Interplay of structured and random interactions in many-species ecological dynamics

Juan Giral Martínez,[1] Matthieu Barbier,[2,3,4] and Silvia De Monte[1,5]

[1] *Institut de Biologie de l'École Normale Supérieure, Département de Biologie, École Normale Supérieure, PSL Research University, Paris, France*
[2] *CIRAD, UMR PHIM, 34090 Montpellier, France.*
[3] *PHIM Plant Health Institute, University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, 34090 Montpellier, France.*
[4] *Institut Natura e Teoria en Pirenèus, Surba, France.*
[5] *Max Planck Institute of Evolutionary Biology, Plön, Germany.*

A growing body of 'disordered systems' theory investigates minimal ecosystem models with random interactions, where simple statistics suffice to predict dynamical outcomes and macroecological patterns. However, interaction networks commonly possess large-scale structures, such as hierarchies or functional groups, that have also been used to explain community-level patterns. Here, we ask which results from random interaction models can be expected to be fragile or robust to the presence of such large-scale network structures. We consider a simple superposition of structured and random interactions in a classic population dynamics model, and study its impact on macroscopic observables, abundance distributions and dynamical regimes. Randomness and structure combine in a surprisingly yet deceptively straightforward way: the studied patterns can almost be partitioned into contributions from each of these two components, and yet even this minimal interplay can lead to non-trivial effects. We conclude that whether interaction structure matters or not depends on which patterns we are testing for its consequences: we expect that the phenomenology associated with random interactions is less robust in static patterns of species presence and abundance, and more robust in the nature of possible dynamical regimes.

I. INTRODUCTION

Ecological communities, from microbiomes to food webs, may appear intractably complex due to the large number of entities and processes they harbour over a wide range of scales. One particular facet of this complexity is that, besides spatial, environmental or intra-species processes, community-level patterns and dynamics stem from interactions between a vast number of different species.

The founding work of Robert May [1] proposed to represent such complexity by choosing interactions randomly, thus focusing on predictable statistical consequences of intractably complex species relations. Inserting random interactions into classic population dynamics models, like the generalized Lotka-Volterra Equations (gLVEs), allowed to leverage methods from the statistical physics of disordered systems. Only a few statistics of these random variables turned out to be determinant for explaining patterns of ecological communities, such as abundance distributions and transitions between dynamical regimes [2–4]. At the meantime, the dimension of the relevant parameters' space is drastically reduced, thus facilitating comparison with observational data. Such models have proved appealing because of the simplicity and range of their predictions, and have found qualitative corroboration in recent experiments [5, 6].

However, it is unclear how widely the conclusions drawn from "disordered models" can be expected to hold. Indeed, because they treat all interactions as independent and species as statistically equivalent, they openly neglect dynamically significant ecological structure. This assumption may hold within a guild of relatively similar species, but broader ecological networks are usually conspicuously structured, and there is evidence that this structure is reflected in dynamics (e.g. highly unstable species dynamics, yet clear stability at the level of higher taxonomic [7, 8] or functional groups [9–11]).

Our goal here is to investigate which predictions of fully random models are altered by structure in species interactions. Among possible predictions, only equilibrium stability has received extensive attention along this line so far: a range of studies went beyond May's original results by combining randomness with various specific structures, e.g. trophic hierarchy, spatial clusters, sparse interactions or complex correlations [12–17]. One particular type of structure, bipartite networks of consumers and resources, has been more extensively explored in combination with disorder, e.g. with predictions on species abundances or dynamics [2, 18–20]. These diverse studies have generally found disparate results. Complementing these results in a more general context that encompasses several possible structures, we examine which properties of a disordered model are likely to be robust or fragile.

We consider a dynamical model of interacting populations, and assume that their interactions derive from many ecological processes, which can be partitioned into a large set of *hidden* processes, and a small set of *structural* processes. The hidden processes are represented by drawing interactions independently at random, as they arise from many different mechanisms [5] that are fixed but unknown, and may vary from context to context (hereafter, different "realizations" of the randomness). The structural processes reflect the modeller's ecological knowledge of the community. They reflect broad eco-

logical functions or mechanisms that many or all species participate in: for instance, interactions between any two species may be shaped by their impact on and response to key public goods (e.g. resources, pH), by their belonging to distinct functional groups, or through trophic relations mediated by traits (e.g. body size). A reduction in system complexity is key to predictability in both fully structured and fully random models. This is achieved by coarse-graining, *i.e.* identifying a few macroscopic variables, for instance the total abundance of functional groups, that account for static or dynamical properties of the community. Here, we study whether a coarse-grained description still exists as we interpolate between those two extremes, and what form it takes.

Besides the theoretical challenge, evaluating the relative importance of structured and uncharacterized variability on community-level functions has strong empirical relevance. At a time where highly diverse communities can be studied in quantitative terms by means of high-throughput genomic techniques, statistical methods are increasingly developed for separating species or environmental variables in clusters of ecological significance [7, 9, 21–23]. However, criteria are still lacking to decide whether conclusions on how such sets affect one another are robust to the unavoidable hidden variation within each cluster.

We show that the mathematical approach of disordered systems can be applied to a general model that combines structure and randomness. The community dynamics is therefore epitomized by a few community-level variables and parameters, associated to the structuring functions and to disorder. These macroscopic variables obey closed relations that constrain, as in the purely random case, species-level distributions at equilibrium, and whose dynamics converges, on the opposite limit, to that of the structural degrees of freedom. At intermediate levels, randomness and structure can moreover interfere in ways that unexpectedly impact community-level outcomes, as we discuss when the structure-driven dynamics is out-of-equilibrium.

II. METHODS

A. Model for the dynamics of species-rich, interacting communities

We describe the dynamics of the rescaled abundances $x_i(t)$ of S interacting species, labeled $i = 1, \ldots, S$, in the framework of generalized Lotka-Volterra Equations. These are the simplest non-linear model capturing the effect of ecosystem composition on individual demographic rates, which are influenced by both intra-specific competition and inter-specific interactions.

In order to focus on heterogeneity in the species' interactions, we consider an adimesional version of the equations, which is derived in the SI 1.1 from a more general formulation in terms of absolute abundances. The rescaling is possible under simplifying but non-restrictive assumptions.

The variation in time of the abundance of species i

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left[1 - x_i + \sum_j A_{ij} x_j \right] + m. \tag{1}$$

depends on the *interaction coefficients* A_{ij} that account for the impact of species j on the instantaneous growth rate of species i. Moreover, a small *immigration rate* $m \ll 1$ is added to avoid species getting permanently excluded from the community [3, 24]. Species that persist at the immigration level can therefore invade as soon as the community context becomes favorable. At any given time in the simulations, however, species whose abundance is smaller than a multiple of m are counted as extinct in order for immigration not to artificially enhance species diversity.

We model the interaction matrix $\mathbf{A}$ as the superposition of a *structural* and a *random* component, where the first encodes our ecological knowledge of the community and the second any hidden, unstructured process. Heterogeneity in the latter may reflect within-taxon variance, physiological variability, environmentally-mediated effects or any other feature too intricate to be accurately characterized. When the matrix is completely unstructured, eq. (1) reduces to the so-called disordered gLVEs, whose dynamics has been extensively characterized under different assumptions on the statistics of the random interactions [3, 24–26]. The structural part, on the other hand, is expected to be robust against changes in microscopic details and independent of unknown environmental variables. It reflects known factors, e.g. the belonging to a trophic or size class, the use of metabolic substrates or the production of public goods [7, 21, 27], in their effect on the fixed pair-wise interaction coefficients. The two terms of such decomposition can generally vary in their relative importance, bridging between two extreme cases that have been previously addressed with distinct mathematical approaches, as we discuss in Section III A.

B. Structured interactions

We define structure as resulting from the existence of a set of n_F community-level functions. We will index functions f_λ with Greek letters, $\lambda = 1, \ldots, n_F$. Every species can contribute to a community function and can be affected by it. This relationship is characterized by two sets of species-specific functional *traits* (Figure 2, a)[28]. Its *effect trait* $v_i^{(\lambda)}$ quantifies the per-capita contribution of species i to function λ. Its *response trait* $u_i^{(\lambda)}$ gives the impact of function λ on the per-capita growth rate of species i. For instance, as illustrated in Fig. 1

a, interactions between functional groups may be measurable, and community dynamics described in terms of the total abundance of species belonging to each group - while within-group variability is ignored. Alternatively, interactions could be mediated by a number of shared resources [29, 30], whose concentrations constitute the functions relevant for describing the community dynamics (Fig. 1 b). The effect traits are then the consumption rates, while the response traits are the rate of transformation into units of species abundance.

For simplicity, we consider the total *magnitude of the function* f_λ at time t to be a linear combination of the species' abundances,

$$f_\lambda(t) = \overline{v^{(\lambda)} x(t)} := \frac{1}{S} \sum_i v_i^{(\lambda)} x_i(t). \tag{2}$$

Here and in the following, the overline defines an average over species for a fixed set of parameters (interaction coefficients). Due to the factor S^{-1}, each species's contribution to a given function is small in comparison to that of the rest of the community, in accordance with our focus on collective functions – which cannot be achieved by one species alone. Similarly, the total impact of all functions on a species' growth rate is assumed to be a linear combination of the functional magnitudes, i.e. $\sum_\lambda u_i^{(\lambda)} f_\lambda(t)$. Our analytical approach, however, can be applied to more complicated dependencies, including non-linearities.

The structure bestowed on the GLVs by the existence of collective functions is therefore determined by the *structural matrix* (Figure 2, b)

$$\mu_{ij} = \frac{1}{S} \sum_\lambda u_i^{(\lambda)} v_j^{(\lambda)} \tag{3}$$

where the interaction between any two species is based on their functional traits. Conversely, any interaction matrix with collective interactions can be written in this way via its Singular Value Decomposition. Even though the functional traits obtained this way need not correspond to known ecological processes, the formal description is the same.

Figure 2 a and b illustrate how the structural matrix is derived from species functional traits in the case where there are three equally sized functional groups (panel a in Fig. 1). Species that belong to group λ have the same effect trait, i.e. $v_i^{(\lambda)} = 1$, whereas all other species do not contribute to that function, hence $v_i^{(\lambda)} = 0$. If species within a functional group also have the same response traits (Fig. 2 a), then the structural matrix is composed of $n_F \times n_F$ homogeneous blocks, as represented by the nine squares of Fig. 2 b.

C. Combining structure and randomness

Structure being thus defined, we now turn to the unstructured part of the interactions, reflecting sources of between-species variation too intricate to be described by simple community functions (Fig. 2 c). These non-observable features are modelled as random terms [1, 25, 31], and are defined as a *random matrix* of entries σz_{ij} (Fig. 2 d). This reflects the idea that, lacking direct knowledge, the most parsimonious way to account for the unstructured part of interactions is to chose them randomly. For simplicity, the z_{ij}'s are chosen as independent, identically distributed variables with zero mean and unit variance. The specific details of the probability distribution do not matter, as long as it has finite variance and doesn't depend on S.

The *total interaction matrix* A_{ij} (illustrated in Fig. 2) is defined as the sum of the structured and of the disordered components:

$$A_{ij} = \mu_{ij} + \frac{\sigma}{\sqrt{S}} z_{ij}. \tag{4}$$

The $S^{-1/2}$ scaling factor plays the same role as the S^{-1} factor in eq. (2), ensuring that single interaction terms are small compared to the aggregated effect of the whole community. By setting the exponent to 1/2, structure and randomness have the same overall importance for the dynamics, as this scaling takes into account cancellations between positive and negative random coefficients.

The parameter σ balances the relative importance of structure and randomnes. Without structure, eq. (1) reduces to the classical disordered generalized LVEs studied in statistical physics. Analogous to the latter, we will consider typical ecosystem outcomes, i.e. those that do not depend on the specific realization of randomness. The expectation is that, when structure is added, the statistical properties of such randomness should be sufficient to characterize behaviours that are independent of the unknown species-level details, and that, in the spirit of macroecology, such generic predictions can be connected with recurrent empirical patterns.

