t)$ of a stochastic differential equation converges in distribution to the stationary probability measure μ on $\mathbb{R}^N$ if for every continuous bounded function f, $\lim_{t\to+\infty}\mathbb{E}(f(\boldsymbol{X}(t))=\int_{\mathbb{R}^N} f(\boldsymbol{x})\mathrm{d}\mu(\boldsymbol{x})$ (see [110]). Long time behaviour of SDEs is the subject of an abundant literature in mathematics [111,112].

The LV eq. (2.1) can arise as limits of Individual Based Models (IBM), see [6], when population sizes are large. We introduce a scaling parameter K referred to as a *carrying capacity* and assume that the N species have initially sizes of order K. The individuals can give birth to new individuals of the same species or die. More precisely, the natural birth and death rates of an individual of species $i\in[N]$ are b_i^K and d_i^K such that $b_i^K-d_i^K=r_i^K$. The competition pressure (or extra death rate) exerted by an individual of species j on an individual of species i is Γ_{ij}^K, if the latter is non positive. In case Γ_{ij}^K is positive, it can be considered as an extra birth rate due to cooperation between the species i and j. Additionally an individual of species i experiences an intra-specific death rate proportional to the size of its species (each individual experiences an extra death term due to the logistic competition and equal to the size of the population i over K). The natural IBMs associated with (2.1) have two levels of stochasticity: (i) matrix Γ^K is random, (ii) the occurrence of birth/death events is random. Hereafter, we work conditionally to Γ^K.

Individual-based models have long been used for simulations in Ecology [113–116], (see [117,118] for software performing IBM simulations). By presenting the fluctuations arising in the convergence of the IBM abundances to LV abundances, we provide a link between the equations considered in this review and these algorithms (see [4, Part 3] for more mathematical details).

[4] One can think of fruit and beak sizes in a plants-birds interaction network.

Denote by $Y_i^K(t)$ the size of species i at time t and by $\boldsymbol{Y}^K(t) = (Y_1^K(t), \cdots, Y_N^K(t))$ the vector of all the species' sizes, with forall $i \in [N]$, the following convergence in probability

$$\lim_{K \to +\infty} \frac{Y_i^K(0)}{K} = x_i(0). \tag{5.1}$$

The stochasticity of the birth and death events gives rise to an additional noise process compared to (2.1). More precisely, $\boldsymbol{Y}^K(t)$ now satisfies the following SDE: for all $i \in [N]$,

$$\mathrm{d}Y_i^K(t) \quad = \quad \big(r_i^K - Y_i^K(t) - (\Gamma^K \boldsymbol{Y}^K(t))_i\big) Y_i^K(t)\, \mathrm{d}t + \mathrm{d}M_i^K(t)\,, \tag{5.2}$$

where M_i^K is a martingale process, with $\mathbb{E}(M_i^K(t)) = 0$, $\mathrm{Cov}(M_i^K(t), M_j^K(t)) = 0$ and

$$\mathrm{Var}(M_i^K(t)) \quad = \quad \frac{1}{K}\mathbb{E}\Big(\int_0^t \big(b_i^K + d_i^K + Y_i^K(s) + (\Gamma^K \boldsymbol{Y}^K(s))_i\big) Y_i^K(s)\, \mathrm{d}s\Big)\,, \tag{5.3}$$

(this is the analogue of Eq. (3.4) derived for $N = 1$ in [4, Part 3]).

We now detail two different limits that can be derived from the IBM depending on the chosen parameters and rescalings: (i) the Lotka-Volterra ODEs (2.1) with a fluctuation process of Ornstein-Uhlenbeck type when $K \to +\infty$ without time rescaling; (ii) the Feller-type diffusions when the birth-death dynamics is nearly critical and when time is also rescaled by K, that is when considering the process at times Kt, $t \in [0, T]$.

(a) Large population limit and fluctuation around the ODE (2.1)

First, we let $K \to +\infty$ without rescaling time and with a fixed number N of species. We consider the rescaled process:

$$\boldsymbol{X}^K(t) := \frac{\boldsymbol{Y}^K(t)}{K}\,.$$

Here, for all $i, j \in [N]$,

$$b_i^K = b_i\,, \quad d_i^K = d_i\,, \quad \Gamma_{ij}^K = \frac{\Gamma_{ij}}{K}\,, \tag{5.4}$$

where the quantities b_i, d_i and Γ_{ij} do not depend on K and $r_i := b_i - d_i$. Notice that the competition term Γ_{ij}^K can be understood as the extra death rate exerted by an individual of the species j on an individual of the species i. When the population is large and of order K, the interaction between the individuals of a given pair are weaker and therefore the competition term is rescaled in $1/K$.

An averaging phenomenon appears (similarly to the law of large numbers): from (5.3), we can see heuristically that the noise disappear and the evolution equations (5.2) can be approximated by Lotka-Volterra ODEs (see e.g. [119, Theorem 2.1 p.456], or for more generalizations to measure-valued processes in [6,108,120–122] for rigorous proofs). Indeed, the variance of the martingale part in X_i^K is of order $1/K$ and converges to zero when $K \to +\infty$.

Proposition 5.1. *Assume* (5.1), (5.4) *and assume that for all* $i \in [N]$,

$$\sup_{K \in \mathbb{N}} \mathbb{E}\left[\left(X_i^K(0)\right)^3\right] < +\infty, \tag{5.5}$$

then, when $K \to +\infty$, *the abundance processes* $(\boldsymbol{X}^K(t))_{t \geq 0}$ *converges uniformly on every compact time interval* $[0, T]$ *with* $T > 0$ *(as* $\mathbb{R}^N$*-valued processes) and in probability to population densities* $(\boldsymbol{x}(t))_{t \geq 0}$,

for $i \in [N]$, whose evolution is described by the system of ODEs (2.1):

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i\left(r_i - x_i + (\Gamma \boldsymbol{x})_i\right), \qquad i \in [N]. \tag{5.6}$$

Put formally, this means that

$$\forall \varepsilon, T > 0\,, \qquad \mathbb{P}\left\{\sup_{t \le T}\left\|\boldsymbol{X}^K(t) - \boldsymbol{x}(t)\right\| > \varepsilon\right\} \xrightarrow[K\to\infty]{} 0\,.$$