III. RESULTS

In the following section III A, we obtain a coarse-grained description of the properties of a community by applying Dynamical Mean Field Theory (DMFT, [3, 24, 32]) to eq. (1) with interactions that combine structure and randomness. We then use it in section III B to draw conclusions on how these two components combine at equilibrium. We first derive a set of closed equations for community-level degrees of freedom, that determine both community- and species-level states. We then illustrate their consequences by focusing on Species Abundance Distributions, a pattern classically studied in theoretical as well as applied ecology [33]. Finally, in

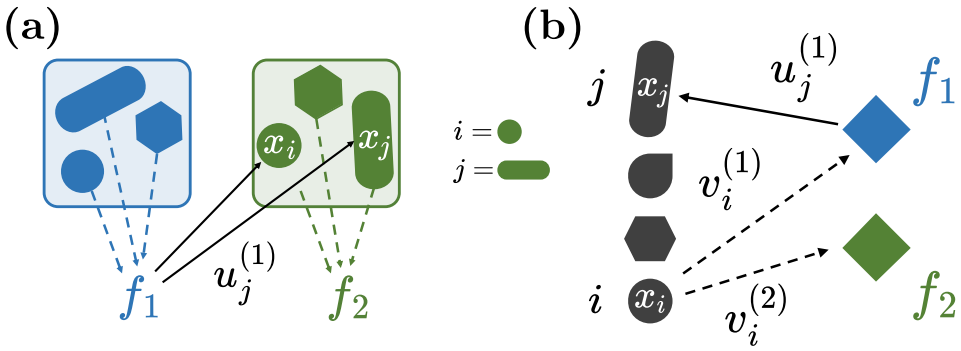

FIG. 1. **Examples of ecosystem structure and their formalization.** **(a)** The community is divided in two functional groups, whose total abundances f_1, f_2 are the relevant collective functions. Species in group $\lambda \in \{1, 2\}$ contribute to the abundance of their group with equal effect traits, $v_i^{(\lambda)} = 1$. Conversely, a group acts upon the rest of the community through its total biomass f_λ, whose effect on any species is weighted by the response traits $u_i^{(\lambda)}$ (shown here only for the effect of the blue group on two species of the green group). In this formalization, disordered gLVEs [3] – where the total biomass is the only collective function – corresponds to the existence of a single group. **(b)** Interactions between species are mediated by two resources of concentrations f_1, f_2, which provide the collective functions. Each species affects the function by either consuming the resource or producing a resource, as encoded by the effect traits $v^{(\lambda)}$ (dotted lines, shown only for species i). Conversely, it is affected by these resources through the response traits $u^{(\lambda)}$ (shown only for species j).

section III C we discuss how such community equilibrium can lose stability, and the respective roles of randomness and structure in out-of-equilibrium species coexistence. For the sake of readability, the main text only presents the main mathematical keystones, while the analytical details can be found in Supplementary Information (SI).

A. Effective community dynamics reflects both structure and disorder

In the absence of randomness, $\sigma = 0$, the Lotka-Volterra system's behaviour is entirely determined by the structured relations between species. We can rewrite eq. (1) as

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left[1 - x_i + \sum_\lambda u_i^{(\lambda)} f_\lambda(t) \right] + m \tag{5}$$

and thus express the dynamics in terms of the functional magnitudes f_λ. Such reduction to a small number of effective variables is implicit in many low-complexity models in ecology, where a number of species (e.g. 'plants', 'herbivores' and 'carnivores', 'phytoplankton' and 'zooplankton') are treated in bulk and considered equivalent within classes [34]. After a transient, coarse-grained variables then follow ODEs that are parametrized at the collective level in terms of group-level interactions, typically much fewer than if one had to account for every species within functional classes.

Once randomness is introduced, coarse-graining is less straightforward. DMFT provides an appropriate framework for studying aggregated properties that are shared by typical realizations of the disordered matrix [3]. This approach describes aggregated, *community-level* variables, usually called "self-averaging", that are the same for any (large) community whose fixed set of interactions is randomly sampled from a distribution with given variance. Contrary to such coarse-grained quantities, *microscopic variables* such as the abundance of a single species generally change from one to another realization (sampling) of randomness.

DMFT can be generalized to encompass structure and disorder in the interactions matrix eq. (4). We summarize in the following the essential steps of the derivation of such Structured DMFT, and leave the details for SI Sec. S3. A key feature is that, due to the large number of species, a formal equivalence can be established between the original dynamical equations with random interaction coefficients and a set of S uncoupled Stochastic Differential Equations (SDE)

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left[1 - x_i + \sum_\lambda u_i^{(\lambda)} f_\lambda(t) + \sigma \zeta_i(t) \right] + m \tag{6}$$

where the dynamics of species i depends, via the species-specific response traits, on the magnitude of functions f_λ and on a Gaussian colored noise ζ_i [1]. The noise term in eq. (6) depends on the abundances of all species in the community through the relation $\langle \zeta_i(t)\zeta_j(t') \rangle = \delta_{ij} C(t, t')$, where $C(t, t')$ is the two-time average of abun-

[1] The most significant difference with prior studies is that, while DMFT in the usual setting yields a low-dimensional model, i.e. a single equation representing the distribution of possible trajectories of any species picked at random in the system, here equations eq. (6) are still as numerous as the initial model, because species are intrinsically different from each other due to their distinct u's and v's; what has been gained is that the equations are now uncoupled.

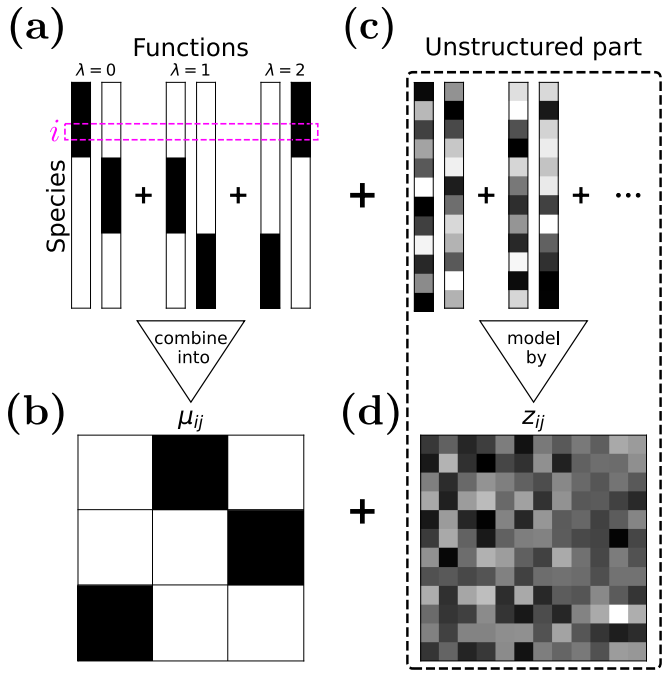

FIG. 2. **Example of interaction matrix that superposes structure and randomness.** The community composed of three homogeneous functional groups, so that species traits $u_i^{(\lambda)}$ and $v_i^{(\lambda)}$ ($\lambda \in \{0,1,2\}$) take the same value within a group **(a)**. The resulting structural matrix μ_{ij} **(b)** is low-rank, its entries are constant by block. Differences between species may stem from other traits that are to fine-grained to be known **(c)**. We model them by adding a random matrix z_{ij} **(d)**, weighted by the relative intensity of the random component.

dance, or *community correlator*,

$$C(t,t') = \overline{x_i(t)x_i(t')} = \overline{\langle x_i(t)x_i(t')\rangle}. \tag{7}$$

Here, brackets indicate averages over a single species in multiple realizations of the disorder matrix, while the overline was defined in eq. (2) as an average over species for a given realization. Owing to the law of large numbers, averages over species equal averages over species and randomness. This implies that, as for the ordinary DMFT, the correlator doesn't depend on the matrix realization.

Similarly, for a fixed value of randomness intensity σ, functional magnitudes are self-averaging and can thus be directly computed from the abundances averaged over realizations of the disorder:

$$f_\lambda(t) = \overline{v^{(\lambda)}\,\langle x(t)\rangle}. \tag{8}$$

Such averaging makes them independent of species-specific details represented by the random contributions z_{ij}, while they retain the structural features through the effect traits $v_i^{(\lambda)}$.

Due to the non-linear nature of eq. (1), the simplicity of the equations does not entail simplicity in the dynamics. Every species is influenced by $n_F + 1$ community-level, self-averaging quantities: the functional magnitudes eq. (8) and the correlator eq. (7). Just as the functions are sufficient to capture the effect of structure in the absence of disorder, the effects of disorder are summarized by the sole correlator and are not species-specific. The statistical equivalence between the original dynamics and trajectories obtained from the coarse-grained equations hinges upon such community-level observables being self-consistent, that is matching the microscopic, species-level stochastic processes. Their computation can in principle be achieved numerically [24], as in the unstructured case. The algorithm however would have important computational costs, exacerbated in our case by the fact that different species have different statistics. We therefore concentrate here on exploiting analytical results that can be obtained at equilibrium, and use them to gain insight also on the onset out-of-equilibrium regimes.

B. A closed set of community-level variables determines equilibrium patterns

Macroscopic quantities are easily obtained by aggregating microscopic ones, while precisely deducing microscopic variables from macroscopic ones is in general impossible. However, a sufficient description by community-level statistics, where the dynamics of coarse-grained observables is independent of species-specific details, can be attained when the community is at equilibrium, as we explain next.

At the species level, the stationary state is defined by the equilibrium species abundances $x_i^\star = \lim_{t\to\infty} x_i(t)$. We will specifically discuss the Species Abundace Distribution (SAD), *i.e.* the fraction of species of a given abundance (class) as a function of abundance $x_i^\star$. A long tradition in ecology looks for universal patterns in such distributions, and compares them with the predictions of theoretical models that typically encompass randomness but not structure [35–37] (but see [38]).

The n_F functional magnitudes and the correlator correspondingly attain the community-level equilibrium

$$\begin{aligned} f_\lambda^\star &= \lim_{t\to\infty} f_\lambda(t) \\ C^\star &= \lim_{t\to\infty} C(t,t). \end{aligned} \tag{9}$$

For convenience, we define the *modal abundance* of species i as follows:

$$x_i^+ = 1 + \sum_\lambda u_i^{(\lambda)} f_\lambda^\star. \tag{10}$$

As discussed in Section III A, unlike the species-specific equilibrium abundance, the magnitudes of ecological functions are self-averaging quantities. Therefore, the modal x_i^+ are independent of the realization of randomness. Moreover, let us introduce the parameter

$$\gamma = \sigma\sqrt{C^\star} \tag{11}$$

that increases with the intensity of disorder and with the correlator – another self-averaging quantity.

When interactions are dominated by structure ($\sigma = 0$), the equilibrium values of species abundances coincide with the modal abundance, and are readily obtained from eq. (5) as $x_i^\star|_{\sigma=0} = \max\left(0, x_i^+\right)$. Equilibrium functional magnitudes then obey the closed set of n_F equations:

$$f_\lambda^\star|_{\sigma=0} = \frac{1}{S}\sum_i v_i^{(\lambda)} \max\left(0, 1 + \sum_\mu u_i^{(\mu)} f_\mu^\star\right) \tag{12}$$

that only involve structural traits. The equations above are strongly nonlinear at $x_i^+ = 0$, when species go extinct. Therefore, they generally admit multiple equilibrium solutions, which may differ in abundance as well as stability. The distribution of species abundances in the absence of randomness can be easily visualized when the community is structured in homogeneous functional groups (Figs. 1 a and 2). Every species belonging to the same group has the same effect traits $u_j^{(\lambda)}$, therefore exactly the same equilibrium abundance, which corresponds simply to the solution of the low-dimensional Lotka-Volterra equations for the functional groups. In Figure 3 a we illustrate the SAD for a community structured in four groups, where abundances of extant species concentrate over just three modal values, with relative frequency equal to the fraction of species in each group. Species with negative modal abundance go extinct.

The presence of unstructured variation modifies this solution, and species will have a different equilibrium abundance even if they share the same structural traits. The effect of increasing the variance σ is to spread the abundances around the modal abundance. Species that share the same $u_j^{(\lambda)}$ will have the same x_i^+, consistent with the idea that they respond coherently to the function λ, despite individual heterogeneity. More precisely, in SI Sec. S5 we show that, when $\sigma > 0$, the equilibrium abundance of species i is randomly distributed according to a truncated-Gaussian random variable,

$$x_i^\star = \max\left(0, x_i^+ + \gamma\xi_i^\star\right), \tag{13}$$

where $\xi_i^\star$ are independent standard Gaussian variables that reflect the diversity of traits among different species.

Remarkably, moreover, community-level variables are related by a closed set of $n_F + 1$ relations[2]

$$\begin{aligned} f_\lambda^\star &= \frac{\gamma}{S}\sum_i v_i^{(\lambda)}\, \omega_1\left(\frac{x_i^+}{\gamma}\right) \\ 1 &= \frac{\sigma^2}{S}\sum_i \omega_2\left(\frac{x_i^+}{\gamma}\right) \end{aligned} \tag{14}$$

[2] In the case without disorder, eq. (12), the interaction matrix A_{ij} was low-rank, allowing direct coarse-graining, whereas here a low-dimensional description exists despite (A_{ij}) being full-rank. Despite accounting for most of the dimensionality of the matrix, random interactions thus modify the system in a minimal way, by adding a single macroscopic relation on top of those provided by structure.

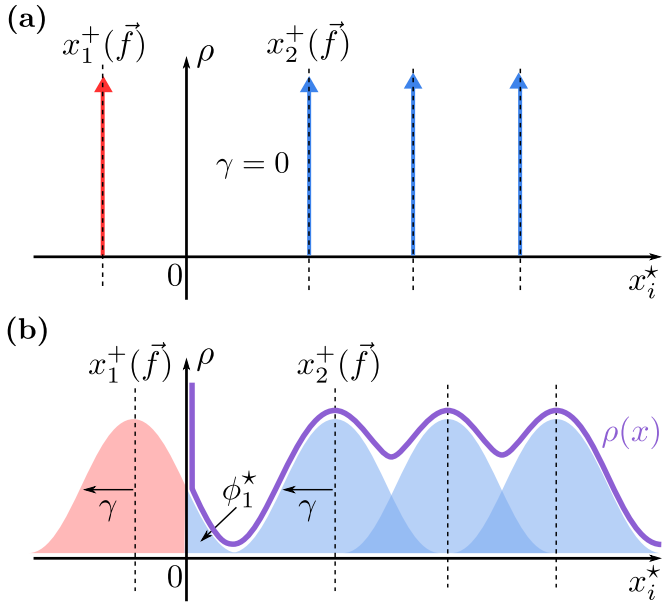

FIG. 3. **Schematic representation of the Species Abundance Distribution (SAD) for functionally structured communities.** The SAD eq. (16) results from the overlap of structure – as defined by a small number of functions – and randomness – quantified by the standard deviation γ. **(a)** In the absence of randomness ($\gamma = 0$), abundances of species within functional groups coincide with those of the low-dimensional Lotka-Volterra equations where within-group abundances are summed up. **(b)** When the x_i^+s are sufficiently separated compared to γ, the distribution is multimodal. The values x_i^+ the group abundance distributions are centered on are in general different with respect to those with pure structure.

where we have defined the mathematical functions

$$\omega_k(w) = (2\pi)^{-1/2}\int_{-w}^{\infty} \mathrm{e}^{-z^2/2}\,(w+z)^k\,\mathrm{d}z. \tag{15}$$

We can now exploit the self-consistent relations eq. (14) to derive the exact distribution of the microscopic abundances. The SAD turns out to be a convolution of a 'structure-driven SAD', reflecting the deterministic biases given by the modal abundances x_i^+, and of a Gaussian distribution of typical width γ (Fig. 3 b):

$$\rho(x) = (1-\phi^\star)\,\delta(x) + \frac{\Theta(x)}{\gamma}\sum_i \psi\left(\frac{x - x_i^+}{\gamma}\right). \tag{16}$$

Here, $\psi(w) = \exp(-w^2/2)/\sqrt{2\pi}$ is the Probability Density Function of a standard Gaussian, Θ is the Heaviside function and $\phi^\star$ is the *diversity*, that is, the fraction of extant species.

In Fig. 4 a and b we provide an example, considering a simulated community composed of two functional groups: each group contains a different number of species, and all species within one group share the same structural traits (structural interaction matrices are represented in the insets). The SAD is trivial in the absence of randomness,

as one group excludes the other at equilibrium (Fig. 4 a). When random variation is added, the excluded group is rescued, and the two peaks widen until they largely overlap (Fig. 4 b). Yet, the existence of an underlying structure remains detectable if the mean abundance of each group is plotted (solid vertical line). The structure-driven SAD can also feature more complex profiles that reflect the distribution of traits within the community or the functional groups. For instance, if the total abundance is the only community function and the response traits are distributed as a power-law, the structure-driven SAD will have a power-law decay (Fig. 5 a). This progressively turns into an Gaussian decay as randomness is increased (Fig. 5 b).
The SAD is therefore very sensitive to the presence of structure in the interactions. Its exact shape is dictated by the interplay of structure and randomness, and it is not expected to follow universal rules, other than those possibly imposed by universal scalings of species traits.

Even though increasing randomness blurs the SAD, the effect of structure never vanishes in the self-consistent relations eq. (14). These can be rescaled into $g_\lambda^\star = S^{-1}\sum_i v_i^{(\lambda)}\ \omega_1\Big(1/\gamma + \sum_\lambda u_i^{(\lambda)} g_\lambda^\star\Big)$, where $g_\lambda = f_\lambda/\gamma$. When γ is large and increases, so must do the functions.

Drops in functional magnitude can however occur at intermediate levels of randomness, as in the example of Fig. 4 c.
Increasing randomness also affects the number and nature of surviving species. According to eq. (15), $\omega_0(x_i^+/\gamma) = \phi_i^\star$ represents the probability, measured over realizations of disorder, that species i survives. For $\sigma > 0$ this is different from zero and one. Hence, species that would go extinct based solely on structure can be 'rescued' by randomness and hence contribute to functions, leading to non-trivial effects even in the simple case of two functional groups illustrated in Figure 4. A group that initially has very low abundance can take advantage of randomness, and drive a community-wide increase in diversity (Figure 4 d). This increases the competitive pressure on the group that would be expected to dominate, causing some of its species to go extinct. Eventually, the functional dominance in the community can get reversed despite the general positive effect of randomness on functional magnitude (Fig. 4 c). Such effect is the consequence of extinctions biasing interactions towards mutualism [39].

The diversity of the community

$$\phi^\star = \overline{\omega_0\left(\frac{x_i^+}{\gamma}\right)} \tag{17}$$

generally cannot be predicted from the diversity in the purely structured limit, which suggests it is a poor indicator of how community function may change under different sources of species-level variation.

In summary, the addition of randomness preserves the existence of deterministic relations between macroscopic functions, but it changes those functions through extinctions and rescue.

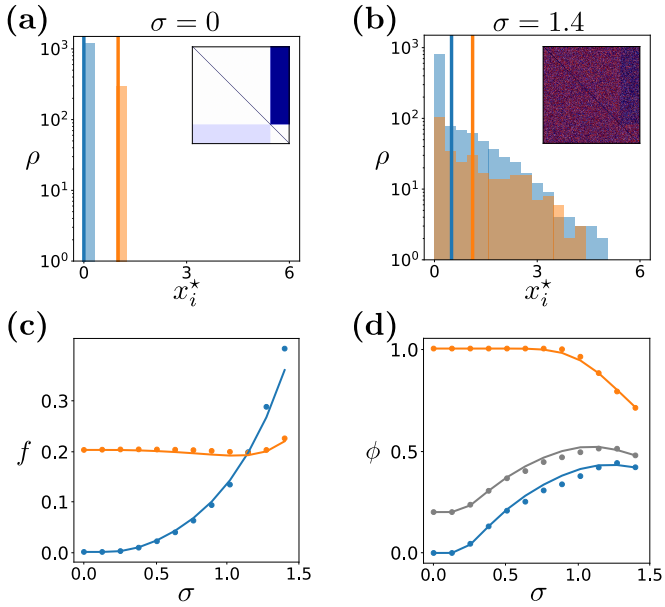

FIG. 4. **Species abundance distributions and community-level observables at equilibrium**. SADs for a community structured in two functional groups, as illustrated by the interaction matrix in the inset, in the absence **(a)** and in the presence of random variation **(b)**. The SADs for the two groups are separately plotted. Their mean abundances (solid vertical lines) highlight that the signature of structure is still visible when randomness is so strong that the total SAD would be featureless. **(c)** The functional magnitudes change with the intensity σ of random variation (dots, numerical simulations; solid lines, theoretical prediction from eqs. (14)). Different functional groups can witness opposite changes in the value of their defining function, but when randomness is strong enough, they will all see an increase in function due to the asymmetry induced by constraining abundances to be positive. In this case, the total abundance of the class increases much faster for species that were rare in the absence of randomness. **(d)** The number of surviving species (as predicted by eq. (15), solid lines vs simulations, dots) for both groups as well as for the whole community (grey). By rescuing the blue group from extinction, randomness initially increases the survival fraction. For large σ, however, the effect of extinctions dominates and the survival fraction decreases, despite the increase in total abundance.

C. Randomness can drive oscillating communities to stable equilibrium coexistence

The previous analytical results rely on the existence of a stable fixed point. Equation (1) with structured or random interactions is however known to also display out-of equilibrium dynamics, including limit cycles, neutral cycles, and chaotic behaviour [40]. In this section, we explore the effect of introducing random variability between strains in a model describing the ecology of plankton species that is known to exhibit chaotic dynamics [41].

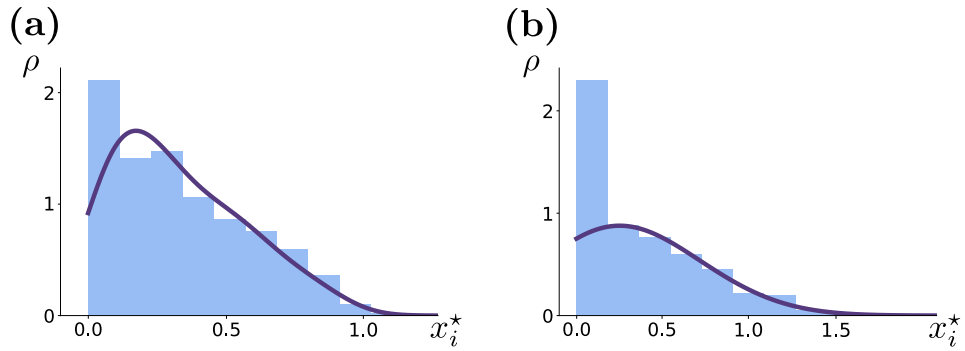

FIG. 5. **Species abundance distributions for communities with a power-law distributed structural trait**. Comparison of the SAD of a simulated community with $S = 1000$ (histograms) and the theoretical predictions of eq. (13) (lines) for two different levels of random variation in the interactions (**(a)** $\sigma = 0.25$ and **(b)** $\sigma = 0.75$). Species abundances are affected by a single function: the total biomass (meaning that they all have the same effect trait $v_j^{(\lambda)} = 1$). However, the response trait $u_i^{(\lambda)}$ of a species i is drawn from a power-law distribution with exponent $\alpha = 2$ on the interval $[0, 3]$. The SAD that reflects the power-law distribution of the traits, visible for small σ, turns into a truncated Gaussian as the magnitude of disorder increases, progressively erasing the structural features. The shape of the SAD is well predicted from the sole knowledge of α, σ, without knowledge of the traits of each species in the community.

There, a combination of strong interactions and weak immigration causes a persistent turnover between a small set of dominant species and a large set of rare species, thus reproducing dynamic and static patterns that are characteristic of plankton communities when observed at a high taxonomic resolution [42, 43].

We ask whether this behavior changes when one accounts for the possible uncharacterized variation of interactions between strains that belong to a same species. The recent availability of data at such a taxonomic resolution, as well as theoretical investigations, indeed indicate that, analogous to functional groups, species may harbour a degree of unknown genetic and trait variation, with dynamical consequences [43–46].

In keeping with previous notations, we consider a set of n_F species and divide each of them into s strains, for a total of $S = s \times n_F$ strains that constitute the microscopic variables of our model. Figure 6 displays the asymptotic regimes of a community with fixed structural interactions and for increasing values of the intensity of within-species variability. When $\sigma = 0$, all strains within a species have the same chaotic trajectory as the species they belong to (Fig. 6 a). When the intensity σ of random variability increases, strains remain largely synchronized within species, so that the two levels of description have the same qualitative dynamics. Their trajectories get however progressively modified and the community crosses different dynamical regimes (Fig. 6 b), until it reaches a stable equilibrium (Fig. 6 c) where numerous species coexist even in the absence of immigration. For even larger σ, the strains start to oscillate again chaotically, but are desynchronized. At the level of species, hence, fluctuations cancel out (Fig. 6 d). these observations suggest that microscopic variability has the effect of stabilizing the dynamics at the species level while destabilizing them at the strain level. These two effects stem from two different ways in which randomness can modify the Jacobian at an equilibrium.

The strain-level instability occurring for strong randomness is akin to that characterized in gLVEs with purely random interactions [3]. Such *disorder-driven* collective phase transition features a concomitant bulk instability of the Jacobian matrix and of the reduced interaction matrix (where only extant species are considered). The dynamics thus looses stability when

$$\phi^\star \sigma^2 - 1 = 0, \tag{18}$$

where $\phi^\star$ is the fraction of surviving species, as defined in eq. (17). This expression is the same as for purely random systems, however here $\phi^\star$ depends on σ, so that structure affects the location of the transition. For the community we simulated, disorder-driven microscopic chaos is predicted for $\sigma \gtrsim 2.25$.