We now consider the fluctuation process associated with this convergence:

$$\boldsymbol{\eta}^K(t) = \sqrt{K}\left(\boldsymbol{X}^K(t) - \boldsymbol{x}(t)\right). \tag{5.7}$$

It is a $\mathbb{R}^N$-valued vector process whose ith coordinate is $\sqrt{K}(X_i^K(t) - x_i(t))$. Another reformulation is that the stochastic process can be expressed as:

$$\boldsymbol{X}^K(t) = \boldsymbol{x}(t) + \frac{\boldsymbol{\eta}^K(t)}{\sqrt{K}}\,.$$

Applying [119, Theorem 2.3, Chapter 11], we obtain that:

Proposition 5.2. *Under the assumption of Proposition 5.1 and assuming that, in distribution,*

$$\lim_{K\to+\infty} \boldsymbol{\eta}_0^K = \boldsymbol{\eta}_0 \in \mathbb{R}^N, \tag{5.8}$$

then, when $K \to +\infty$, the process $(\eta^K(t))_{t\ge 0}$ converges in distribution, and for the topology of uniform convergence on $[0, T]$ for $T > 0$, to the solution of the Ornstein-Uhlenbeck SDE:

$$\begin{aligned}\mathrm{d}\boldsymbol{\eta}(t) = &\left(\boldsymbol{r} - 2\boldsymbol{x}(t) - \Gamma\boldsymbol{x}(t)\right) \circ \boldsymbol{\eta}(t)\,\mathrm{d}t + \boldsymbol{x}(t) \circ \left(\Gamma\boldsymbol{\eta}(t)\right)\mathrm{d}t \\ &+ \mathrm{diag}\Big(\left(\boldsymbol{b} + \boldsymbol{d} + \boldsymbol{x}(t) + \Gamma\boldsymbol{x}(t)\right) \circ \boldsymbol{x}(t)\Big)\,\mathrm{d}\boldsymbol{B}(t)\,,\end{aligned} \tag{5.9}$$

with initial condition $\boldsymbol{\eta}_0$ defined in (5.8), $\boldsymbol{B}$ a standard N-dimensional Brownian motion, $\circ$ the Hadamard product and $\mathrm{diag}(\cdot)$ the diagonal matrix with diagonal entries the components of the vector in the argument. Equivalently, the componentwise definition of $\boldsymbol{\eta} = (\eta_i)$ is given for all $i \in [N]$ by:

$$\begin{aligned}\mathrm{d}\eta_i(t) = &\Big(r_i - 2x_i(t) - (\Gamma\boldsymbol{x})_i(t)\Big)\eta_i(t)\,\mathrm{d}t + x_i(t)\Big(\Gamma\boldsymbol{\eta}(t)\Big)_i\,\mathrm{d}t \\ &+ \Big(b_i + d_i + x_i(t) + (\Gamma\boldsymbol{x})_i(t)\Big)\,x_i(t)\,\mathrm{d}B_i(t)\,.\end{aligned}$$

Conditionally on matrix Γ, the solution $\boldsymbol{x}(t)$ of (2.1) is deterministic and hence the stochastic differential equation (5.9) is of Ornstein-Uhlenbeck type, i.e.

$$\mathrm{d}\boldsymbol{\eta}(t) = A(t)\boldsymbol{\eta}(t)\mathrm{d}t + \Sigma(t)\mathrm{d}\boldsymbol{B}(t)$$

where

$$\begin{aligned} A(t) &= \mathrm{diag}\Big(\boldsymbol{r} - 2\boldsymbol{x}(t) - \Gamma\boldsymbol{x}(t)\Big) + \mathrm{diag}\Big(\boldsymbol{x}(t)\Big)\,\Gamma, \\ \Sigma(t) &= \mathrm{diag}\Big(\left(\boldsymbol{b} + \boldsymbol{d} + \boldsymbol{x}(t) + \Gamma\boldsymbol{x}(t)\right) \circ \boldsymbol{x}(t)\Big)\,. \end{aligned}$$

The solution is a centered Gaussian process with covariance function:

$$\mathrm{Cov}\left(\boldsymbol{\eta}(t), \boldsymbol{\eta}(s)\right) = \int_0^{t\wedge s} e^{-\int_0^v A(u)\mathrm{d}u}\Sigma^2(v)e^{-\int_0^v A(u)\mathrm{d}u}\mathrm{d}v\,,$$

a $N \times N$ matrix-valued function.

This central limit theorem quantifies the convergence rate in Proposition 5.1 and allows for example to compute confidence intervals.

(b) Noisy versions of Lotka-Volterra equations

(i) Limiting Feller diffusion for large population with time rescaling

Another way to exhibit SDEs is to consider diffusive time rescaling in the almost-critical case (e.g. [6, Section 4.2], when the growth rates are close to zero). More precisely, if

$$b_i^K = K\sigma_i + b_i, \quad d_i^K = K\sigma_i + d_i, \quad \text{and} \quad \Gamma_{ij}^K = \frac{\Gamma_{ij}}{K}, \tag{5.10}$$

where σ_i, b_i, d_i and Γ_{ij} do not depend on K. As usual, we denote by $\boldsymbol{b} = (b_i)$, $\boldsymbol{d} = (d_i)$, $\boldsymbol{r} = (b_i - d_i)$ $N \times 1$ vectors and define the matrix Σ as the $N \times N$ diagonal matrix with entries (σ_i), i.e. $\Sigma = \mathrm{diag}(\sigma_i)$. The fact that the birth and death rates are of order K corresponds to accelerating the time proportionally to the factor K that also rescaling the population size. That the species i is close to criticality appears in the fact that both the birth and the death rate have the same leading term in $K\sigma_i$ with the same coefficient σ_i. This coefficient can depend on the species.

Proposition 5.3. *Assume* (5.1), (5.5) *and the rates* (5.10). *Then, when* $K \to +\infty$, *the process* $(\boldsymbol{X}^K(t))_{t\geq 0}$ *converges uniformly on every compact time interval* $[0, T]$ *with* $T > 0$ *and in distribution to the solution of the following Feller stochastic differential equation:*

$$\mathrm{d}\boldsymbol{X}(t) = (\boldsymbol{r} - \boldsymbol{X}(t) + \Gamma\boldsymbol{X}(t)) \circ \boldsymbol{X}(t)\,\mathrm{d}t + \sqrt{2\Sigma\boldsymbol{X}(t)} \circ \mathrm{d}\boldsymbol{B}_t \tag{5.11}$$

where $\boldsymbol{B}$ *is a* N*-dimensional standard Brownian motion independent of* Γ, *and where the function* $x \mapsto \sqrt{x}$ *is applied elementwise to the vector* $\Sigma\boldsymbol{X}(t)$.

The random noise appearing in (5.11) comes from the rapid successions of birth and death events in this accelerated time-scale. See [4, Part 4(a)] for details.

(ii) Variations around the Feller equations

In [123], Biroli et al. added an immigration factor $\lambda > 0$. The SDEs they consider are:

$$\mathrm{d}\boldsymbol{X}(t) = \boldsymbol{X}(t) \circ (\mathbf{1}_N - \boldsymbol{X}(t) + \Gamma\boldsymbol{X}(t))\,\mathrm{d}t + \lambda\mathbf{1}_N\mathrm{d}t + f(\boldsymbol{X}(t)) \circ \mathrm{d}\boldsymbol{B}(t), \tag{5.12}$$

where $\boldsymbol{B}(t)$ is a N-dimensional standard Brownian motion independent of Γ, and where the function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is applied elementwise to any vector $\boldsymbol{x} = (x_i)$, i.e. $f(\boldsymbol{x}) = (f(x_i))$.

In the framework of a symmetric interaction matrix, Biroli et al. [123] unveil, by the replica method, the large-N system behavior of $\boldsymbol{X}$, recovering the parameter regions for which the system has a single equilibrium or multiple equilibria. Another generalization was done by Roy et al. in [124], where Γ can follow the general elliptic model. These authors study the large-N limit of the SDE (5.12) by relying on a dynamical mean field approach based on the dynamical cavity method, as detailed in [125].

A better understanding of these results from a mathematical perspective, as well as their generalization to more sophisticated models than the elliptical model, are interesting and useful research directions which have not been undertaken so far to the best of our knowledge. One of the first mathematical formalizations of this class of problems goes back to the work of Ben Arous and Guionnet [126], who were interested in the dynamics of mean-field spin glasses. In that setting, their analogue of our SDE (5.12) is a Langevin version of the so-called Sherrington-Kirkpatrick model for the spin glass dynamics. In the same line of thought, Faugeras et al. [127,128] used the approach of Ben Arous and Guionnet to study a diffusion version of the so-called Hopfield model for biological neural networks. Details are given in [4, Part 4(b)].

6. Discussion

Our guided tour through large LV models has highlighted the importance of some features of matrix Γ to understand the dynamics of complex ecological systems. Among these features,

the first two statistical moments of the interactions, μ/N and $\sigma^2/(N\alpha^2)$, are of paramount importance, as well as the scaling of the normalisation applied to matrix Γ ($1/N$ for the deterministic part of the matrix, $1/\sqrt{N}$ or $1/\alpha_N\sqrt{N}$ for its random part).

Our review covers some key topics on the analysis of large LV models, i.e. conditions leading to a feasible equilibrium, to a unique maximal equilibrium and the link between the interaction matrix and equilibrium stability (Sec. 3). In Sec. 4 and [4, Section 5], we mention some sophistications (block, trait-based, sparse matrices) of the model that can make it more realistic. Theoretical results obtained on kernel matrices (many traits, many species) [4, Thm. 5.3] suggest that the study of LV models obtained from trait-based interaction matrices could be simplified by assuming that many interaction matrix spectra will resemble that of a Wishart matrix. However, this result does not yield any direct conclusion on the feasibility of an equilibrium of the LV model, nor does it help generalize the heuristics 3.5 to such structured models.

We have not considered models based on fully organized interaction matrices, which are not amenable to an RMT analysis but sometimes allow some direct analysis through classic analytical tools (e.g. Lyapunov functions, see also [129]). For instance, modelling food webs as strictly organized by trophic levels (i.e. species from trophic level k can only be positively affected by some species of level $k-1$ and negatively affected by some species of level $k+1$), it is possible to express coexistence conditions as mathematical conditions on the covering of the food web by pairs of interacting species [130,131]. This finding, which highlights the difficulty of having a feasible equilibrium in level-organized food webs, hints at the potential importance of omnivory (i.e. that interactions are not strictly organized by levels, such that a predator can also feed on the prey of its prey species) to explain the stability of real food webs. An interesting endeavour could be to formalize omnivory in the context of random interaction matrices in order to tease apart the effect of omnivory from that of the food web being acyclic (which would also break the trophic level-based nature of the interaction matrix used by [130,131]).

(a) Empirical data and real interaction networks

There is a huge literature on statistical models for ecological networks, see [132–139]. The variety of data available and the specificity of each ecosystem explain this vast corpus, which in itself would deserve an entire review.

Many papers consider variations of the LV system, for instance integrating functional responses [140] (and references therein). The inference and empirical testing of models is limited by the type and volume of data that can be obtained through the observation of natural systems or controlled experiments. For most systems, we can only access abundances x_i (often in a single snapshot that may not be at equilibrium, less frequently in time series). For some systems, especially with predator-prey or pollinator-plant interactions, we can also know which interactions Γ_{ij} are zero or nonzero, as species may be limited in which partners they can interact with, but quantitative estimates of nonzero interactions remain out of reach. Therefore, exhaustive information on parameters r_i and Γ_{ij} is not accessible without additional assumption [10]: the main method for estimating them reliably is to observe the growth rates and/or equilibrium abundances of each species in isolation and in combinations, which is only possible in small-N experiments. Abundance distributions have been exploited to fit many other models, e.g neutral theory [13]. Additional data may help estimation and model validation [69,141,142]. Finally, notice that although the models discussed above have a high number of parameters, their distributions have low degrees of freedom, see for instance the elliptic model.