Counter the intuition built on species-rich community models [1, 3], randomness can also have a stabilizing effect at the species level, by two means. First, randomness-induced strain extinctions may be enough to reduce the number of effectively interacting strains so much that the systems stops oscillating. Second, and less trivially, *structure-driven* transitions to equilibrium can occur as bifurcations of community-level aggregated variables. Instead of being related to the bulk of the Jacobian matrix, these bifurcations involve its outliers. Directly derived from the low-rank community structure bestowed by functions, they are modified by randomness. We show in SI Sec. 7 that structure-driven transitions can be located using the 'pseudo-Jacobian' matrix

$$\mathcal{J}_{ij} = -x_i^\star \left(\delta_{ij} - \mu_{ij}\right), \tag{19}$$

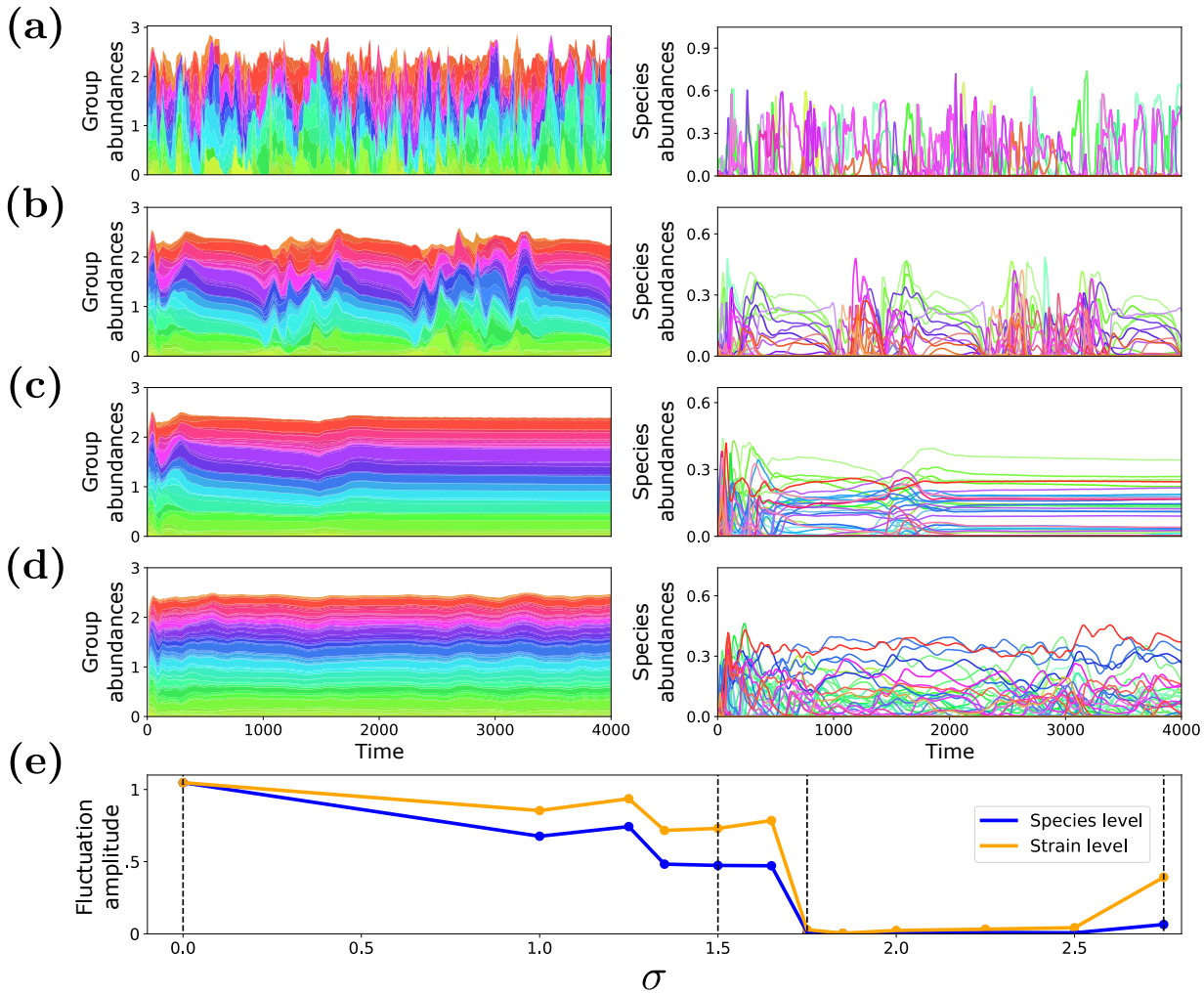

FIG. 6. **Influence of strain variability on species and strain dynamics.** A matrix of species-species interactions is chosen as in [41], with $n_F = 150$ species. Each species contains $s = 40$ strains, for a total $S = 6000$. The structural part of strain interactions is that of the species they belong to, while instraspecific variability between strains is random and of intensity σ. *(a-d)* The variation in time of community composition at the species level (left) and at the strain level (right, three strains per species) is plotted for $\sigma = 0, 1, 1.75$ and 2.75 (represented by dotted lines in **(e)**). When strains within a species are identical, they synchronize on a chaotic trajectory **(a)**. As σ is increased, a series of structure-driven transitions occur that similarly affect both species and strains. Starting with the fastest ones, oscillatory modes progressively stabilize, slowing down the dynamics **(b)** and eventually stabilizing it, until the system reaches a fixed point both at the species and the strain level **(c)**. Increasing σ further, a disorder-driven transition is crossed, after which strain trajectories oscillate in an asynchronous manner, resulting in negligible oscillations at the species level **(d)**. **(e)** The average amplitude of the oscillations at species (blue) and strain (yellow) level illustrates the community stabilization via structure-driven bifurcations and then disorder-driven decoupling of strain and species dynamics.

which explicitly depends on the structural part of the interactions, but not on randomness. Randomness, however, impacts $\mathcal{J}$ indirectly through the equilibrium abundances $x_i^\star$ Equation (13).
As σ becomes larger, the diagonal matrix in Equation (19) admits a set of increasingly different positive values. Eigenvalues with a non-zero imaginary part and a positive real part, *i.e.* oscillating dynamical modes, will see their real part decrease and possibly vanish.

Whereas eq. (18) is satisfied when an extensive number of bulk modes diverge at once, here the crossing of a single pair of conjugate outliers generically leads to a low-codimensional bifurcation with a two-dimensional neutral manifold, *i.e.* a Hopf bifurcation. For a deeper mathematical analysis of this phenomenon, the reader is directed to our concurrent paper [47].
The interplay between the abundances at equilibrium and the outliers of the structural interaction matrix can thus result into the stabilization of the community dynamics, a phenomenon that transcends both purely random and purely deterministic models.

IV. DISCUSSION

Models of ecological dynamics with random species interactions are increasingly explored as a paradigm for simple and parsimonious explanations of species abundances and dynamics [4]. Yet, we do not know to which extent their predictions are robust to the presence of structure in species interactions, such as the existence of functional groups or other macroscopic features, evidenced in many empirical interaction networks and also exploited to explain community-level patterns [7]. In particular, it is important to ascertain whether including both randomness and structure in species interactions may still allow for simple predictions, or whether we should expect the increased complexity to severely hamper generalization.

As a first step toward understanding the possible interplay of randomness and structure, we have considered a model for community dynamics where we know that parsimonious predictions are possible both when interactions are purely structured and when they are randomly assigned. When structure is the only determinant, simplicity arises from the existence of a limited number of macroscopic observables, which represent broad ecological functions of the community and control how every species affects and responds to others. On the other hand, species-rich communities with purely random interactions display statistical regularities captured by collective-level degrees of freedom, which aggregate microscopic heterogeneity.

In general, however, structuring processes play out in parallel to a vast number of other, unknown, mechanisms. Modelling them by the superposition of independent random fluctuations to structured species interactions, we derived a single effective description of the community dynamics that integrates the two extreme low-dimensional representations. We then investigated two main aspects of the interplay between randomness and structure, associated to different mathematical properties of the dynamical model: first, static patterns of species abundances (SADs) and macroscopic variables at equilibrium; second, transitions to other dynamical regimes, characterized by spectral properties of the interaction and Jacobian matrices.

The equilibrium state of a structured community is described at a macroscopic level in terms of community-level, coarse-grained variables rather than individual species abundances, like the fully random case. This description takes the form of a set of relations, eqs. (14), which involve the ecological functions that structure the interactions and a single additional degree of freedom, summarizing random variation.

The effect of randomness on macroscopic ecological functions is nonetheless better understood by looking, at the microscopic level, at the abundance of single species. Predictions from purely random models for Species Abundance Distributions are readily altered when structure is present. This is consistent with prior findings that a small, low-rank perturbation of the interaction matrix is sufficient to modulate and fully determine species abundances [5, 48]. In particular, features observed in the SAD reflect the distribution of traits that make up the structural matrix. For instance, if response traits are distributed as a power-law, as in Fig. 5 (see also [49]), then the SAD will have a power law tail. On the other hand, the mapping from traits to equilibrium abundances is non-invertible: many different structural matrices may produce the same SAD. These results indicate that general functional classes of SADs, may reflect, if anything, statistical regularities of the relation between species traits and ecosystem functions, rather than being universal patterns. Alternatively, the universality of patterns may indicate a limited role for interactions, structured or not, in determining community dynamics [37].

Randomness causes species abundances to spread normally around their values predicted from the coarse-grained equations. These randomness-induced variations in abundance have four main consequences: in the SAD, they smooth out the structural patterns, hence reducing the importance of finely-varying structural traits; they allow some species to become highly dominant and magnify any ecological function these contribute to (thus typically favoring functions that receive contributions from more species); they moderate the impact of species extinctions, as any species has a finite probability of surviving for one realization of disorder; and finally, they induce a bias toward overall less competitive interactions among survivors at equilibrium, since species experiencing more negative effects are more likely to go extinct due to disorder – consistent with previous work on randomly assembled communities [39].

As for dynamics, randomness can qualitatively modify the community dynamics dictated by structure in essentially two independent ways, as revealed by analysing the stability of equilibria. First, randomness can induce a phase transition that decouples microscopic (oscillating) and macroscopic (close-to-equilibrium) degrees of freedom, similar to purely random models [3, 24]. This is associated to the destabilization of the bulk of eigenvalues, that is robust to the addition of structure.

Second, in virtue of its effects on species abundances, randomness can modify the community dynamics induced by structure by displacing the outlier eigenvalues associated to the collective modes. Abundances thus play a nontrivial role in determining the stability of the community attractors and mediate between the microscopic heterogeneity in time scales and that that emerges at the community level. Being an essential feature of ecological models based on per-capita growth rates, microscopic time scales are increasingly invoked as key ingredients underpinning patterns of commonness and rarity and stability [50] of communities. Moreover, their mismatch in intrinsically out-of-equilibrium systems such as populations of globally coupled oscillators induces a synchronous gradual unfolding at the microscopic and macroscopic scale [51], suggesting a possible

analogy between randomness-induced stabilization and the phenomenon of amplitude death.

While simple structure and pure randomness have emerged as contrasting paradigms, both promising understanding and prediction of broad phenomenological features from limited information, we believe that current developments call for a synthesis of these two approaches. In this work, we have given a theoretical view on what can generically be expected under several simplifying assumptions. Some of them, such as the homogeneity of single-species parameters (the net growth rate r and carrying capacity K) are easily lifted, whereas others, such as correlations in randomness, pose greater theoretical challenges. Our choice to disregard correlations reflects the idea that species symmetries are already represented in the structural matrix, and that such information would anyway be unavailable for the hidden processes. One path towards generalizing our approach is to include correlations between transpose elements A_{ij} and A_{ji} [25], therefore setting the proportion of symmetric and anti-symmetric random interactions. Perhaps progress could also be made for more complicated correlations, as those addressed in random matrix theory [52] and with gLVEs [53].

Finally, this work has focused on theoretical principles, but provides some hints on possible signatures of the presence of hidden interactions in data. For instance, communities structured in functional groups may be compared across environments that preserve group-level interactions. This may teach us the role of hidden variability and whether we should trust qualitative and quantitative conclusions of models where the natural complexity is drastically reduced.

ACKNOWLEDGMENTS

The authors are very grateful to Jean-François Arnoldi for his participation in the initial stages of the project, and Emil Mallmin for fruitful discussions and for sharing his simulation code. This research was co-funded by the European Union (GA#101059915 - BIOcean5D) and the EABIS Graduate Program in Life Sciences and Biodiversity. JGM acknowledges the support of the Frontiers in Research and Education program.