It is important to keep in mind that real interaction networks such as food webs are likely not randomly structured. Interactions between species play a crucial role in their respective fitness, so that when given sufficient variation in traits governing interaction strength, coevolution of species will lead to particular network structures [30,31,101,143]. In the sequential invasion simulations of food webs [31], the spectrum of the Jacobian matrix of the underlying LV system does not converge towards an ellipsis as the number of invasion attempts increases, but rather becomes bimodal, thus highlighting a possible structuring effect of evolution sensu lato. In another

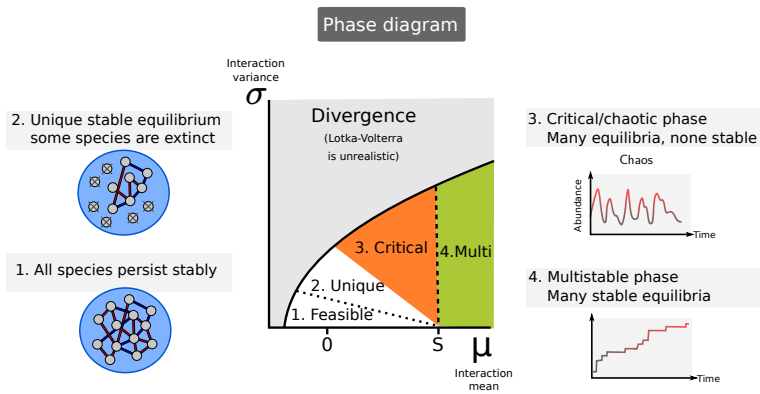


Figure 9: Phase diagram representing the qualitative dynamical regimes of the reference random model (first obtained numerically in [148]).

sequential invasion model, Bastolla et al. [143] found that the evolution of food webs enforces two constraints on species diversity through competition among species preying on the same prey species and through an increase in competition strength at higher trophic levels due to the propagation of fluctuations and energy dissipation. In food web models incorporating realistic, non-invasion-based evolution, species tend to naturally group into trophic levels [30,144,145], an absent feature in most random matrix models. The various structures described in section 4 can be used to investigate their effects on equilibria and their stability, but it is still a long way from assessing the effects of evolutionary processes on these properties, e.g. the effect of the speciation-to-invasion ratio on food web susceptibility to further invasions [146].

(b) Open mathematical and modelling problems

Multiple equilibria and non-equilibrium attractors

While our review focuses on conditions under which there exists a unique stable equilibrium (Sec. 3), other regimes including multiple equilibria or non-equilibrium attractors (e.g. chaos or cycles) have been studied using physics tools [16,123,147], see for instance Fig. 9. A mathematical understanding of these regimes and the associated thresholds is challenging and would represent an important step in the understanding of LV systems.

Quantitative metrics for stability

As hinted in Sec. 5 and [4, Sec. 3], other dynamical properties of LV systems (e.g. dynamics of fluctuations around equilibria), can also be studied using stochastic differential equation analogues to the LV ODEs. More quantitative metrics of stability are also of ecological interest, e.g. quantifying how much attractors change in response to perturbations of model parameters [149], assessing how much model parameters can change without changing the dynamical attractor (i.e. structural stability, [54]), or characterizing the time and trajectory of return to an attractor after a perturbation in the abundances (e.g. chap. 58 in [150]). When perturbations are small, one can linearize the dynamics around an attractor, and compute stability metrics for the i.i.d. or elliptic model [19,151,152]. Beyond the linearized regime, tools such as Freidlin-Wentzell theory can help quantify the basins of dynamical attractors and transition times between them [153].

Counting the number of coexisting species

In Section 3(e), we provide heuristics to compute the proportion of vanishing species for a given equilibrium and refer to articles relying on techniques from Physics. A mathematical computation still remains out of reach and could be phrased as understanding the properties of solutions of Linear Complementarity Problems with random matrices as input. Part of the challenge lies on the fact that the properties of interest of the solution of a LCP (proportion of non-vanishing components) do not only rely on the spectral properties of the random matrices at hand. Table 1 in Section 3 lists a number of open questions when studying a specific LV model.

A mathematical understanding of the cavity method

The cavity method, originally introduced in the context of spin glasses [125], has become extremely versatile and in general yields to closed-form expressions accurately matching simulation results. There are quite few mathematically rigorous counterparts of the results obtained through the cavity method (one can cite e.g. [154] for sparse graphs). In the context of large ecosystems, we have worked in this direction in [155,156], using mathematical results on the Approximate Message Passing algorithm to recover rigorously some equations obtained in [58] through the cavity method. There is still much to do to get a wider mathematical formalization of the cavity method for the analysis of non-spectral properties of large random matrix observables.

Acknowledgements. All authors are supported by the CNRS 80 prime project KARATE. I.A., M.C., W.H., M.M., J.N. are supported by GdR MEGA. I.A. and M.M. are funded by Labex CEMPI (ANR-11-LABX-0007-01). M.C., W.H., J.N. and V.C.T. are supported by Labex Bézout (ANR-10-LABX-0058). F.M. and V.C.T. are funded by ANR EcoNet (ANR-18-CE02-0010). I.A. and V.C.T. acknowledges support of the "Chaire Modélisation Mathématique et Biodiversité (MMB)" of Veolia Environnement-Ecole Polytechnique-Museum National d'Histoire Naturelle-Fondation X. V.C.T. is partly funded by the European Union (ERC-AdG SINGER-101054787). M.M. and V.C.T. thank for its hospitality the CRM Montréal (IRL CNRS 3457) where part of this work was completed.

References

1. Lotka AJ. 1925 *Elements of Physical Biology*. Baltimore, MD: Williams and Watkins.
2. Volterra V. 1931 Variations and fluctuations of the number of individuals in animal species living together. In Chapman RN, editor, *Animal Ecology* pp. 409–448 New York. McGraw-Hill.
3. Odum E, Barrett G. 1971 *Fundamentals of ecology* vol. 3. Saunders Philadelphia.
4. Akjouj I, Barbier M, Clenet M, Hachem W, Maïda M, Massol F, Najim J, Tran VC. 2023 Supplementary Material / Complex systems in Ecology: a guided tour with large Lotka-Volterra models and random matrices. .
5. Marrow P, Law R, Cannings C. 1992 The coevolution of predator-prey interactions : ESSS and Red Queen dynamics. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **250**, 133–141.
6. Champagnat N, Ferrière R, Méléard S. 2006 Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models via timescale separation. *Theoretical Population Biology* **69**, 297–321.
7. May RM. 1972 Will a large complex system be stable?. *Nature* **238**, 413–414.
8. Wigner EP. 1967 Random matrices in physics. *SIAM Review* **9**, 1–23.
9. Tang S, Pawar S, Allesina S. 2014 Correlation between interaction strengths drives stability in large ecological networks. *Ecology Letters* **17**, 1094–1100.
10. Jacquet C, Moritz C, Morissette L, Legagneux P, Massol F, Archambault P, Gravel D. 2016 No complexity-stability relationship in empirical ecosystems. *Nature Communications* **7**, 12573.
11. Neutel AM, Heesterbeek JAP, van de Koppel J, Hoenderboom G, Vos A, Kaldeway C, Berendse F, de Ruiter PC. 2007 Reconciling complexity with stability in naturally assembling food webs. *Nature* **449**, 599–602.
12. James A, Plank MJ, Rossberg AG, Beecham J, Emmerson M, Pitchford JW. 2015 Constructing Random Matrices to Represent Real Ecosystems. *The American Naturalist* **185**, 680–692.