[1] R. M. May, Nature **238**, 413 (1972).
[2] Y. Yoshino, T. Galla, and K. Tokita, Journal of Statistical Mechanics: Theory and Experiment **2007**, P09003 (2007).
[3] G. Bunin, Phys. Rev. E **95**, 042414 (2017).
[4] M. Barbier, J.-F. Arnoldi, G. Bunin, and M. Loreau, Proceedings of the National Academy of Sciences **115**, 2156 (2018), https://www.pnas.org/doi/pdf/10.1073/pnas.1710352115.
[5] M. Barbier, C. De Mazancourt, M. Loreau, and G. Bunin, Physical Review X **11**, 011009 (2021).
[6] J. Hu, D. R. Amor, M. Barbier, G. Bunin, and J. Gore, Science **378**, 85 (2022), https://www.science.org/doi/pdf/10.1126/science.abm7841.
[7] J. E. Goldford, N. Lu, D. Bajić, S. Estrela, M. Tikhonov, A. Sanchez-Gorostiaga, D. Segrè, P. Mehta, and A. Sanchez, Science **361**, 469 (2018), https://www.science.org/doi/pdf/10.1126/science.aat1168.
[8] C. M. Mutshinda, Z. V. Finkel, C. E. Widdicombe, and A. J. Irwin, Functional Ecology **30**, 1714 (2016), https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2435.12641.
[9] S. Louca, S. M. S. Jacques, A. P. F. Pires, J. S. Leal, D. S. Srivastava, L. W. Parfrey, V. F. Farjalla, and M. Doebeli, Nature Ecology & Evolution **1**, 0015 (2016).
[10] S. Louca, L. W. Parfrey, and M. Doebeli, Science **353**, 1272 (2016), https://www.science.org/doi/pdf/10.1126/science.aaf4507.
[11] Q. Zeng, A. Lebreton, L. Auer, X. Man, L. Jia, G. Wang, S. Gong, V. Lombard, M. Buée, G. Wu, Y. Dai, Z. Yang, and F. M. Martin, Microbiome **11**, 217 (2023).
[12] S. Allesina, J. Grilli, G. Barabás, S. Tang, J. Aljadeff, and A. Maritan, Nature communications **6**, 7842 (2015).
[13] D. Gravel, F. Massol, and M. A. Leibold, Nature communications **7**, 12457 (2016).
[14] P. V. Aceituno, T. Rogers, and H. Schomerus, Physical Review E **100**, 010302 (2019).
[15] L. Poley, J. W. Baron, and T. Galla, Physical Review E **107**, 024313 (2023).
[16] J. I. Park, D.-S. Lee, S. H. Lee, and H. J. Park, Incorporating heterogeneous interactions for ecological biodiversity (2024), arXiv:2403.15730 [physics.bio-ph].
[17] S. Marcus, A. M. Turner, and G. Bunin, Local and extensive fluctuations in sparsely-interacting ecological communities (2023), arXiv:2308.01828 [q-bio.PE].
[18] M. Tikhonov and R. Monasson, Physical review letters **118**, 048103 (2017).
[19] M. Advani, G. Bunin, and P. Mehta, Journal of Statistical Mechanics: Theory and Experiment **2018**, 033406 (2018).
[20] Z. Feng, R. Marsland III, J. W. Rocks, and P. Mehta, PLOS Computational Biology **20**, e1011675 (2024).
[21] S. Louca, M. F. Polz, F. Mazel, M. B. N. Albright, J. A. Huber, M. I. O'Connor, M. Ackermann, A. S. Hahn, D. S. Srivastava, S. A. Crowe, M. Doebeli, and L. W. Parfrey, Nature Ecology & Evolution **2**, 936 (2018).
[22] X. Shan, A. Goyal, R. Gregor, and O. X. Cordero, Nature Ecology & Evolution **7**, 716 (2023).
[23] A. Skwara, K. Gowda, M. Yousef, J. Diaz-Colunga, A. S. Raman, A. Sanchez, M. Tikhonov, and S. Kuehn, Nature Ecology & Evolution **7**, 1823 (2023).
[24] F. Roy, G. Biroli, G. Bunin, and C. Cammarota, Journal of Physics A: Mathematical and Theoretical **52**, 484001 (2019).
[25] T. Galla, Europhysics Letters **123**, 48004 (2018).
[26] A. Altieri, F. Roy, C. Cammarota, and G. Biroli, Phys.

Rev. Lett. **126**, 258301 (2021).
[27] L. S. Bittleston, M. Gralka, G. E. Leventhal, I. Mizrahi, and O. X. Cordero, Nature Communications **11**, 1440 (2020), number: 1 Publisher: Nature Publishing Group.
[28] P. Chesson, Annual Review of Ecology, Evolution, and Systematics **31**, 343 (2000).
[29] R. MacArthur, Ecology **36**, 533 (1955).
[30] M. S. Advani, G. Bunin, and P. Mehta, Journal of Statistical Mechanics: Theory and Experiment **2018** (2018).
[31] E. R. Garcia, M. J. Crumpton, and T. Galla, Phys. Rev. E **108**, 034120 (2023).
[32] T. Arnoulx de Pirey and G. Bunin, arXiv e-prints , arXiv (2023).
[33] A. E. Magurran, Current Biology **31**, R1174 (2021).
[34] J. Steele and E. Henderson, The American Naturalist **117**, 676 (1981).
[35] B. J. McGill, R. S. Etienne, J. S. Gray, D. Alonso, M. J. Anderson, H. K. Benecha, M. Dornelas, B. J. Enquist, J. L. Green, F. He, A. H. Hurlbert, A. E. Magurran, P. A. Marquet, B. A. Maurer, A. Ostling, C. U. Soykan, K. I. Ugland, and E. P. White, Ecology Letters **10**, 995 (2007), https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1461-0248.2007.01094.x.
[36] T. J. Matthews and R. J. Whittaker, Journal of Applied Ecology **52**, 443 (2015), https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2664.12380.
[37] J. Grilli, Nature Communications **11**, 10.1038/s41467-020-18529-y (2020).
[38] A. E. Magurran and P. A. Henderson, Nature **422**, 714 (2003).
[39] G. Bunin, arXiv preprint arXiv:1607.04734 (2016).
[40] Y. Takeuchi, *Global dynamical properties of Lotka-Volterra systems* (World Scientific, 1996).
[41] E. Mallmin, A. Traulsen, and S. D. Monte, Proceedings of the National Academy of Sciences **121**, e2312822121 (2024), https://www.pnas.org/doi/pdf/10.1073/pnas.2312822121.
[42] E. Ser-Giacomi, L. Zinger, S. Malviya, C. de Vargas, E. Karsenti, C. Bowler, and S. de Monte, Nature Ecology & Evolution **2**, 10.1038/s41559-018-0587-2 (2018).
[43] A. M. Martin-Platero, B. Cleary, K. Kauffman, S. P. Preheim, D. J. McGillicuddy, E. J. Alm, and M. F. Polz, Nature Communications **9**, 10.1038/s41467-017-02571-4 (2018).
[44] N. Kashtan, S. E. Roggensack, S. Rodrigue, J. W. Thompson, S. J. Biller, A. Coe, H. Ding, P. Marttinen, R. R. Malmstrom, R. Stocker, M. J. Follows, R. Stepanauskas, and S. W. Chisholm, Science (New York, N.Y.) **344**, 416 (2014).
[45] M. T. Pearce, A. Agarwala, and D. S. Fisher, Proceedings of the National Academy of Sciences **117**, 14572 (2020), publisher: Proceedings of the National Academy of Sciences.
[46] A. Goyal, L. S. Bittleston, G. E. Leventhal, L. Lu, and O. X. Cordero, eLife **11**, e74987 (2022).
[47] J. Giral Martínez, S. de Monte, and M. Barbier, Stabilization of macroscopic dynamics by fine-grained disorder in many-species ecosystems (2024), arXiv:2411.14969 [q-bio.PE].
[48] J. Fraboul, G. Biroli, and S. De Monte, Journal of Theoretical Biology , 111557 (2023).
[49] D. Gupta, S. Garlaschi, S. Suweis, S. Azaele, and A. Maritan, Phys. Rev. Lett. **127**, 208101 (2021).
[50] J.-F. Arnoldi, M. Loreau, and B. Haegeman, Ecology letters **22**, 1557 (2019).
[51] S. De Monte, F. d'Ovidio, H. Chaté, and E. Mosekilde, Phys. Rev. Lett. **92**, 254101 (2004).
[52] J. W. Baron, T. J. Jewell, C. Ryder, and T. Galla, Phys. Rev. Lett. **128**, 120601 (2022).
[53] J. Giral Martínez, A symmetry-based approach to species-rich ecological communities (2024), arXiv:2407.13444 [q-bio.PE].
[54] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, 1998).

Supplementary Material

A. DERIVATION OF THE DIMENSIONLESS EQUATIONS

Let us consider a community of S interacting species labeled $i = 1, \ldots, S$, of respective abundances $N_i(t)$. The Lotka-Volterra equations are the simplest non-linear model capturing the effect of ecosystem composition on individual demographic rates [54]. They reproduce qualitative features of population dynamics, such as exponential growth or decay, stable survival and extinction. Within this framework, each species is endowed with an intrinsic net growth rate r_i. Species with larger r_i have faster dynamics, and therefore react faster to changes in community composition. The species' carrying capacity K_i is the maximum abundance it can attain in isolation as a result of competition between individuals of the same species.
Interactions with other species are characterized by a matrix of interaction coefficients a_{ij} that quantify the impact of species j on the instantaneous growth rate of species i. In our convention, a negative a_{ij} implies a deleterious effect on growth and thus competition; on the contrary, a positive coefficient indicates a facilitative interaction. The abundance of every species then changes in time according to

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = r_i N_i \left[1 - \frac{N_i - \sum_j a_{ij} N_j}{K_i}\right] + m_i, \tag{S1}$$

Migration into the community has also been added through the *immigration rates* m_i [3, 24]. In numerical simulations, $m_i \ll K_i$. Some species that persist at the immigration level can therefore invade as soon as the community context becomes favorable. At any given time, we count as extinct those species whose abundance is smaller than a multiple of m_i.

In order to focus on the effect of variability in the interaction parameters, we make the simplifying assumption that some parameters are uniform across species and rescale eq. (S1) so that variables become independent of the carrying capacity. The analytical tools however generalize to cases when this transformation is not realized. Assuming that r_i is constant across species, it can be set to one, without loss of generality, by rescaling the time coordinate. We also set $m_i = m$ for all species. Finally, by factoring out K_i, one can rewrite the Generalized Lotka-Volterra Equations (GLVs) in terms of the normalized abundances $x_i = N_i/K_i$ and the rescaled interaction coefficients $A_{ij} = a_{ij}K_j/K_i$ as

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left[1 - x_i + \sum_j A_{ij} x_j\right] + m. \tag{S2}$$

B. PROPERTIES OF THE INTERACTION MATRIX

A. Scaling

We aim to work in a regime where ecological processes stem from aggregating the contribution from many different species. Given the dynamical equation

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i \left(1 - x_i + \sum_j A_{ij} x_j\right), \tag{S3}$$

this means that the term $A\vec{x}$ needs to be $\mathcal{O}(1)$ even in the presence of an extensive number of species that all have abundances $\mathcal{O}(1)$. Given that

$$\mathbb{E}\left[\sum_j A_{ij} x_j\right] \sim \sum_j \mathbb{E}\left[A_{ij}\right] \sim S\mathbb{E}\left[A_{ij}\right] \tag{S4}$$

$$\mathbb{V}\mathrm{ar}\left[\sum_j A_{ij} x_j\right] \sim \sum_j \mathbb{V}\mathrm{ar}\left[A_{ij}\right] \sim S\mathbb{V}\mathrm{ar}\left[A_{ij}\right] \tag{S5}$$

we need both the mean and the variance of the interaction coefficients to go as $1/S$. This leads to the decomposition of the main text,

$$A_{ij} = \mu_{ij} + \frac{\sigma}{\sqrt{S}} z_{ij}, \tag{S6}$$

where $\mu_{ij} \sim 1/S$.