13. Hubbell SP. 2001 *The unified neutral theory of biodiversity and biogeography (MPB-32)* vol. 32. Princeton University Press.
14. Hastings A, Abbott KC, Cuddington K, Francis T, Gellner G, Lai YC, Morozov A, Petrovskii S, Scranton K, Zeeman ML. 2018 Transient phenomena in ecology. *Science* **361**.
15. Allesina S, Tang S. 2012 Stability criteria for complex ecosystems. *Nature* **483**, 205–208.
16. Roy F, Barbier M, Biroli G, Bunin G. 2020 Complex interactions can create persistent fluctuations in high-diversity ecosystems. *PLoS computational biology* **16**, e1007827.
17. Lehman C, Tilman D. 2000 Biodiversity, stability, and productivity in competitive communities. *American Naturalist* **156**, 534–552.
18. Kokkoris G, Jansen V, Loreau M, Troumbis A. 2002 Variability in interaction strength and implications for biodiversity. *Journal of Animal Ecology* **71**, 362–371.
19. Arnoldi JF, Haegeman B. 2016 Unifying dynamical and structural stability of equilibria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **472**, 20150874.
20. Loreau M, de Mazancourt C. 2013 Biodiversity and ecosystem stability: a synthesis of underlying mechanisms. *Ecology Letters* **16**, 106–115.
21. Wang S, Loreau M. 2014 Ecosystem stability in space: α, β and γ variability. *Ecology Letters* **17**, 891–901.
22. Gardner MR, Ashby WR. 1970 Connectance of large dynamic (cybernetic) systems: critical values for stability. *Nature* **228**, 784–784.
23. Roberts A. 1974 The stability of a feasible random ecosystem. *Nature* **251**, 607–608.
24. Gilpin ME. 1975 Stability of feasible predator-prey systems. *Nature* **254**, 137–139.
25. Allesina S, Tang S. 2015 The stability–complexity relationship at age 40: a random matrix perspective. *Population Ecology* **57**, 63–75.
26. Otto SB, Rall BC, Brose U. 2007 Allometric degree distributions facilitate food-web stability. *Nature* **450**, 1226–1229.
27. Gravel D, Massol F, Leibold MA. 2016 Stability and complexity in model meta-ecosystems. *Nature Communications* **7**, 12457.
28. Mougi A, Kondoh M. 2016 Food-web complexity, meta-community complexity and community stability. *Scientific Reports* **6**, 24478.
29. Eklöf A, Jacob U, Kopp J, Bosch J, Castro-Urgal R, Chacoff NP, Dalsgaard B, de Sassi C, Galetti M, Guimarães PR, Lomáscolo SB, Martín González AM, Pizo MA, Rader R, Rodrigo A, Tylianakis JM, Vázquez DP, Allesina S. 2013 The dimensionality of ecological networks. *Ecology Letters* **16**, 577–583.
30. Allhoff KT, Ritterskamp D, Rall BC, Drossel B, Guill C. 2015 Evolutionary food web model based on body masses gives realistic networks with permanent species turnover. *Scientific Reports* **5**.
31. Låstad SB, Haerter JO. 2022 The geometry of evolved community matrix spectra. *Scientific Reports* **12**, 14668. (10.1038/s41598-022-17379-6)
32. Tang S, Allesina S. 2014 Reactivity and stability of large ecosystems. *Frontiers in Ecology and Evolution* **2**.
33. Bordenave C, Chafaï D. 2012 Around the circular law. *Probability surveys* **9**, 1–89.
34. MacArthur R. 1969 Species packing, and what competition minimizes. *Proceedings of the National Academy of Sciences* **64**, 1369–1371.
35. Redheffer R. 1985 Volterra multipliers I. *SIAM Journal on Algebraic Discrete Methods* **6**, 592–611.
36. Bai Z, Silverstein JW. 2010 *Spectral analysis of large dimensional random matrices* vol. 20. Springer.
37. Girko V. 1986 Elliptic law. *Theory of Probability & Its Applications* **30**, 677–690.
38. Sommers HJ, Crisanti A, Sompolinsky H, Stein Y. 1988 Spectrum of large random asymmetric matrices. *Physical Review Letters* **60**, 1895–1898.
39. O'Rourke S, Renfrew D. 2014a Low rank perturbations of large elliptic random matrices. *Electronic Journal of Probability* **19**.
40. O'Rourke S, Renfrew D. 2014b Low rank perturbations of large elliptic random matrices. *Electronic Journal of Probability* **19**, 1 – 65. (10.1214/EJP.v19-3057)
41. Takeuchi Y. 1996 *Global dynamical properties of Lotka-Volterra systems*. World Scientific.
42. Hofbauer J, Sigmund K. 1988 *The Theory of Evolution and Dynamical Systems: Mathematical Aspects of Selection*. London Mathematical Society Stundent Texts. Cambridge University Press.
43. Hofbauer J, Sigmund K et al.. 1998 *Evolutionary games and population dynamics*. Cambridge university press.

44. Law R, Morton RD. 1996 Permanence and the assembly of ecological communities. *Ecology* **77**, 762–775.
45. Metz J, Geritz S, Meszéna G, Jacobs F, Heerwaarden JV. 1996 Adaptative Dynamics, a Geometrical Study of the Consequences of Nearly Faithful Reproduction. *S.J. Van Strien & S.M. Verduyn Lunel (ed.), Stochastic and Spatial Structures of Dynamical Systems* **45**, 183–231.
46. Robinson JF, Dickerson JE. 1987 Does invasion sequence affect community structure?. *Ecology* **68**, 587–595.
47. Champagnat N, Méléard S. 2011 Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields* **151**, 45–94.
48. Rummel JD, Roughgarden J. 1985 A theory of faunal buildup for competition communities. *Evolution* **39**, 1009–1033.
49. Taper ML, Case TJ. 1992 Models of character displacement and the theoretical robustness of taxon cycles. *Evolution* pp. 317–333.
50. Calcagno V, Jarne P, Loreau M, Mouquet N, David P. 2017 Diversity spurs diversification in ecological communities. *Nature Communications* **8**, 15810.
51. Romanuk TN, Binzer A, Loeuille N, Carscallen WMA, Martinez ND. 2019 Simulated evolution assembles more realistic food webs with more functionally similar species than invasion. *Scientific Reports* **9**, 18242.
52. Case TJ. 1990 Invasion resistance arises in strongly interacting species-rich model competition communities. *Proceedings of the National Academy of Sciences* **87**, 9610–9614.
53. Bastolla U, Lässig M, Manrubia SC, Valleriani A. 2005 Biodiversity in model ecosystems, I: coexistence conditions for competing species. *Journal of Theoretical Biology* **235**, 521–530.
54. Saavedra S, Rohr RP, Bascompte J, Godoy O, Kraft N, Levine JM. 2017 A structural approach for understanding multispecies coexistence. *Ecological Monographs* **87**, 470–486.
55. Murty K, Yu FT. 1997 *Linear Complementarity, Linear and Nonlinear Programming*. Internet Edition.
56. Cottle RW, Pang JS, Stone RE. 2009 *The linear complementarity problem*. SIAM.
57. Clenet M, El Ferchichi H, Najim J. 2022 Equilibrium in a large Lotka-Volterra system with pairwise correlated interactions. *Stochastic Processes and their Applications* **153**.
58. Bunin G. 2017 Ecological communities with Lotka-Volterra dynamics. *Physical Review E* **95**, 042414.
59. Goh BS. 1977 Global stability in many-species systems. *American Naturalist* **111**, 135–143.
60. Champagnat N, Jabin PE, Raoul G. 2010 Convergence to equilibrium in competitive Lotka–Volterra and chemostat systems. *Comptes Rendus Mathematique* **348**, 1267–1272.
61. Dougoud M, Vinckenbosch L, Rohr R, Bersier LF, Mazza C. 2018 The feasibility of equilibria in large ecosystems: A primary but neglected concept in the complexity-stability debate. *PLoS computational biology* **14**, e1005988.
62. Bizeul P, Najim J. 2021 Positive solutions for large random linear systems. *Proceedings of the American Mathematical Society* **149**, 2333–2348.
63. Geman S. 1986 The spectral radius of large random matrices. *The Annals of Probability* **14**, 1318–1328.
64. Geman S, Hwang CR. 1982 A chaos hypothesis for some large systems of random equations. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **60**, 291–314.
65. Galla T. 2018 Dynamically evolved community size and stability of random Lotka-Volterra ecosystems (a). *EPL (Europhysics Letters)* **123**, 48004.
66. Clénet M, Massol F, Najim J. 2023 Equilibrium and surviving species in a large Lotka-Volterra system of differential equations. *Journal of Mathematical Biology* **87**. (10.1007/s00285-023-01939-z)
67. Clenet M, Massol F, Najim J. 2022 Surviving species in a Large Lotka-Volterra system of differential equations. In *28e Colloque sur le traitement du signal et des images* vol. 001-0257 pp. p. 1029–1032 Nancy. GRETSI - Groupe de Recherche en Traitement du Signal et des Images.
68. Bunin G. 2016 Interaction patterns and diversity in assembled ecological communities. *arXiv preprint arXiv:1607.04734*.
69. Barbier M, De Mazancourt C, Loreau M, Bunin G. 2021 Fingerprints of high-dimensional coexistence in complex ecosystems. *Physical Review X* **11**, 011009.
70. Baron JW, Jewell TJ, Ryder C, Galla T. 2022 Non-Gaussian random matrices determine the stability of Lotka-Volterra communities. *arXiv preprint arXiv:2202.09140*.
71. Pimm SL. 1979 The structure of food webs. *Theoretical Population Biology* **16**, 144–158. (https://doi.org/10.1016/0040-5809(79)90010-8)