B. Rank and eigenvectors of the structure matrix

For the reasons stated in the main text, the structure matrix is assumed to be low-rank. Although our main justification is based on the notion of macroscopic structure, low-rank is also an automatic consequence of the scaling chosen in eq. (S6). Indeed, we have

$$\mathrm{Tr}\left[\mu^T \mu\right] = \sum_{ij} \mu_{ij}^2 \sim \mathcal{O}(1). \tag{S7}$$

On the other hand this trace is also the sum of the singular values of μ, therefore only a finite number of them can be of order 1, i.e. μ is low-rank. Now we turn to the eigenvectors of μ. First, let $\vec{x}$ be a localized vector. Then

$$[\mu\vec{x}]_i = \sum_{j:x_j \neq 0} \mu_{ij} x_j \ll 1. \tag{S8}$$

Therefore the structure part doesn't act on localized directions. Similarly, assume that the vector is sampled from a uniform distribution on the unit sphere

$$\|\mu\vec{x}\|^2 = \sum_{ij} x_i \mu_{ki} \mu_{kj} x_j = \mathcal{O}(S^{-2}). \tag{S9}$$

Therefore neither does it act on 'typically random' directions. In particular, eigenvectors of μ with non-zero eigenvalue are collective and need to have non-random features in their entries (for instance, having more positive than negative entries, or be structured in groups).
Conversely, assume that the matrix is low-rank and write it using its SVD

$$\mu_{ij} = S^{-1} \sum_{\lambda} \lambda u_i^{(\lambda)} v_j^{(\lambda)}, \tag{S10}$$

where the λ's are the (positive) singular values, which are $\mathcal{O}(1)$ and the $u^{(\lambda)}$ and $v^{(\lambda)}$ are non-normalized. Then given that $u_i^{(\lambda)}, v_j^{(\lambda)} \sim \mathcal{O}(1)$, the scaling of the interaction matrix is $\mathcal{O}(S^{-1})$. Furthermore, if the eigenvectors were randomly distributed then the matrix would only act on the very low-dimensional space spanned by those vectors. In particular, the projection of any fixed vector with positive entries on this low-dimensional is almost surely zero. Given that the interaction matrix is applied to abundance vectors, which have positive entries, such matrices would effectively imply vanishing interactions in the Lotka-Volterra model. Hence eigenvectors of low-rank matrices need to be non-random for them to be meaningful.
Therefore we see that there is a quantitative equivalence between the $1/S$ scaling for the structural part and the fact of having a low rank and non-random eigenvectors. In other words, collective structure and this scaling go hand-in-hand, therefore providing a mathematical justification for our choice. This is the complete opposite of what happens in the random part of the interactions, where the scaling can be upgraded to $1/\sqrt{S}$ precisely because the eigenvectors are random and the rank is full.

C. NUMERICAL SIMULATIONS AND SOLUTIONS

A. Numerical simulations

All numerical calcualtions where performed in Python. Numerical simulations where done using Scipy's `solve_ivp` function with the `RK45` method. Interactions for simulations shown in the main text were drawn from Gaussian distributions although other distributions were also tested. Initial conditions were sampled uniformly between zero

and 1. The solver was stopped after a fixed amount of time, generally $T = 1000$. This is more than enough for convergence in the UFP phase and also enough to observe a large-enough amount of cycles in the cyclic phase. The immigration threshold was set to $\varepsilon = 10^{-8}$.
To classify the numerical simulations between `divergent, ufp, cyclic` and `unstable`, we used several criteria. First, if the simulation stopped before T with the solver throwing an integration error then it was classified as `divergent`. This is because in this phase trajectories diverge in finite time. Otherwise, the simulation was clipped to discard the first $T_{\rm tr} = 250$ time units to keep only the stationary regime. To assess convergence, the following criterion was used

$$\frac{1}{S}\sum_i \sqrt{\frac{1}{\Delta T}\int_{T_{\rm tr}}^{T} {\rm d}t x_i(t)^2 - \left(\frac{1}{\Delta T}\int_{T_{\rm tr}}^{T} {\rm d}t x_i(t)\right)^2} < \frac{0.01}{\sqrt{\Delta T}}, \tag{S11}$$

where $\Delta T = T - T_{\rm tr}$. If positive the traectory was classified as `ufp`. If negative, we used a measure of cyclicity to assess whether the trajectory belonged to the oscillating regime or not. More precisely, for each species, we computed the (moduli of the) Fourier transform of the trajectory, isolated its peaks and discarded any peaks with magnitude less than 5% of the maximum peak. To get rid of numerical effects, the minimal considered frequency was 3 times the sampling frequency. Denote the frequencies of the peaks as $f_1, \ldots, f_p$ and the values of the transforms at those peaks as $\hat{F}_1, \ldots, \hat{F}_p$. We then calculated the ratios $r_i = (f_{i+1} - f_i)/f_1$ and used the following measure of cyclicity

$$\texttt{cyc} = 1 - 2\frac{\sum_j \hat{F}_j \times \{r_j\}}{\sum_j \hat{F}_j}, \tag{S12}$$

where $\{x\}$ is the distance from x to the closest integer to x. Under the assumption that the trajectory is cyclic, all frequencies are multiples of the fundamental frequency and therefore $\texttt{cyc} = 1$. Then the following criterion was used,

$$\frac{1}{S}\sum_i \texttt{cyc}_i > 0.8, \tag{S13}$$

where $\texttt{cyc}_i$ is the cyclicity measure of the trajectory of species i. If met the trajectory was classified as `cyclic`, otherwise as `unstable`.
Although these criteria work in most cases, some misclassification can occur especially close to the the borders between the different phases. This (rather than finite-S effects) is responsible for most outliers in the experimental phase diagram of the main text. In particular, `unstable` trajectories close to the `ufp` phase can have very slow fluctuations, which sometimes leads to them being classified as fixed points or cyclic.

B. Numerical solutions

Numerical solutions for the fixed point equations where obtained by using an iterative scheme. The equations where rewritten in vector form as $\vec{X} = f(\vec{X})$ and the iteration ran as $\vec{X}_{n+1} = (1-r)\vec{X}_n + rf(\vec{X}_n)$ for a fixed number of iterations. We used $r = 0.25$. The initial condition was taken to be $\vec{X}_0 = \vec{1}$. Although we have no mathematical guarantee of convergence for this solver beyond the unstructured case, we find that it to converge in practice as long as a solution exists. Performance-wise, the solver runs extremely fast due to the low-dimensionality of the space on which the equations are rewritten. Except very close to the divergence transition, we find that 200 iterations were more than enough to converge to the solution, although the examples of the main text were solved with 20000 iterations.
For the examples of the article, this simple scheme was enough. For more complicated cases, several solutions to the fixed point equations can exist, corresponding to different structural regimes. In that case, the initial condition plays a role in determining towards which solution the solver converges. For those cases, we used a combination of multiple initial conditions and intuitive guessing to find the solutions, but a more automatic solver should be implemented for future work.

D. OBTAINING THE EFFECTIVE PROCESS WITH DYNAMICAL MEAN FIELD THEORY

We now describe how to go from the deterministic Lotka-Volterra equations with random prescription for the interaction matrix to the set of stochastic single-species processes with dynamical noise. Recalling that the trajectories of single species are random due to the randomness in the interactions the idea is to express their probability distribution

and simplify it to a point where it can be recognized as the probability density associated to a stochastic process. Start by writing the probability of a trajectory $\vec{x}(t)$ as

$$\mathbb{P}\left[\vec{x}(t)\right] = \int \prod_{ij} \mathrm{d}z_{ij} \mathbb{P}\left[z_{ij}\right] \prod_i \delta_f \left(\frac{\dot{x}_i}{x_i} - 1 + x_i - \sum_j \left(\mu_{ij} + \frac{\sigma}{\sqrt{S}} z_{ij} \right) x_j \right), \tag{S14}$$

where δ_f stands for a functional Dirac delta function, enforcing the dynamics at each time step. Formally, this is just a change of variables. Seeing the functional delta as a product of deltas for each time step, we can use the Fourier representation

$$\delta\left(x\right) = \int \frac{\mathrm{d}\hat{x}}{\sqrt{2\pi}} e^{i\hat{x}x}, \tag{S15}$$

so that the probability of a trajectory becomes

$$\mathbb{P}\left[\vec{x}(t)\right] = \int \mathcal{D}\hat{x}(t) \int \prod_{ij} \mathrm{d}z_{ij} \mathbb{P}\left[z_{ij}\right] \exp\left(i \sum_i \int \mathrm{d}t \hat{x}_i(t) \left(\frac{\dot{x_i}(t)}{x_i(t)} - 1 + x_i(t) - \sum_j \left(\mu_{ij} + \frac{\sigma}{\sqrt{S}} z_{ij} \right) x_j(t) \right) \right), \tag{S16}$$

where $\mathcal{D}\hat{x}(t)$ formally stands for a product of pointwise measures $\mathcal{D}\hat{x}(t) = \prod_t \mathrm{d}\hat{x}(t)$. In writing this and in the following, we ignore the prefactors as we will not be interested in eventual normalizing constants. With this representation we can now do the integral over the random matrix entries. Indeed,

$$\int \prod_{ij} \mathrm{d}_{zij} \mathbb{P}\left[z_{ij}\right] \exp\left(i \sum_{ij} \int \frac{\sigma z_{ij}}{\sqrt{S}} \mathrm{d}t \hat{x}_i(t) x_j(t) \right) = \exp\left[-\frac{S\sigma^2}{2} \int \mathrm{d}t\mathrm{d}t' \left(\frac{1}{S} \sum_i \hat{x}_i(t) \hat{x}_i(t') \right) \left(\frac{1}{S} \sum_i x_i(t) x_i(t') \right) \right]. \tag{S17}$$

Define the quantities

$$\mathcal{A}(t,t') = S^{-1} \sum_i x_i(t) x_i(t') \tag{S18}$$

$$\mathcal{B}(t,t') = S^{-1} \sum_i \hat{x}_i(t) \hat{x}_i(t') \tag{S19}$$

which we will later introduce in eq. (S16) using again a Fourier representation for functional deltas,

$$\int \mathcal{D}\hat{\mathcal{A}}(t,t') \mathcal{D}\mathcal{A}(t,t') \mathcal{D}\hat{\mathcal{B}}(t,t') \mathcal{D}\mathcal{B}(t,t') \exp\left(iS \int \mathrm{d}t\mathrm{d}t' \hat{\mathcal{A}}(t,t') \left(\mathcal{A}(t,t') - \frac{1}{S} \sum_i x_i(t) x_i(t') \right) \right) \tag{S20}$$

$$\times \exp\left(iS \int \mathrm{d}t\mathrm{d}t' \hat{\mathcal{B}}(t,t') \left(\mathcal{B}(t,t') - \frac{1}{S} \sum_i \hat{x}_i(t) \hat{x}_i(t') \right) \right). \tag{S21}$$

By assumption, the structural matrix μ_{ij} has low rank. We can therefore write it as

$$\mu_{ij} = S^{-1} \sum_{\lambda=1}^{\mathrm{rk}(\mu)} u_i^{(\lambda)} v_j^{(\lambda)}. \tag{S22}$$

This decomposition is not unique. It can stem from knowledge of the different functions whose superposition gives the matrix μ_{ij}, or by a singular value decomposition. In any case, we define the following quantities

$$f_\lambda(t) = S^{-1} \sum_i v_i^{(\lambda)} x_i(t) \tag{S23}$$

$$n_\lambda(t) = S^{-1} \sum_i u_i^{(\lambda)} \hat{x}_i(t) \tag{S24}$$

which we introduce similarly in eq. (S16) using a Fourier representation,

$$\prod_\lambda \int \mathcal{D}\hat{m}_\lambda(t) \mathcal{D}f_\lambda(t) \mathcal{D}\hat{n}_\lambda(t) \mathcal{D}n_\lambda(t) \exp\left(iS \int \mathrm{d}t \hat{m}(t) \left(m(t) - \frac{1}{S} \sum_i v_i^{(\lambda)} x_i(t) \right) \right) \tag{S25}$$

$$\times \exp\left(iS \int \mathrm{d}t \hat{n}(t) \left(n(t) - \frac{1}{S} \sum_i u_i^{(\lambda)} \hat{x}_i(t) \right) \right). \tag{S26}$$

All in all the probability of a trajectory becomes

$$\mathbb{P}\left[\vec{x}(t)\right] \propto \int \mathcal{D}\mathcal{A}\mathcal{D}\hat{\mathcal{A}}\mathcal{D}\mathcal{B}\mathcal{D}\hat{\mathcal{B}} \prod_{\lambda} \mathcal{D}f_\lambda \mathcal{D}\hat{m}_\lambda \mathcal{D}n_\lambda \mathcal{D}\hat{n}_\lambda \tag{S27}$$

$$\times \exp\left(iS \int \mathrm{d}t\mathrm{d}t' \left(\hat{\mathcal{A}}\mathcal{A} + \hat{\mathcal{B}}\mathcal{B} + \sum_\lambda f_\lambda \hat{m}_\lambda + n_\lambda \hat{n}_\lambda\right)\right) \tag{S28}$$

$$\times \exp\left(-\frac{S\sigma^2}{2} \int \mathrm{d}t\mathrm{d}t' \mathcal{A}\mathcal{B} - S\sum_\lambda \int \mathrm{d}t f_\lambda n_\lambda\right) \tag{S29}$$

$$\times \exp\left(\sum_i \psi_i\right), \tag{S30}$$

where we have used the shorthand

$$\psi_i = \ln\left[\int \mathcal{D}\hat{x}_i \exp\left(i\int \mathrm{d}t\hat{x}_i \left(\frac{\dot{x}_i}{x_i} - 1 + x_i - \sum_\lambda \left(v_i^{(\lambda)}\hat{m}_\lambda + u_i^{(\lambda)}\hat{n}_\lambda\right)\right) - i\int \mathrm{d}t\mathrm{d}t' \left(x_i(t)\hat{\mathcal{A}}(t,t')x_i(t') + \hat{x}_i(t)\hat{\mathcal{B}}(t,t')\hat{x}_i(t')\right)\right)\right] \tag{S31}$$