72. Grilli J, Adorisio M, Suweis S, Barabás G, Banavar JR, Allesina S, Maritan A. 2017 Feasibility and coexistence of large ecological communities. *Nature communications* **8**, 14389.
73. Marcus S, Turner AM, Bunin G. 2022 Local and collective transitions in sparsely-interacting ecological communities. *PLoS computational biology* **18**, e1010274.
74. Dunne JA, Williams RJ, Martinez ND. 2002 Food-Web Structure and Network Theory: The Role of Connectance and Size. *Proceedings of the National Academy of Sciences - PNAS* **99**, 12917–12922.
75. Busiello DM, Suweis S, Hidalgo J, Maritan A. 2017 Explorability and the origin of network sparsity in living systems. *Scientific Reports* **7**. (10.1038/s41598-017-12521-1)
76. Bollobás B. 2001 *Random graphs*. Cambridge University Press 2 edition.
77. Durrett R. 2007 *Random graph dynamics*. New York: Cambridge University Press.
78. der Hofstad RV. 2017 *Random Graphs and Complex Networks* vol. 1*Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge University Press.
79. Alt J, Ducatez R, Knowles A. 2023 Poisson statistics and localization at the spectral edge of sparse Erdős–Rényi graphs. *The Annals of Probability* **51**, 277 – 358. (10.1214/22-AOP1596)
80. Akjouj I, Najim J. 2022 Feasibility of sparse large Lotka-Volterra ecosystems. *Journal of Mathematical Biology* **85**, 66.
81. Holland P, Laskey K, Leinhardt S. 1983 Stochastic blockmodels: some first steps. *Social Networks* **5**, 109–137.
82. Abbe E. 2018 Community detection and stochastic block models: recent development. *Journal of Machine Learning Research* **18**, 1–86.
83. Lee C, Wilkinson DJ. 2019 A review of stochastic block models and extensions for graph clustering. *Applied Network Science* **4**.
84. Matias C, Miele V. 2017 Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **79**, 1119–1141.
85. Baskerville EB, Dobson AP, Bedford T, Allesina S, Anderson TM, Pascual M. 2011 Spatial Guilds in the Serengeti Food Web Revealed by a Bayesian Group Model. *PLOS Computational Biology* **7**, 1–11. (10.1371/journal.pcbi.1002321)
86. Benaych-Georges F, Bordenave C, Knowles A. 2020 Spectral radii of sparse random matrices. *Ann. Inst. H. Poincaré Probab. Statist.* **56**, 2141–2161.
87. Thébault E, Fontaine C. 2010 Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* **329**, 853–856.
88. Barber MJ. 2007 Modularity and community detection in bipartite networks. *Phys. Rev. E* **76**, 066102. (10.1103/PhysRevE.76.066102)
89. Stouffer DB, Bascompte J. 2011 Compartmentalization increases food-web persistence. *Proceedings of the National Academy of Sciences - PNAS* **108**, 3648–3652.
90. Billiard S, Leman H, Rey T, Tran V. 2022 Continuous limits of large plant-pollinator random networks and some applications. *MathematicS In Action*.
91. Lovász L. 2012 *Large networks and graph limits* vol. 60*American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI. (10.1090/coll/060)
92. Molloy M, Reed B. 1995 A critical point for random graphs with a given degree sequence. *Random structures and algorithms* **6**, 161–180.
93. Okuyama T, Holland JN. 2008 Network structural properties mediate the stability of mutualistic communities. *Ecology letters* **11**, 208–216.
94. Chung F, Lu L, Vu V. 2003 Spectra of Random Graphs with Given Expected Degrees. *Proceedings of the National Academy of Sciences - PNAS* **100**, 6313–6318.
95. Marčenko VA, Pastur LA. 1967 The spectrum of random matrices. *Teor. Funkciĭ Funkcional. Anal. i Priložen. Vyp.* **4**, 122–145.
96. Dieckmann U, Doebeli M. 1999 On the origin of species by sympatric speciation. *Nature* **400**, 354–357.
97. Roughgarden J. 1979 *Theory of population genetics and evolutionary ecology: an introduction*. New York: Macmillan.
98. Meszéna G, Gyllenberg M, Pásztor L, Metz JA. 2006 Competitive exclusion and limiting similarity: A unified theory. *Theoretical population biology* **69**, 68–87.
99. Nuismer SL, Jordano P, Bascompte J. 2013 Coevolution and the architecture of mutualistic networks. *Evolution* **67**, 338–354.
100. Kisdi E. 1999 Evolutionary branching under asymmetric competition. *J. Theor. Biol.* **197**, 149–162.

101. Siliansky de Andreazzi C, Astegiano J, Guimarães Jr PR. 2020 Coevolution by different functional mechanisms modulates the structure and dynamics of antagonistic and mutualistic networks. *Oikos* **129**, 224–237.
102. Santamaria L, Rodriguez-Girones MA. 2007 Linkage rules for plant-pollinator networks: Trait complementarity or exploitation barriers?. *PLoS biology* **5**, 354–362.
103. Rohr RP, Naisbit RE, Mazza C, Bersier LF. 2016 Matching-centrality decomposition and the forecasting of new links in networks. *Proceedings of the Royal Society. B, Biological sciences* **283**.
104. Serván CA, Capitán JA, Miller ZR, Allesina S. 2020 Effects of phylogeny on coexistence in model communities. *bioRxiv*. (10.1101/2020.09.04.283507)
105. Fyodorov Y, Khoruzhenko B. 2016 Nonlinear analogue of the May-Wigner instability transition. *Proceedings of the National Academy of Sciences* **113**, 6827–6832.
106. Ben Arous G, Fyodorov Y, Khoruzhenko B. 2021 Counting equilibria of large complex systems by instability index. *Proceedings of the National Academy of Sciences* **118**, e2023719118.
107. Tarnowski W, Neri I, Vivo P. 2020 Universal transient behavior in large dynamical systems on networks. *Phys. Rev. Res.* **2**, 023333. (10.1103/PhysRevResearch.2.023333)
108. Bansaye V, Méléard S. 2015 *Stochastic models for structured populations* vol. 1*Mathematical Biosciences Institute Lecture Series. Stochastics in Biological Systems*. Springer, Cham; MBI Mathematical Biosciences Institute, Ohio State University, Columbus, OH. Scaling limits and long time behavior (10.1007/978-3-319-21711-6)
109. Krumbeck Y, Yang Q, Constable GW, Rogers T. 2021 Fluctuation spectra of large random dynamical systems reveal hidden structure in ecological networks. *Nature Communications* **12**, 3625.
110. Billingsley P. 1999 *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, Inc. second edition edition.
111. Meyn S, Tweedie R. 1993a Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes. *Advances in Applied Probability* **25**, 518–548.
112. Meyn S, Tweedie R. 1993b Stability of Markovian Processes II: Continuous time processes and sampled chains. *Advances in Applied Probability* **25**, 487–517.
113. DeAngelis DL, Grimm V. 2014 Individual-based models in ecology after four decades. *F1000prime reports* **6**.
114. DeAngelis DL. 2018 *Individual-based models and approaches in ecology: populations, communities and ecosystems*. CRC Press.
115. Ferriere R, Tran VC. 2009 Stochastic and deterministic models for age-structured populations with genetically variable traits. In *ESAIM: Proceedings* vol. 27 pp. 289–310. EDP Sciences.
116. Grimm V, Berger U, Bastiansen F, Eliassen S, Ginot V, Giske J, Goss-Custard J, Grand T, Heinz SK, Huse G et al.. 2006 A standard protocol for describing individual-based and agent-based models. *Ecological modelling* **198**, 115–126.
117. Giorgi D, Kaakai S, Lemaire V. 2020 IBMPOPSIM R Package. https://cran.r-project.org/package=IBMPopSim.
118. Legendre S. 2020 ZEN, Eco-evolutionary software. https://www.biologie.ens.fr/ legendre/zen/zen.html.
119. Ethier S, Kurtz T. 1986 *Markov Processus, Characterization and Convergence*. New York: John Wiley & Sons.
120. Fournier N, Méléard S. 2004 A Microscopic Probabilistic Description of a Locally Regulated Population and Macroscopic Approximations. *Ann. Appl. Probab.* **14**, 1880–1919.
121. Haegeman B, Loreau M. 2011 A mathematical synthesis of niche and neutral theories in community ecology. *Journal of theoretical biology* **269**, 150–165.
122. Chazottes JR, Collet P, Méléard S. 2019 On time scales and quasi-stationary distributions for multitype birth-and-death processes. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **55**, 2249–2294.
123. Biroli G, Bunin G, Cammarota C. 2018 Marginally stable equilibria in critical ecosystems. *New Journal of Physics* **20**, 083051.
124. Roy F, Biroli G, Bunin G, Cammarota C. 2019 Numerical implementation of dynamical mean field theory for disordered systems: Application to the Lotka–Volterra model of ecosystems. *Journal of Physics A: Mathematical and Theoretical* **52**, 484001.
125. Mézard M, Parisi G, Virasoro M. 1986 *Spin Glass Theory and Beyond*. World Scientific. (10.1142/0271)
126. Ben Arous G, Guionnet A. 1995 Large deviations for Langevin spin glass dynamics. *Probability Theory and Related Fields* **102**, 455–509.

127. Faugeras O, Touboul J, Cessac B. 2009 A constructive mean-field analysis of multi population neural networks with random synaptic weights and stochastic inputs. *Frontiers in Computational Neuroscience* **3**, 1. (10.3389/neuro.10.001.2009)
128. Cabana T, Touboul J. 2013 Large deviations, dynamics and phase transitions in large stochastic and disordered neural networks. *Journal of Statistical Physics* **153**, 211–269.
129. Gouzé JL. 1993 Global behavior of n-dimensional Lotka–Volterra systems. *Mathematical Biosciences* **113**, 231–243.
130. Haerter JO, Mitarai N, Sneppen K. 2016 Food web assembly rules for generalized Lotka-Volterra equations. *PloS Computational Biology* **12**, e1004727.
131. Haerter JO, Mitarai N, Sneppen K. 2017 Existence and construction of large stable food webs. *Physical Review E* **96**, 032406.
132. Botella C, Dray S, Matias C, Miele V, Thuiller W. 2022 An appraisal of graph embeddings for comparing trophic network architectures. *Methods in Ecology and Evolution* **13**, 203–216.
133. Cirtwill A, Eklöf A, Roslin T, Wootton K, Gravel D. 2019 A quantitative framework for investigating the reliability of empirical network construction. *Methods in Ecology and Evolution* **10**, 902–911.
134. de Siqueira Santos S, Fujita A, Matias C. 2021 Spectral density of random graphs: convergence properties and application in model fitting. *Journal of Complex Networks* **9**.
135. Matias C, Rebafka T, Villers F. 2018 A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika* **105**, 665–680.
136. Miele V, Matias C. 2017 Revealing the hidden structure of dynamic ecological networks. *Royal Society Open Science* **4**, 170251.
137. Miele V, Matias C, Ohlmann M, Poggiato G, Dray S, Thuiller W. 2021 Quantifying the overall effect of biotic interactions on species communities along environmental gradients. The code for ELGRIN is part of the Econetwork R package.
138. Momal R, Robin S, Ambroise C. 2021 Accounting for missing actors in interaction network inference from abundance data. *Journal of the Royal Statistical Society Series C: Applied Statistics* **70**, 1230–1258.
139. Momal R, Robin S, Ambroise C. 2020 Tree-based inference of species interaction networks from abundance data. *Methods in Ecology and Evolution* **11**, 621–632. (https://doi.org/10.1111/2041-210X.13380)
140. Bansaye V, Billiard S, Chazottes JR. 2018 Rejuvenating functional responses with renewal theory. *Journal of the Royal Society Interface* **15**, 20180239.
141. Hu J, Amor DR, Barbier M, Bunin G, Gore J. 2022 Emergent phases of ecological diversity and dynamics mapped in microcosms. *Science* **378**, 85–89.
142. Fort H. 2018 Quantitative predictions from competition theory with an incomplete knowledge of model parameters tested against experiments across diverse taxa. *Ecological Modelling* **368**, 104–110.
143. Bastolla U, Lässig M, Manrubia SC, Valleriani A. 2005 Biodiversity in model ecosystems, II: species assembly and food web structure. *Journal of Theoretical Biology* **235**, 531–539.
144. Drossel B, Higgs PG, McKane AJ. 2001 The influence of predator-prey population dynamics on the long-term evolution of food web structure. *Journal of Theoretical Biology* **208**, 91–107. (10.1006/jtbi.2000.2203)
145. Loeuille N, Loreau M. 2005 Evolutionary emergence of size-structured food webs. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5761–5766.
146. Tokita K, Yasutomi A. 2003 Emergence of a complex and stable network in a model ecosystem with extinction and mutation. *Theoretical Population Biology* **63**, 131–146.
147. Ros V, Roy F, Biroli G, Bunin G, Turner AM. 2023 Generalized Lotka-Volterra Equations with Random, Nonreciprocal Interactions: The Typical Number of Equilibria. *Physical Review Letters* **130**. (10.1103/physrevlett.130.257401)
148. Kessler DA, Shnerb NM. 2015 Generalized model of island biodiversity. *Physical Review E* **91**, 042705.
149. Barabás G, Meszéna G, Ostling A. 2014 Fixed point sensitivity analysis of interacting structured populations. *Theoretical Population Biology* **92**, 97–106.
150. Trefethen LN, Embree M. 2005 *Spectra and pseudospectra: the behavior of nonnormal matrices and operators*. Princeton University Press.
151. Arnoldi JF, Bideault A, Loreau M, Haegeman B. 2018 How ecosystems recover from pulse perturbations: A theory of short-to long-term responses. *Journal of theoretical biology* **436**, 79–92.

152. Arnoldi JF, Loreau M, Haegeman B. 2019 The inherent multidimensionality of temporal variability: how common and rare species shape stability patterns. *Ecology letters* **22**, 1557–1567.
153. Rodríguez-Sánchez P, Van Nes EH, Scheffer M. 2020 Climbing Escher's stairs: a way to approximate stability landscapes in multidimensional systems. *PLoS computational biology* **16**, e1007788.
154. Bordenave C, Lelarge M. 2010 Resolvent of large random graphs. *Random Structures Algorithms* **37**, 332–352. (10.1002/rsa.20313)
155. Akjouj I, Hachem W, Maïda M, Najim J. 2023 Equilibria of large random Lotka-Volterra systems with vanishing species: a mathematical approach. .
156. Hachem W. 2023 Approximate Message Passing for sparse matrices with application to the equilibria of large ecological Lotka-Volterra systems. .