All three exponents in eq. (S30) are of order $\mathcal{O}(S)$, so that in the limit $S \to \infty$ we can calculate the functional integrals via the saddle-point method. We start by extremizing w.r.t. the unhatted variables,

$$\frac{\partial}{\partial \mathcal{A}(t,t')} = i\hat{\mathcal{A}}(t,t') - \frac{\sigma^2}{2}\mathcal{B}(t,t') = 0 \Rightarrow \hat{\mathcal{A}} = -\frac{i\sigma^2}{2}\mathcal{B} \tag{S32}$$

$$\frac{\partial}{\partial \mathcal{B}(t,t')} = i\hat{\mathcal{B}}(t,t') - \frac{\sigma^2}{2}\mathcal{A}(t,t') = 0 \Rightarrow \hat{\mathcal{B}} = -\frac{i\sigma^2}{2}\mathcal{A} \tag{S33}$$

$$\frac{\partial}{\partial f_\lambda(t,t')} = i\hat{m}_\lambda(t,t') - n_\lambda(t,t') = 0 \Rightarrow \hat{m}_\lambda = n_\lambda \tag{S34}$$

$$\frac{\partial}{\partial n_\lambda(t,t')} = i\hat{n}_\lambda(t,t') - f_\lambda(t,t') = 0 \Rightarrow \hat{n}_\lambda = f_\lambda. \tag{S35}$$

For the hatted variables we get

$$\frac{\partial}{\partial \hat{\mathcal{A}}(t,t')} = i\mathcal{A}(t,t') - iS^{-1}\sum_j \left\langle x_j(t)x_j(t')\right\rangle_i = 0 \Rightarrow \mathcal{A} = S^{-1}\sum_j \left\langle x_j(t)x_j(t')\right\rangle_j \tag{S36}$$

$$\frac{\partial}{\partial \hat{\mathcal{B}}(t,t')} = i\mathcal{B}(t,t') - iS^{-1}\sum_j \left\langle \hat{x}_j(t)\hat{x}_j(t')\right\rangle_i = 0 \Rightarrow \mathcal{B} = S^{-1}\sum_j \left\langle \hat{x}_j(t)\hat{x}_j(t')\right\rangle_j \tag{S37}$$

$$\frac{\partial}{\partial \hat{m}_\lambda(t)} = if_\lambda(t) - iS^{-1}\sum_j v_j^{(\lambda)} \left\langle x_j(t)\right\rangle_j = 0 \Rightarrow f_\lambda = S^{-1}\sum_j v_j^{(\lambda)} \left\langle x_j\right\rangle_j \tag{S38}$$

$$\frac{\partial}{\partial \hat{n}_\lambda(t)} = in_\lambda(t) - iS^{-1}\sum_j u_j^{(\lambda)} \left\langle \hat{x}_j(t)\right\rangle_j = 0 \Rightarrow n_\lambda = S^{-1}\sum_j u_j^{(\lambda)} \left\langle \hat{x}_j\right\rangle_j, \tag{S39}$$

where for some functional $\mathcal{O}[x_j, \hat{x}_j]$ we have introduced the notation

$$\left\langle \mathcal{O}\right\rangle_j = \frac{\int \mathcal{D}\hat{x}_j \mathcal{O}[x_j,\hat{x}_j] \exp\left(i\int \mathrm{d}t\hat{x}_j \left(\frac{\dot{x}_j}{x_j} - 1 + x_j - \sum_\lambda \left(v_j^{(\lambda)}\hat{m}_\lambda + u_j^{(\lambda)}\hat{n}_\lambda\right)\right) - i\int \mathrm{d}t\mathrm{d}t' \left(x_j(t)\hat{\mathcal{A}}(t,t')x_j(t') + \hat{x}_j(t)\hat{\mathcal{B}}(t,t')\hat{x}_j(t')\right)\right)}{\int \mathcal{D}\hat{x} \exp\left(i\int \mathrm{d}t\hat{x}_j \left(\frac{\dot{x}_j}{x_j} - 1 + x_j - \sum_\lambda \left(v_j^{(\lambda)}\hat{m}_\lambda + u_j^{(\lambda)}\hat{n}_\lambda\right)\right) - i\int \mathrm{d}t\mathrm{d}t' \left(x_j(t)\hat{\mathcal{A}}(t,t')x_j(t') + \hat{x}_j(t)\hat{\mathcal{B}}(t,t')\hat{x}_j(t')\right)\right)} \tag{S40}$$

In particular one has

$$\left\langle \hat{x}_j(t)\hat{x}_j(t')\right\rangle_j = \left\langle \hat{n}_\lambda\right\rangle = 0, \tag{S41}$$

and therefore, plugging in the values obtained from the saddle-point conditions

$$\left\langle \mathcal{O}\right\rangle_j = \frac{\int \mathcal{D}\hat{x}_j \mathcal{O}[x_j,\hat{x}_j] \exp\left(i\int \mathrm{d}t\hat{x}_j \left(\frac{\dot{x}_j}{x_j} - 1 + x_j - \sum_\lambda u_j^{(\lambda)} f_\lambda\right) - \frac{\sigma^2}{2}\int \mathrm{d}t\mathrm{d}t' \hat{x}_j(t)\mathcal{A}(t,t')\hat{x}_j(t')\right)}{\int \mathcal{D}\hat{x}_j \exp\left(i\int \mathrm{d}t\hat{x}_j \left(\frac{\dot{x}_j}{x_j} - 1 + x_j - \sum_\lambda u_j^{(\lambda)} f_\lambda\right) - \frac{\sigma^2}{2}\int \mathrm{d}t\mathrm{d}t' \hat{x}_j(t)\mathcal{A}(t,t')\hat{x}_j(t')\right)}. \tag{S42}$$

Similarly the probability simplifies to

$$\mathbb{P}\left[\vec{x}(t)\right] \propto \prod_j \int \mathcal{D}\hat{x}_i \exp\left(i\int \mathrm{d}t\hat{x}_j\left(\frac{\dot{x}_j}{x_j} - 1 + x_j - \sum_\lambda u_j^{(\lambda)} f_\lambda\right) - \frac{\sigma^2}{2}\int \mathrm{d}t\mathrm{d}t'\hat{x}_j(t)\mathcal{A}(t,t')\hat{x}_j(t')\right). \tag{S43}$$

In particular the averages defined in eq. (S42) are in fact averages with respect to the probability measure induced by the trajectories themselves, i.e. averages like those defining f_λ and $\mathcal{A}$ are to be self-consistently obtained from the effective dynamics that we are deriving with DMFT.
Because eq. (S43) appears as a product over all species, it follows that each species undergoes stochastic dynamics independent from the rest. Moreover, focusing on species j, the probability corresponds to the following stochastic process

$$\dot{x}_j = x_j\left(1 - x_j + \sum_\lambda u_j^{(\lambda)} f_\lambda + \sigma\zeta_j(t)\right), \tag{S44}$$

where ζ_j is a colored Gaussian process with correlation $\langle\zeta_j(t)\zeta_j(t')\rangle = \mathcal{A}(t,t')$. The dynamics are now simplified to a set of independent one-dimensional stochastic processes. Yet, they are not fully solved since the variables $\mathcal{A}(t,t')$ and $f_\lambda(t)$ are self-consistently defined from the stochastic processes themselves. Solving this loop analytically is not possible.

E. RELATIONSHIP BETWEEN DIFFERENT TYPES OF AVERAGES

Throughout our calculationn we will encounter three kinds of averages:

1. with respect to the randomness in the interaction coefficients.
2. with respect to the effective stochastic process, denoted by $\langle\cdot\rangle$.
3. with respect to species, denoted by $\bar{\cdot}$.

Because the stochastic process is an effective representation of the randomness in the interaction, averages w.r.t. it imply averages over the randomness in the interaction coefficients. If the initial abundances of the system are chosen randomly, then the average over the stochastic process also includes that kind of randomness. Otherwise, averages w.r.t. the stochastic process are equivalent to averages w.r.t. the interactions. This is a crucial ingredient of DMFT. Moreover, in situations where memory of the initial conditions is lost (such as in the Unique Fixed Point phase), averages this equivalence is true regardless of how initial conditions have been chose (randomly or not).
On the other hand, averages w.r.t. species are done for a given realization of the system, i.e. for a given realization of the randomness in the coefficients, and involve expressions of the form

$$\overline{f} = \frac{1}{S}\sum_i f_i, \tag{S45}$$

where f_i is any function of species i. These are in principle different from averages w.r.t. interactions/stochastic processes. However, because the stochastic processes of different species are independent, the central limit theorem ensures that

$$\overline{f} = \frac{1}{S}\sum_i f_i = \frac{1}{S}\sum_i \langle f_i\rangle = \overline{\langle f\rangle}. \tag{S46}$$

In other words, averages over species imply averages over the stochastic process. In unstructured ecosystems, the statistical equivalence of species implies that the converse is also true. In our case, structure makes species different, so that the equivalence no longer holds.

F. OBTAINING THE EQUILIBRIUM EQUATIONS

We now show how to go from the effective process eq. (S44) to the equilibrium equations. Under the assumption that the system indeed reaches a fixed point, the aforementioned loop can be solved mostly analytically.

Suppose that the system admits a fixed point with abundances $\{x_j^\star\}$. From the definition of $\mathcal{A}$,

$$\mathcal{A}(t,t') \xrightarrow[t,t'\to\infty]{} C^\star \equiv S^{-1}\sum_j \left\langle \left(x_j^\star\right)^2 \right\rangle_j = \overline{(x^\star)^2}, \tag{S47}$$

where the overline now indicates an average over the species. Note that because the processes are independent across species, averaging over species implies averaging over the dynamics of each species. Similarly,

$$f_\lambda(t) \xrightarrow[t\to\infty]{} f_\lambda^\star \equiv S^{-1}\sum_j v_j^{(\lambda)} \left\langle x_j^\star \right\rangle_j = \overline{v^{(\lambda)}x^\star}. \tag{S48}$$

Since $\mathcal{A}(t,t')$ is the correlator of the processes ζ_j, its convergence implies that all the processes ζ_j converge at infinity to some (random) value $\zeta_j^\star$. Let $\xi_j^\star = \zeta_j^\infty/\sqrt{C^\star}$. Then equating to zero the derivative in eq. (S44) gives

$$x_j^\star = 1 + \sum_\lambda u_j^{(\lambda)} f_\lambda^\star + \gamma\xi_j^\star \text{ or } x_j^\star = 0. \tag{S49}$$

We recall that $\gamma = \sigma\sqrt{C^\star}$. The first solution is only feasible if it is positive. On the other hand, it can be easily checked that the second solution is unstable as soon as the first one is positive. Hence the general solution is

$$x_j^\star = \max\left(0, 1 + \sum_\lambda u_j^{(\lambda)} f_\lambda^\star + \gamma\xi_j^\star\right), \tag{S50}$$

In particular, $\xi_j^\star$ is a Gaussian random variable with zero mean and unit variance.
Now that the law of each species's equilibrium value is obtained as a function of $f_\lambda^\star$ and γ, we obtain the self-consistent equations for these two variables. By definition

$$f_\lambda^\star = \overline{v^{(\lambda)}x^\star} = \overline{\int \mathcal{D}z \max\left(0, 1 + \sum_\lambda u^{(\lambda)} f_\lambda^\star + \gamma z\right)} = \gamma\overline{v^{(\lambda)}\int \mathcal{D}z \max\left(0, x_+/\gamma + z\right)} \tag{S51}$$

$$C^\star = \overline{(x^\star)^2} = \overline{\int \mathcal{D}z \max\left(0, 1 + \sum_\lambda u^{(\lambda)} f_\lambda^\star + \gamma z\right)^2} = \gamma^2\overline{\int \mathcal{D}z \max\left(0, x_+/\gamma + z\right)^2}, \tag{S52}$$

where x^+, γ are defined in the main text and $\mathcal{D}z$ now represents the probability measure of a standard Gaussian variable. These are the equations given in the main text up to the definition of $\omega_1(\cdot)$ and $\omega_2(\cdot)$.

G. DMFT WITH NON-UNIFORM VARIANCES IN THE RANDOM PART

We now briefly describe how to extend the previous calculation to the case where the random part of the interaction coefficients has a non-uniform variance. In other words, whenever $\langle z_{ij}^2\rangle = \sigma_{ij}^2$. Our calculation can be straightforwardly extended to accomodate virtually any σ_{ij} pattern, but for the reasons stated in the main text this generality is not our goal here. We limit ourselves to the case of a rank-one standard deviation matrix that can be written $\sigma_{ij} = \sigma\kappa_i\nu_j$. The relevance of this case is explained in the main text. In short, it arises whenever carrying capacities are not uniform and a rescaling by the carrying capacities is performed. In that case, the interaction coefficients are rescaled from A_{ij} to $A_{ij}K_j/K_i$, leading to a rescaling of the strength of hidden interactions from $\sigma_{ij} = \sigma$ to $\sigma_{ij} = \sigma K_j/K_i$. Hence the effective equations for this case can be obtained by taking $\kappa_i = 1/K_i$ and $\nu_j = K_j$ in what follows.
The main difference in the DMFT calculation comes from the integral w.r.t. interaction coefficients eq. (S17), which now reads

$$\int \prod_{ij} \mathrm{d}_{z_{ij}}\mathbb{P}\left[z_{ij}\right] \exp\left(i\sum_{ij}\int \frac{\sigma z_{ij}}{\sqrt{S}}\mathrm{d}t\hat{x}_i(t)x_j(t)\right) = \exp\left[-\frac{S\sigma^2}{2}\int \mathrm{d}t\mathrm{d}t'\left(\frac{1}{S}\sum_i \kappa_i^2\hat{x}_i(t)\hat{x}_i(t')\right)\left(\frac{1}{S}\sum_i \nu_i^2 x_i(t)x_i(t')\right)\right]. \tag{S53}$$

This implies a straightforward modification of the parameters $\mathcal{A}$ and $\mathcal{B}$, and the rest of the calculation is essentially unchanged. In the end, the effective stochastic process is

$$\dot{x}_j = x_j\left(1 - x_j + \sum_\lambda u_j^{(\lambda)} f_\lambda + \sigma\kappa_j\zeta_j(t)\right), \tag{S54}$$

and the correlator of ζ_i is now $C_\nu(t,t') \equiv \langle \zeta_i(t)\zeta_i(t')\rangle = \overline{\nu^2 x(t)x(t')}$. In particular the equilibrium distribution is changed to

$$x_j^\star = \max\left(0, 1 + \sum_\lambda u_j^{(\lambda)} f_\lambda^\star + \gamma_\nu \kappa_j \xi_j^\star\right), \tag{S55}$$

where we have defined $\gamma_\nu = \sigma\sqrt{C_\nu^\star}$. The fixed point equations read

$$f_\lambda^\star = \overline{v^{(\lambda)} x^\star} = \overline{v^{(\lambda)} \int \mathcal{D}z \max\left(0, 1 + \sum_\lambda u^{(\lambda)} f_\lambda^\star + \kappa\gamma_\nu z\right)} = \gamma_\nu\ \overline{\kappa v^{(\lambda)} \omega_1(x^+/\kappa\gamma_\nu)} \tag{S56}$$

$$C_\nu^\star = \overline{\nu^2 \left(x^\star\right)^2} = \overline{\nu^2 \int \mathcal{D}z \max\left(0, 1 + \sum_\lambda u^{(\lambda)} f_\lambda^\star + \kappa\gamma_\nu z\right)^2} = \gamma_\nu^2\ \overline{\nu^2\kappa^2\omega_2(x^+/\kappa\gamma_\nu)}. \tag{S57}$$

H. OBTAINING THE STABILITY CONDITIONS

The stability can be analyzed by adding an initial perturbation to the abundances of the fixed point and checking whether the system returns to the fixed point or not. For small perturbations, we do so by linearizing the effective dynamics around the fixed point. This is therefore equivalent to checking the stability of the Jacobian. The linearized stochastic differential equation reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta x_j(t) = \delta x_j(t)\left(1 - x_j^\star + \sum_\lambda u_j^{(\lambda)} f_\lambda^\star + \gamma\xi_j^\star\right) + x_j^\star\left(-\delta x_j(t) + \sum_\lambda u_j^{(\lambda)}\delta f_\lambda(t) + \sigma\delta\zeta_j(t)\right), \tag{S58}$$

where we use the notation $\delta_j x(t) = x_j(t) - x_j^\star$. The fluctuations δf_λ and $\delta\xi_j$ are self-consistently induced by the fluctuations in the $\delta x_j(t)$'s, just like f_λ and ξ_j are induced by the x_j's. As such that they have to obey the relations

$$\delta f_\lambda(t) = \overline{v^{(\lambda)}\delta x(t)} \tag{S59}$$

$$\langle \xi_j(t)\xi_j(t')\rangle = \overline{\delta x(t)\delta x(t')}. \tag{S60}$$

The first term in the r.h.s. of eq. (S58) is zero for extant species, due to eq. (S50). It is negative for extinct species for the same reason. As such, it can never cause an instability. The second term is zero for extinct species and non-zero for extant species. As such, only the second term can cause the instability. We are therefore left to study the following stochastic process *over the set of extant species*

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta x_j(t) = x_j^\star\left(-\delta x_j(t) + \sum_\lambda u_j^{(\lambda)}\delta f_\lambda(t) + \sigma\delta\xi_j(t)\right). \tag{S61}$$

Given that the ξ_j's are Gaussian processes, the $\delta\xi_j$'s are also Gaussian processes. We can therefore fully characterize themselves through their mean and two-point correlations. In particular, taking the mean and using eq. (S60) gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\delta x_j(t)\rangle = x_j^\star\left(-\langle\delta x_j(t)\rangle + \sum_{\lambda,i} u_j^{(\lambda)} v_i^{(\lambda)}\langle\delta x_i(t)\rangle\right) \Rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left\langle\vec{\delta x}\right\rangle = \mathcal{J}\left\langle\vec{\delta x}\right\rangle, \tag{S62}$$

where the matrix $\mathcal{J}$ is defined as

$$\mathcal{J}_{ij} = -x_i^\star\left(\delta_{ij} - \frac{\mu_{ij}}{S}\right). \tag{S63}$$

Positive eigenvalues of $\mathcal{J}$ indicate an instability of the system. Note that it makes no difference to eigenvalues whether we include or exclude the extinct species in the definition of this matrix. As such, we will include them to be able to work with our knowledge of the full structure matrix. Expanding the parenthesis in eq. (S63), the first bit is a diagonal matrix whose entries are the (neg-) equilibrium abundances. As such, all its eigenvalues are negative, thus stable. The second bit is a low-rank matrix given that μ is low-rank. Its eigenvalues are related, but not equal to, the eigenvalues of μ. Our goal is to compute how the eigenvalues of $\mathcal{J}$ relate to the eigenvalues of μ. To do so we probe

the values z for which the inverse $(z-\mathcal{J})^{-1}$ is ill-defined.
We decompose $z-\mathcal{J}$ as

$$z-\mathcal{J}=\underbrace{z+\mathrm{diag}\,(x^\star)}_{\alpha}-\underbrace{\mathrm{diag}\,(x^\star)\,U}_{\beta}\,D\,\underbrace{V}_{\gamma}, \tag{S64}$$

where we have used a decomposition $\mu=UDV$. This can be obtained either from a singular value decomposition or from the knowledge of μ as a sum of rank-one contributions. Note in particular that if the $\mathrm{rank}(\mu)=k$ then D can be taken to be a $k\times k$ matrix. Using Woodbury's matrix identity

$$(z-\mathcal{J})^{-1}=\alpha^{-1}+\alpha^{-1}\beta\left(D^{-1}-\gamma\alpha^{-1}\beta\right)^{-1}\gamma\alpha^{-1}. \tag{S65}$$

In particular, outside the bulk of eigenvalues of α (which are the abundances), the invertibility of $z-\mathcal{J}$ is equivalent to that of $D^{-1}-\gamma\alpha^{-1}\beta$, which is now a $k\times k$ matrix. Rewriting explicitly, z is an outlier eigenvalue of $\mathcal{J}$ iff

$$\left(D^{-1}\right)_{ab}-\overline{v^{(a)}u^{(b)}\frac{x^\star}{z+x^\star}} \tag{S66}$$

defines a singular matrix. As usual, the overline denotes an averages over species indices. The condition in the main text can then be obtained by supposing that the decomposition is of the form $\mu=UV$, in other words $D=\mathbb{I}$. Note that any decomposition of μ works. In particular, the one given here need not be the simplest one to work with.
We now turn our attention to the instability condition associated to the correlations. For this one, it is convenient to take the Laplace transform of the linearized effective process. This yields

$$s\hat{\delta x}_j(s)=x_j^\star\left(-\hat{\delta x}_j(s)+\sum_\lambda u_j^\lambda\hat{\delta f}_\lambda(s)+\sigma\hat{\delta\xi}_j(s)\right)+\delta x_j(0), \tag{S67}$$

where $\delta x_j(0)$ is the initial perturbation to the abundance of species j. Given the Final Value Theorem for Laplace transforms, stability requires that

1. $\hat{\delta x}_j(s)$ has no poles on the right half of the complex plane.
2. $s\hat{\delta x}_j(s)\to 0$ when $s\to 0$.

Without loss of generality, we can assume that we deal with a set of parameters for which the mean is stable in the sense of eq. (S62). In that case $\left\langle s\hat{\delta x}_j(s)\right\rangle\to 0$ as $s\to 0$ and therefore $s\hat{\delta f}_\lambda(s)\to 0$ as well. Now multiply by s, take squared moduli and average in eq. (S67),

$$\left|\frac{s}{x_j^\star}+1\right|^2\left\langle\left|s\hat{\delta x}_j(s)\right|^2\right\rangle=\left|\sum_\lambda u_j^{(\lambda)}s\hat{\delta f}_\lambda(s)\right|^2+\sigma^2\left\langle\left|s\hat{\delta\xi}_j(s)\right|^2\right\rangle+|s|^2\delta x_j(0)^2. \tag{S68}$$

The loop is closed using the following self-consistency condition between the $\hat{\delta x}_j(s)$'s and the $\hat{\delta\xi}_j(s)$'s that stems from eq. (S60)

$$\left\langle\left|\hat{\delta\xi}_j(s)\right|^2\right\rangle=S^{-1}\sum_j\left\langle\left|\hat{\delta x}_j(s)\right|^2\right\rangle=\overline{\left|\hat{\delta x}(s)\right|^2}. \tag{S69}$$

As usual, the overline denotes an average over species, which in turn implies also the average $\langle\rangle$ over the effective process. Averaging over species in eq. (S68) gives

$$\overline{\left|s\hat{\delta x}(s)\right|^2}=\frac{\left|\sum_\lambda u_j^{(\lambda)}s\hat{\delta f}_\lambda(s)\right|^2+|s|^2\overline{\delta x(0)^2}}{\left|1+\frac{s}{x^\star}\right|^2-\sigma^2\phi}. \tag{S70}$$

The fraction of extant species ϕ appears because the average over species includes both extant and extinct species, but extinct species have a null contribution. Now take the limit $s\to 0$. By assumption, the first term in the r.h.s. of eq. (S68) vanishes and we therefore have

$$\overline{\left|s\hat{\delta x}(s)\right|^2}\sim\frac{|s|^2\overline{\delta x(0)^2}}{1-\sigma^2\phi}\to 0, \tag{S71}$$

which is vanishing, unless the condition $1 = \sigma^2\phi$ is met. This signals the instability. For $\sigma^2\phi > 1$ the condition is not met, but the function has a pole on the right half of the complex plane, so there is a divergence nonetheless. For $\sigma^2\phi < 1$ the expression goes to zero and has no poles on the right half of the complex plane, so there is no divergence. Note that eq. (S71) was obtained by taking averages over species, so one could wonder whether instabilities could happen for individual species that are not captured by this analysis. This is not the case given eq. (S68). Indeed,

$$\left\langle \left| s\hat{\delta x}_j(s) \right|^2 \right\rangle \sim_{s\to 0} \frac{\sigma^2 |s|^2 \overline{\delta x(0)^2}}{1-\sigma^2\phi} + |s|^2 \delta x_j(0)^2, \tag{S72}$$

and therefore the same analysis applies to each individual species. In other words, this divergence occurs for all species at the same point.